



Universidad
Politécnica
de Cartagena



ETS
Ingeniería de
Telecomunicación

TRABAJO FIN DE GRADO
EN
INGENIERÍA DE SISTEMAS DE TELECOMUNICACIONES

“Extracción de firmas acústicas de embarcaciones mediante descriptores
frecuenciales”

“Extraction of acoustic signatures of ships through frequency descriptors”

Autora: Rocío Sabariego Arenas

Director: Juan Morales Sánchez

Convocatoria de Septiembre 2022

Índice

1. Resumen-----	4
2. Abstract-----	5
3.Introducción-----	6
3.1 Objetivos-----	6
4. Base de datos y métodos-----	7
4.1 Base de datos-----	7
4.1.1 Diferencias entre las embarcaciones. -----	7
4.2 ¿Qué es el espectro de potencia?-----	9
4.3 Descriptores frecuenciales-----	10
4.3.1 Coeficientes Cepstrales de las Frecuencias de Mel -----	10
4.3.2 Coeficientes cepstrales gammatones.-----	23
4.3.3 Comparación de los GTCC y los MFCC-----	36
4.4. Machine learning o aprendizaje automático -----	37
4.4.1 Aprendizaje supervisado-----	37
4.4.2 Aprendizaje no supervisado -----	38
4.4.3 Aprendizaje por refuerzo -----	39
4.4.4 Métodos de clasificación que se van a usar. -----	39
4.4.4.1 Clasificación mediante árboles de decisión. -----	39
4.4.4.2 Clasificación mediante redes neuronales convolucionales, CNN. -----	40
4.5 Definiciones de métricas -----	43
4.6 Construcción de la base de datos. -----	44
4.5.1 Construcción de la base de datos para descriptores frecuenciales MFCC para árbol de decisión-----	46
4.5.2 Construcción de la base de datos para descriptores frecuenciales GTCC -----	47
4.5.3 Construcción de base de datos para red neuronal convolucional. -----	48
4.5.4 Esquema resumen referido a la construcción de la base de datos. -----	51
5. Entrenamiento y resultados. -----	52
5.1 Ejercicios de entrenamiento con árboles de decisión-----	52
5.1.1 Caso de estudio 1. Caso binario. Hay barco, no hay barco. -----	52
5.1.1.1 Resultados obtenidos para los descriptores MFCC. -----	52
5.1.1.2 Resultados obtenidos para los descriptores GTCC.-----	54
5.1.1.3 Comparación de resultados para el caso binario.-----	55
5.1.2 Caso de estudio 2. No hay barco, es barco pequeño o es barco grande. -----	55
5.1.2.1 Resultados obtenidos para los descriptores MFCC. -----	56
5.1.2.2 Resultados obtenidos para los descriptores GTCC.-----	57

5.1.2.3 Comparación de resultados para el caso de estudio 2.-----	58
5.1.3 Caso de estudio 3. Caso multisalida. Todos los tipos de barcos-----	59
5.1.3.1 Resultados obtenidos para los descriptores MFCC.-----	59
5.1.3.2 Resultados obtenidos para los descriptores GTCC.-----	61
5.1.3.3 Comparación de resultados para el caso de estudio 3.-----	62
5.1.4 Uso de la herramienta SMOTE y resultados para los tres casos de estudio.-----	63
5.2 Ejercicios de entrenamiento con red neuronal convolucional-----	69
5.2.1 Caso de estudio 1. Caso binario. Hay barco no hay barco-----	69
5.2.1.1 Resultados obtenidos para los descriptores MFCC-----	69
5.2.1.2 Resultados obtenidos para los descriptores GTCC-----	71
5.2.1.3 Comparación de resultados para el caso binario-----	73
5.2.2 Caso de estudio 2. No hay barco, hay barco grande o hay barco pequeño-----	74
5.2.2.1 Resultados obtenidos para los descriptores MFCC-----	74
5.2.2.2 Resultados obtenidos para los descriptores GTCC-----	77
5.2.2.3 Comparación de resultados para el caso de estudio 2-----	79
5.2.3 Caso de estudio 3. Caso multisalida. Todos los barcos-----	80
5.2.3.1 Resultados obtenidos para los descriptores MFCC-----	80
5.2.3.2 Resultados obtenidos para los descriptores GTCC-----	83
5.2.3.3 Comparación de resultados para el caso de estudio 3-----	86
6.Conclusiones.-----	88
7. Bibliografía-----	90

1. Resumen

Introducción: Las firmas acústicas de las embarcaciones es un término con el cual se hace referencia a unas combinaciones acústicas de emisores de sonido, en el caso de estudio de este TFG se trata de embarcaciones marítimas. Dichas firmas acústicas serán extraídas para la identificación de embarcaciones y gracias a ello se podrá hacer uso de ellas en futuras aplicaciones de detección y clasificación de embarcaciones.

Objetivos: el objetivo principal ha consistido en la extracción de firmas acústicas de las embarcaciones, mediante el uso de descriptores frecuenciales, para una futura identificación y clasificación de dichas embarcaciones.

Métodos: Para ello se ha analizado una base de datos, de la cual se ha realizado un estudio para poder extraer sus firmas acústicas mediante el uso de los descriptores frecuenciales MFCC Y GTCC. Posteriormente se ha realizado un trabajo de aprendizaje automático con dichos descriptores para poder así realizar una identificación de embarcaciones dentro de la base de datos y a su vez una clasificación de dichas embarcaciones.

Resultados: Con el uso del aprendizaje automático basado en árboles de decisión, se ha alcanzado para el caso de estudio binario en ambos tipos de descriptores, MFCC y GTCC un valor de media respecto a la identificación de un 97% y 98% respectivamente y con el uso de SMOTE, 97% y 97%; en el caso de estudio 2, el cual consiste en clasificar entre si la embarcación es grande, pequeña o no hay embarcación, un valor de media de 85% y 82% respectivamente y con el uso de SMOTE, 81% y 81%; en el caso de estudio 3, correspondiente a la clasificación de todas las clases de embarcación, un valor de media de 65% y 69% respectivamente y con el uso de SMOTE, un 63% y 57%. Con el uso del aprendizaje automático basado en redes neuronales convolucionales, se han alcanzado los siguientes datos para ambos tipos de descriptores MFCC y GTCC. En el caso de estudio binario se han obtenido unos valores de media respecto a la identificación de un 96% y 98% respectivamente; en el caso de estudio 2 se han obtenido unos valores del 73% y 89% respectivamente y en el caso de estudio 3 se han obtenido unos valores medios de clasificación de un 88% y 80% respectivamente

Conclusiones: El uso de estos descriptores frecuenciales es útil para la identificación de embarcaciones, siendo los GTCC generalmente más eficaces a la hora de realizar la identificación y clasificación de las embarcaciones. Al igual de que el uso de redes neuronales convolucionales es más favorecedor a la hora de realizar clasificaciones multisalida con respecto a los árboles de decisión.

2. Abstract

Introduction: Acoustic signatures of ships is a concept that refers to acoustic combinations of sound emitters, in this case of study we are talking about maritime ships. These acoustic signatures will be extracted for the identification of ships and thanks to them, they can be used in future applications for the detection and classification of ships.

Objectives: The main objective was to extract acoustic signatures of ships, using frequency descriptors, for future identification and classification of these ships.

Methods: For this purpose, a database has been analysed and a study has been performed in order to extract their acoustic signatures by using the frequency descriptors MFCC and GTCC. Subsequently, machine learning work has been applied to these descriptors in order to be able to identify the ships within the database and, in turn, to classify these ships.

Results: With the use of machine learning, based on decision trees, we have achieved for the binary case study in both types of descriptors, MFCC and GTCC an average value with respect to the identification of 97% and 98%, respectively, and with the use of SMOTE, 97% and 97%; in study case 2 a mean value of 85% and 82%, respectively, and with the use of SMOTE, 81% and 81%; in study case 3 a mean value of 65% and 69%, respectively, and with the use of SMOTE, 63% and 57%. With the use of machine learning, based on convolutional neural networks, the following data have been achieved for both types of MFCC and GTCC descriptors. In the binary case study, an average identification value of 96% and 98% has been obtained, respectively; in case study 2, an average identification value of 73% and 89% has been obtained, respectively; and in case study 3, an average classification value of 88% and 80% has been obtained, respectively.

Conclusions: The use of these frequency descriptors is useful for ship identification, with GTCCs generally being more effective in performing ship identification and classification. Likewise, the use of convolutional neural networks is more favourable for multi-output classifications than decision trees.

3.Introducción

En referente a este TFG se estudia la implementación, adaptación y evaluación de métodos de extracción de firmas acústicas de embarcaciones mediante el uso de descriptores frecuenciales.

Dichas firmas acústicas pueden ser calculadas mediante diferentes descriptores frecuenciales, los cuales deberán de obtenerse gracias al uso de herramientas como los espectros de potencia y la DFT. Se prestará una mayor atención a las representaciones que estén basadas en el mel-frequency cepstrum, gracias al cual, se obtendrán los descriptores MFCC y GTCC.

Ambos descriptores mencionados con anterioridad serán usados para la detección y a su vez identificación de diferentes tipos de embarcaciones. Para poder realizar este trabajo de identificación de clasificación, será necesario el uso de una base de datos con la cual se deberá de entrenar nuestra red neuronal artificial, la cual está basada en aprendizaje supervisado. Dicha base de datos es la base de datos 'ShipsEar: An underwater vessel noise database', Applied Acoustics. Volume 113, 1 December 2016, Pages 64–69)

La motivación para realizar dicho estudio se debe al posterior perfeccionamiento e implementación de él en diferentes aplicaciones, en las cuales se pueda detectar y clasificar los diferentes tipos de embarcaciones que estén entrando y/o saliendo de un lugar concreto. Un ejemplo podría ser el uso de aplicaciones en tiempo real que, por reconocimiento de diferentes audios obtenidos en la entrada de un puerto, pudiera obtener diferentes tipos de características de la embarcación deseada.

3.1 Objetivos.

Los objetivos principales que se ha querido abordar durante la elaboración de este trabajo han sido los siguientes:

- Estudio sobre los descriptores frecuenciales haciendo uso de herramientas como el espectro de potencia y la transformada de Fourier discreta, DFT.
- Extracción de diferentes tipos de descriptores frecuenciales, en nuestro caso los GTCC y los MFCC, para el estudio de las diferentes firmas acústicas que se pueden ser obtenidas mediante el sonido de las embarcaciones en el fondo marino.
- Análisis de las diferencias y similitudes que podemos observar de dichas firmas acústicas con relación a los diferentes tipos de embarcaciones.
- Comparación y evaluación de los métodos de extracción de las firmas acústicas mediante los diferentes descriptores frecuenciales.
- Comprobación de la capacidad de poder detectar las embarcaciones entre el ruido captado.
- Evaluar las posibilidades de poder identificar los diferentes tipos de embarcación únicamente con el sonido obtenido.

4. Base de datos y métodos

4.1 Base de datos

La base de datos (Santos Domínguez, Torres Guijarro, Cardenal López, & Peña Gimenez, 2016), está formada por un conjunto de ochenta y tres audios, cuyo contenido es el sonido del fondo marino, en los cuales se pueden diferenciar el ruido del mar y del oleaje, diferentes tipos de embarcaciones o el sonido de diferentes fenómenos meteorológicos sobre la superficie marítima.

El conjunto de audios viene distribuido de la siguiente manera en referente al tipo de embarcación o en su defecto fenómeno meteorológico que se pueda observar.

Tipo de embarcación		Tiempo total de cada clase	Duración mínima	Duración máxima
Dresger	Barco dragador	4min 22s	30s	1min 2s
Fishboat	Barco de pesca	8min 34s	53s	2min 57s
Motorboat	Lancha motora	14min 2s	23s	2min 4s
Musselboat	Barco mejillonero	12min 13s	11s	5min 14s
Natural Ambient	Sonido ambiente	17min 38s	34s	2min
Ocean liner	Trasatlántico	15min 42s	9s	6min 40s
Passengers	Crucero, ferry	57min 36s	1min	5min 49s
Pilot ship	Barco piloto	2min 18s	44s	1min 34s
RORO	Mercancías	24min 14s	30s	11min 22s
Sail botat	Barco de vela	6min 48s	19s	4min 57s
Trawler	Yate	2min 43s	2min 43s	2min 43s
Tugboat	Remolcadora	3min 26s	1min 4s	2min 22s

Tabla 1: Tabla ilustrativa de la base de datos

Por lo que se puede ver en la tabla anterior tenemos un conjunto de 11 tipos distintos de embarcación y a su vez un único grupo de donde se engloban todos los fenómenos meteorológicos en cual se denominará “Natural Ambient”

4.1.1 Diferencias entre las embarcaciones.

Para poder intentar identificar las diferentes embarcaciones se debe de tener en cuenta que diferencias hay entre ellas.

Respecto a nuestra base de datos el único elemento que se obtiene para poder diferenciar una embarcación de otra depende del tipo de motor que tenga cada una de las embarcaciones.

Las embarcaciones de menor tamaño como pueden ser barcos de recreo o lanchas motoras han de tener un motor de menor potencia respecto a embarcaciones como pueden ser trasatlánticos o barcos de mercancías. Eso da lugar a que los barcos con menor potencia, al transmitir su sonido a través del agua lo haga en un conjunto de frecuencias mayor, mientras que los barcos con un motor que genere una mayor potencia, el sonido que transmitirán será en un ancho de frecuencias más bajo.

Teniendo en cuenta esta información se ha clasificado nuestra base de datos en tres casos distintos de estudios.

- Caso binario: en esta clasificación únicamente tendremos en cuenta si encontramos una embarcación en el registro de audio o es sonido ambiente.

Hay embarcación	Tiempo total: 151min 56s	No hay embarcación	Tiempo total: 17min 38s
Dresger	Barco dragador	Natural Ambient	Sonido ambiente
Fishboat	Barco de pesca		
Motorboat	Lancha motora		
Musselboat	Barco mejillonero		
Ocean liner	Trasatlántico		
Passengers	Crucero, ferry		
Pilot ship	Barco piloto		
RORO	Mercancías		
Sail botat	Barco de vela		
Trawler	Yate		
Tugboat	Remolcadora		

Tabla 2: Clasificación binaria y tiempo total de cada clase

- Caso de estudio 2 donde hay barcos grandes o barcos pequeños según su motor, o no hay barcos.

Grandes	Tiempo total: 104min 12s	Pequeños	Tiempo total: 47min 44s	No hay embarcación	Tiempo total: 17min 38s
Dresger	Barco dragador	Fishboat	Barco de pesca	Natural Ambient	Sonido ambiente
Ocean liner	Trasatlántico	Motorboat	Lancha motora		
RORO	Mercancías	Musselboat	Barco mejillonero		
Tugboat	Remolcadora	Passengers	Crucero, ferry		
		Pilot ship	Barco piloto		
		Sail botat	Barco de vela		
		Trawler	Yate		

Tabla 3: Clasificación caso de estudio 2 y tiempo total de cada clase

En el caso del barco remolcador, aun siendo una embarcación pequeña, lo hemos considerado como grande debido a que la potencia de su motor es elevada y cumple mejor la condición de este tipo de embarcación.

- Caso de estudio 3, es un caso de estudio general en el cual se tratará de diferenciar cada una de las embarcaciones según su tipo.

4.2 ¿Qué es el espectro de potencia?

Las señales que se han obtenido mediante la grabación de audios son señales en tiempo discreto, por lo cual, no se le ha podido realizar como tal la transformada de Fourier, para obtener el comportamiento respecto de la frecuencia de dichas señales. Por ello, se ha hecho uso de la transformada de Fourier discreta, a la cual nos referiremos a ella como DFT.

La DFT es capaz de obtener un muestro de la representación en el dominio de la frecuencia de una señal que se encuentra en el dominio del tiempo y que a su vez es una señal que por definición es discreta y de duración finita. Estas características de la DFT la hacen ideal para el procesamiento de información de señales digitales.

La DFT de una función muestreada $x[n]$ cumple esta relación:

$$X[k] = T \sum_{n=0}^{N-1} x[n] e^{-\frac{j\pi kn}{N}}$$

La DFT también puede ser calculada mediante la transformada rápida de Fourier o también llamada FFT, la cual consiste en un algoritmo capaz de calcular de manera más rápida y eficiente la DFT. La FFT tiene algunas limitaciones respecto a la señal $x[n]$ debe de tener un número de muestras igual a una potencia entera de dos.

Respecto a la ecuación anterior, sabiendo que T es el periodo de muestreo de la señal original $x(t)$, su espectro de potencia vendrá dada por la siguiente relación:

$$P_k = \frac{|X[k]|^2}{NT}$$

Por lo cual podemos decir que el espectro de potencia de una señal nos da que frecuencias contienen la potencia de dicha señal.

Si vemos la figura 1 y la figura 2 tenemos un claro ejemplo de lo que sería el espectro en potencia de una señal. En este caso hemos usado como ejemplo una señal sinusoidal con ruido.

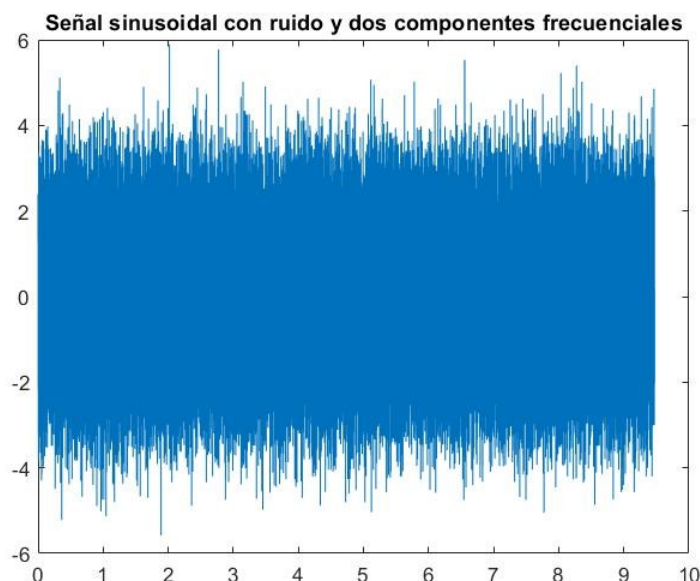


Figura 1: Señal sinusoidal con ruido y dos componentes frecuenciales

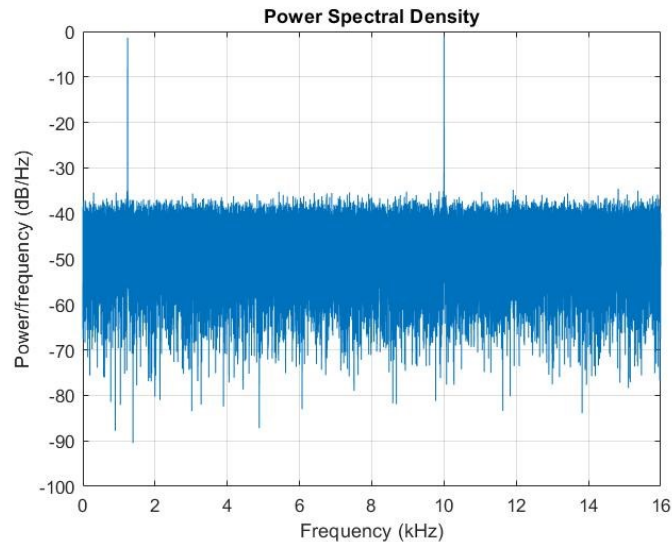


Figura 2: Espectro de potencia

Gracias al espectro de potencia se ha podido observar en la figura 2 cual es la potencia de la señal en dB. Teniendo en cuenta que se trata de una señal sinusoidal con dos componentes frecuenciales, se ha observado como aparecen dos deltas en las frecuencias de la señal sinusoidal, y el resto de potencia que se ha encontrado en las demás frecuencias es originada por el ruido. Esto nos permite obtener una mayor cantidad de información respecto a una señal que cuando hacemos uso simplemente del tiempo.

4.3 Descriptores frecuenciales

4.3.1 Coeficientes Cepstrales de las Frecuencias de Mel

Los coeficientes cepstrales en la escala de Mel, o también reconocidos como Mel Frequency Cepstral Coefficients, a los cuales se va a hacer referencia a ellos como, MFCC, fueron diseñados para representar la amplitud del espectro de potencia del habla de una manera muy precisa, por ello se usan para el reconocimiento del habla. Aun siendo para un uso específico se pretende probar si funcionan o no para este estudio.

Para obtener estos coeficientes primero aplica un filtro de pre-énfasis sobre la señal y se divide en diferentes tramas a las cuales se les aplicará una función de enventanado, la ventana usada en este caso es la ventana de Hamming. El ventaneo es necesario para suavizar los extremos de la señal. Tras realizar este proceso se procede a obtener la DFT de cada trama para así obtener la amplitud del espectro y ser pasada al dominio Mel gracias al banco de filtros. La escala Mel, es una escala lineal por debajo de 1kHz y logarítmica. Gracias a esto podemos obtener el logaritmo de la señal para que por último paso se le aplique una Transformada de Coseno Discreta (DCT), y obtener así un vector del cual se obtendrán los coeficientes por trama.

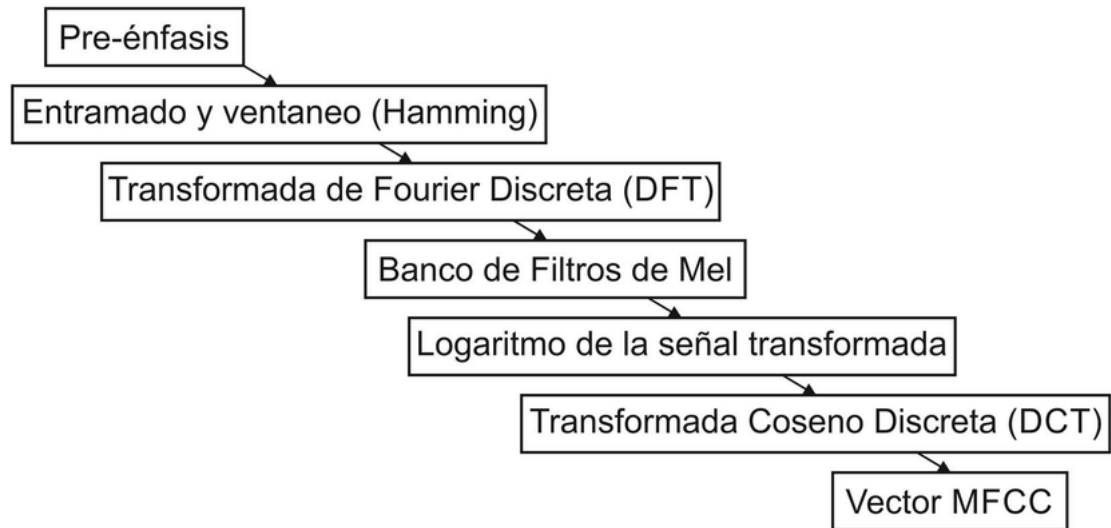


Figura 3: Esquema sobre la realización de los coeficientes MFCC

Para poder obtener los MFCC de las señales de nuestra base de datos mediante Python se ha hecho uso de librerías (Librosa.org, 2013).

Estas librerías se usan para el análisis de audios y a su vez proporcionan los componentes básicos para crear sistemas de recuperación de información musical. Con ella se pueden obtener los coeficientes de MFCC de nuestras señales de entrada gracias a la función `librosa.feature.mfcc`.

Una vez obtenidos los coeficientes se puede obtener el espectrograma de las señales, con relación a este tipo de coeficientes.

Se ha representado un espectrograma de cada uno de los tipos de embarcaciones que se encuentran en la base de datos.

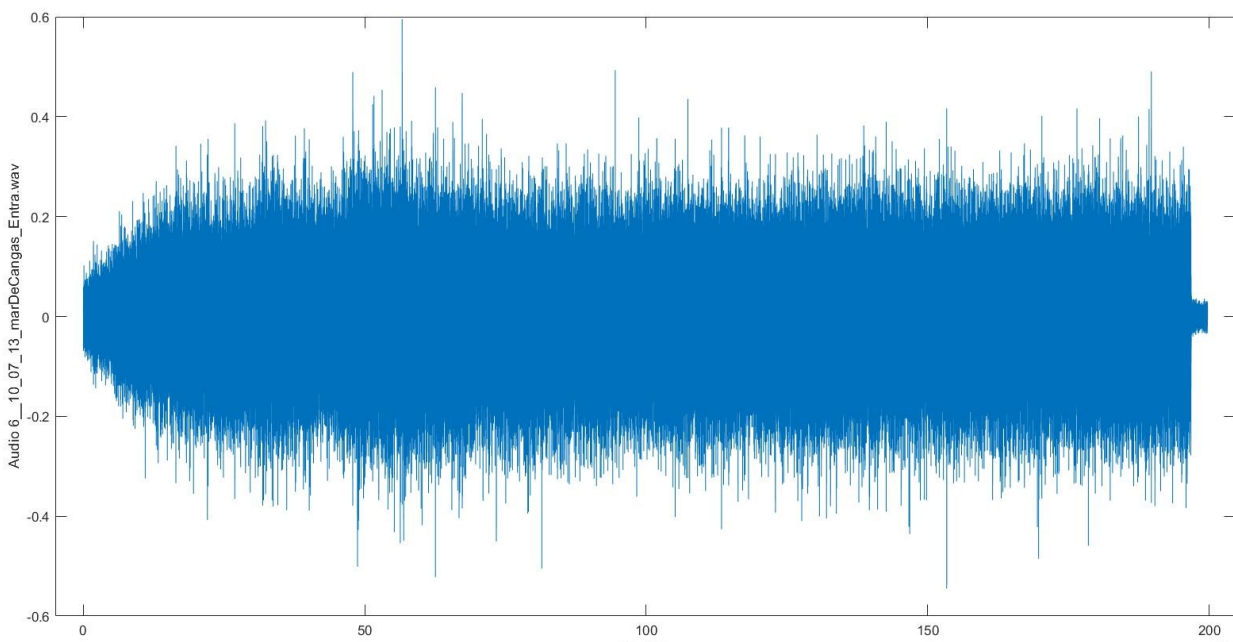


Figura 4: Audio 6__10_07_13_marDeCangas_Entra. Tipo de embarcación: Passenger (Pasajeros)

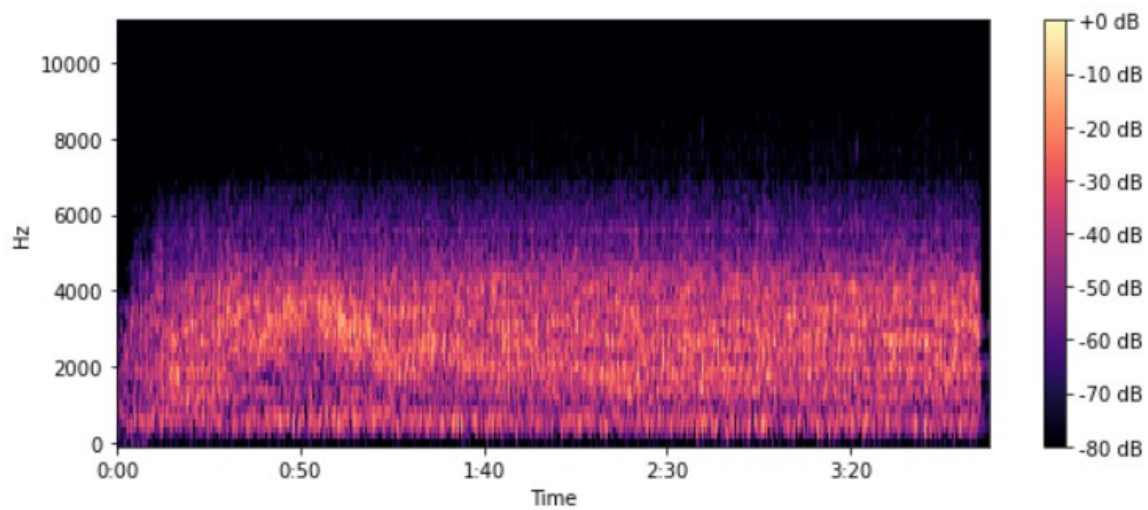


Figura 5: Espectrograma de coeficiente MFCC correspondiente al archivo: Audio 6__10_07_13_marDeCangas_Entra

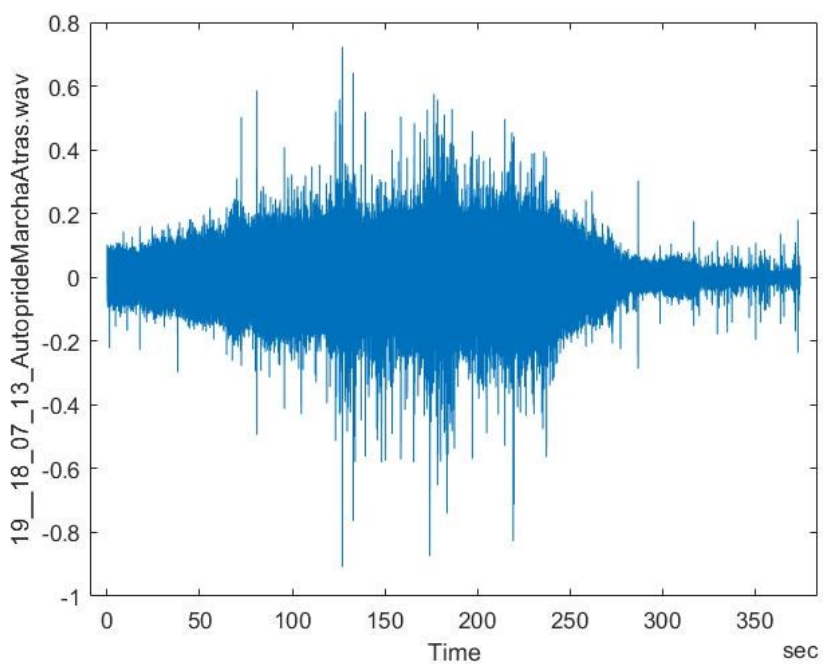


Figura 6: 19__18_07_13_AutoprideMarchaAtras. Tipo de embarcación: RORO (Mercancías)

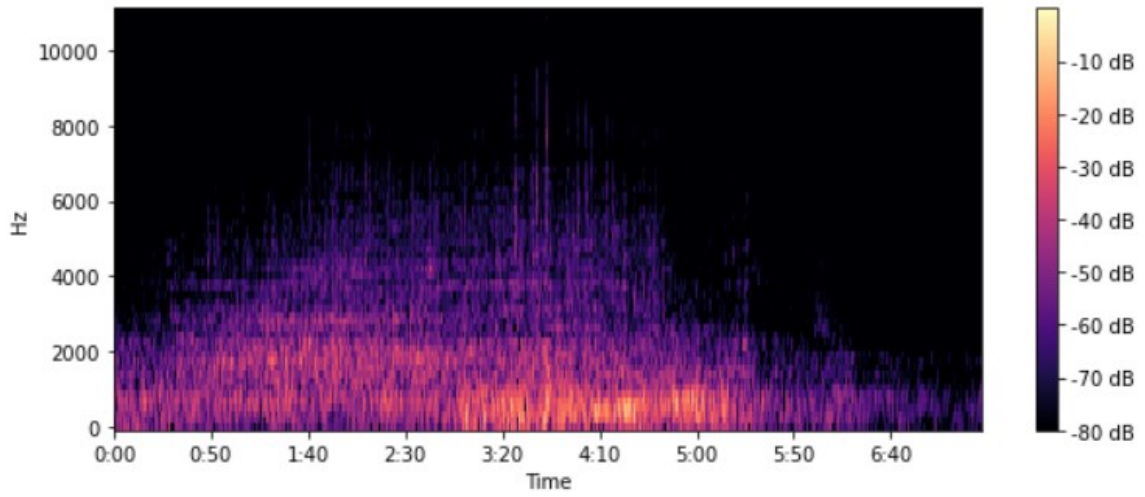


Figura 7: Espectrograma de coeficiente MFCC correspondiente al archivo: 19__18_07_13_AutoprideMarchaAtras

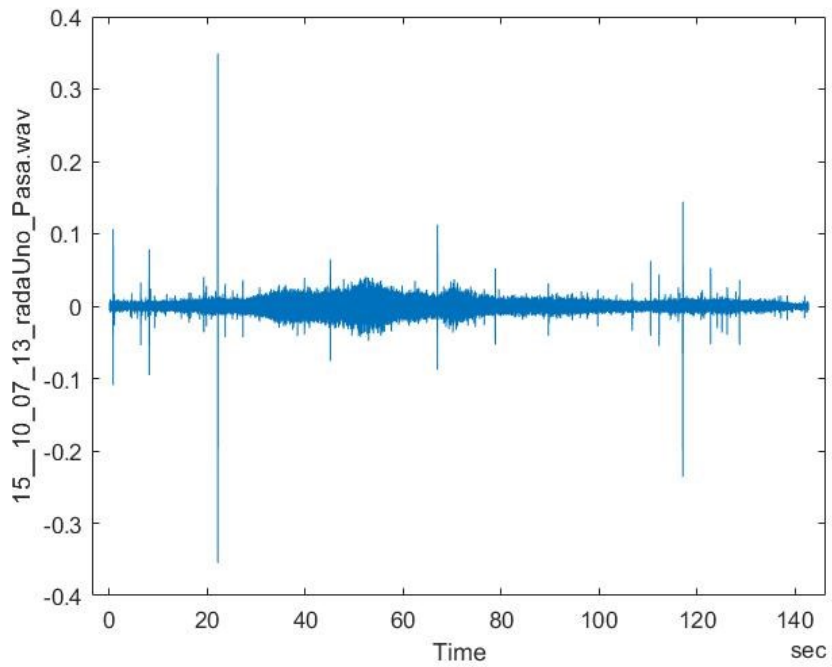


Figura 8: 15__10_07_13_radaUno_Pasa. Tipo de embarcación: Tuboat (Remolcador)

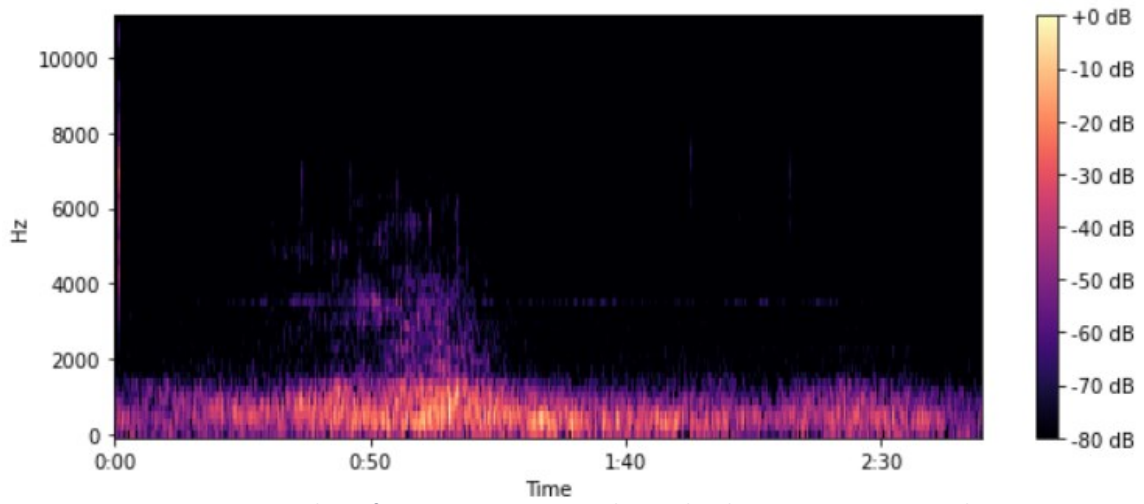


Figura 9: Espectrograma de coeficiente MFCC correspondeinte al archivo: 15__10_07_13_radaUno_Pasa

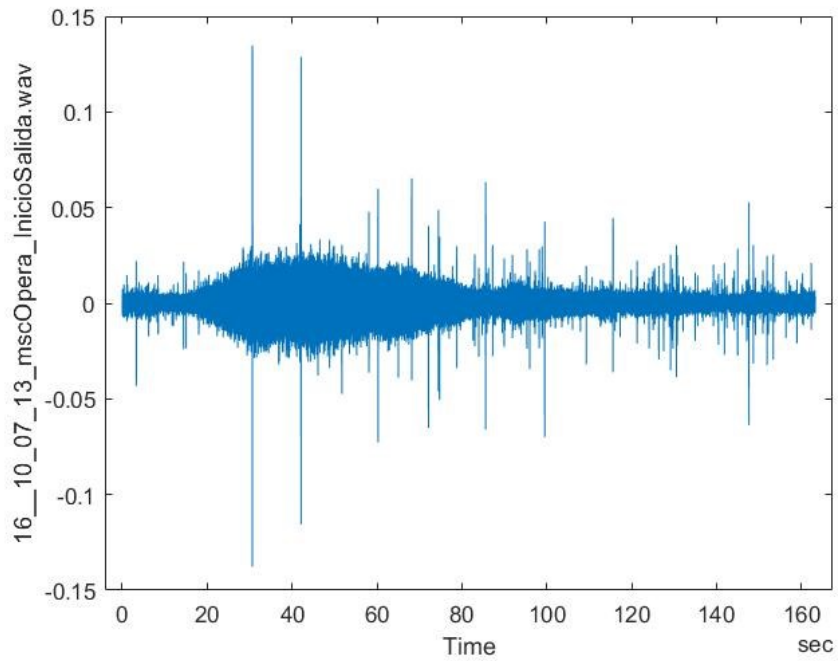


Figura 10: 16__10_07_13_mscOpera_InicioSalida. Tipo de embarcación: Ocean Liner (Trasatlántico)

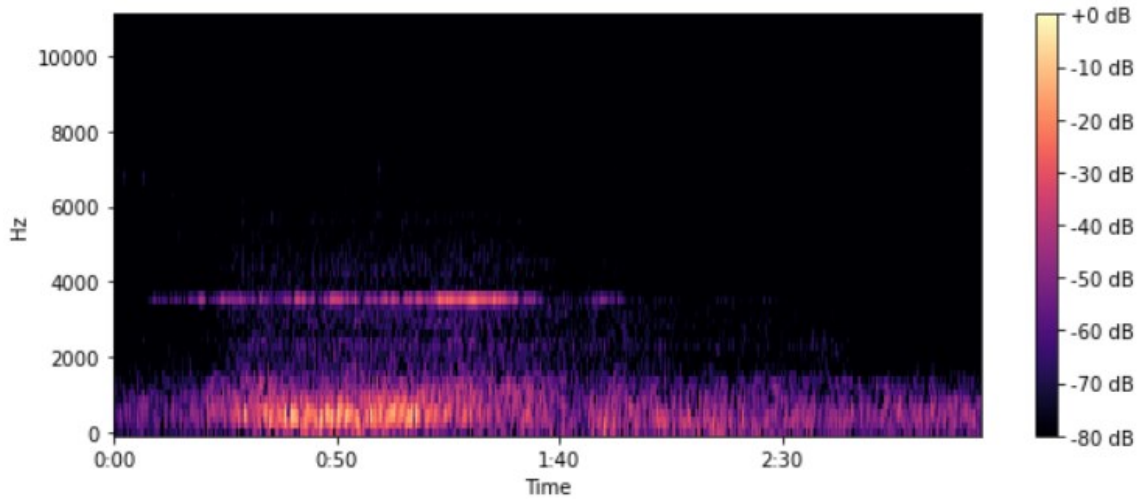


Figura 11: Espectrograma de coeficiente MFCC correspondiente al archivo: 16__10_07_13_mscOpera_InicioSalida

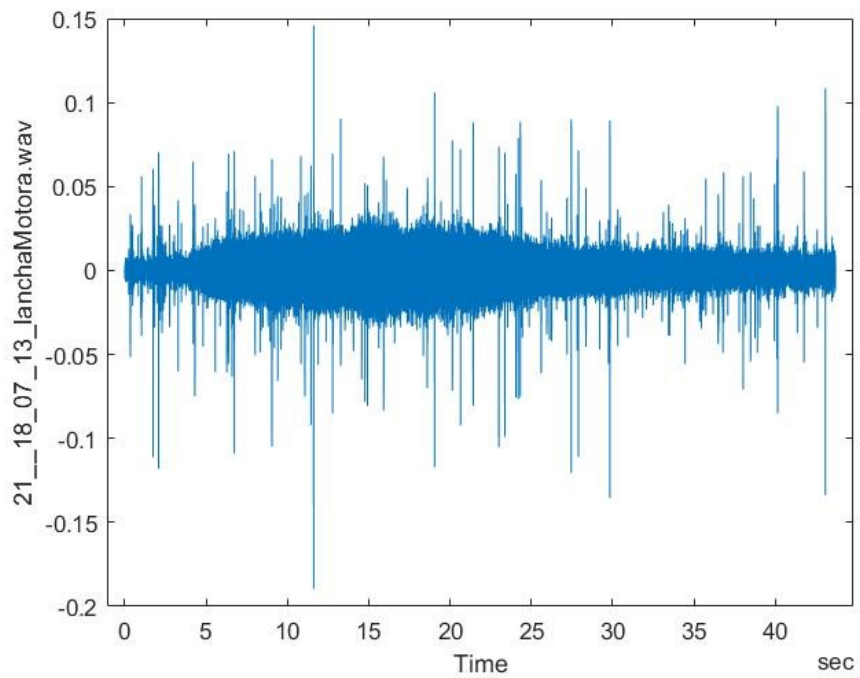


Figura 12: 21__18_07_13_lanchaMotora.Tipo de embarcación: Motorboat (Lancha motora)

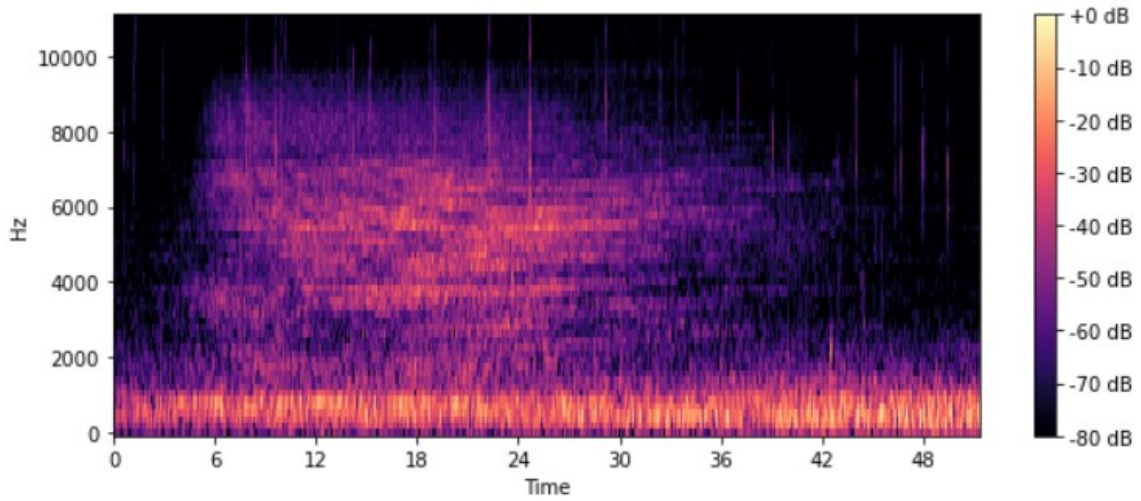


Figura 13: Espectrograma de coeficiente MFCC correspondeinte al archivo: 21_18_07_13_lanchaMotora

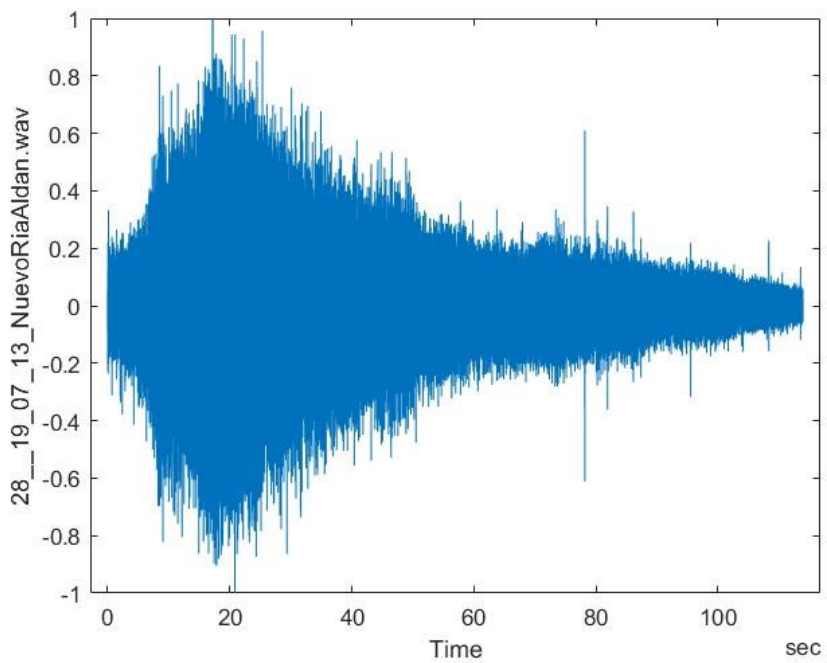


Figura 14: 28_19_07_13_NuevoRiaAldan.Tipo de embarcación: Trawler (Yate)

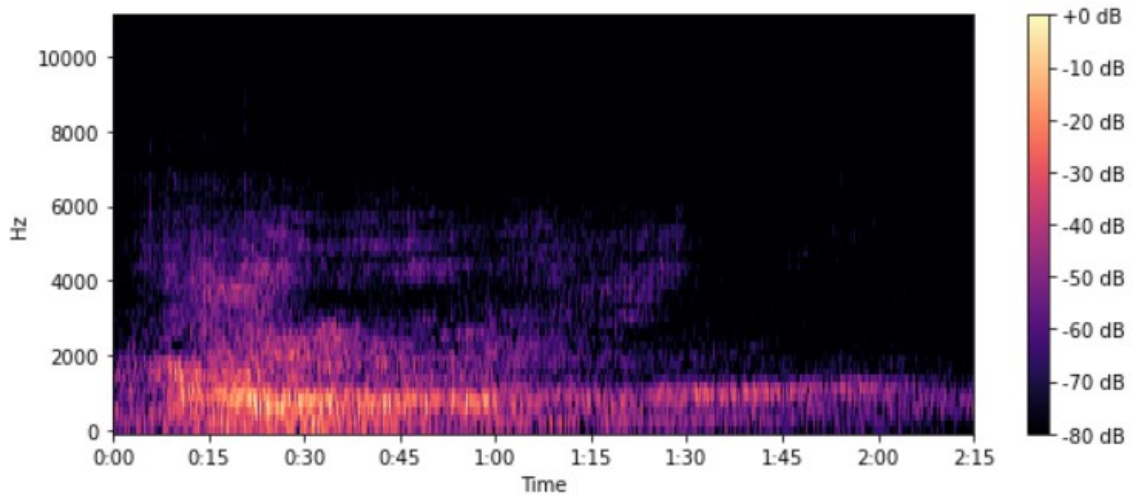


Figura 15: Espectrograma de coeficiente MFCC correspondeinte al archivo: 28__19_07_13_NuevoRiaAldan

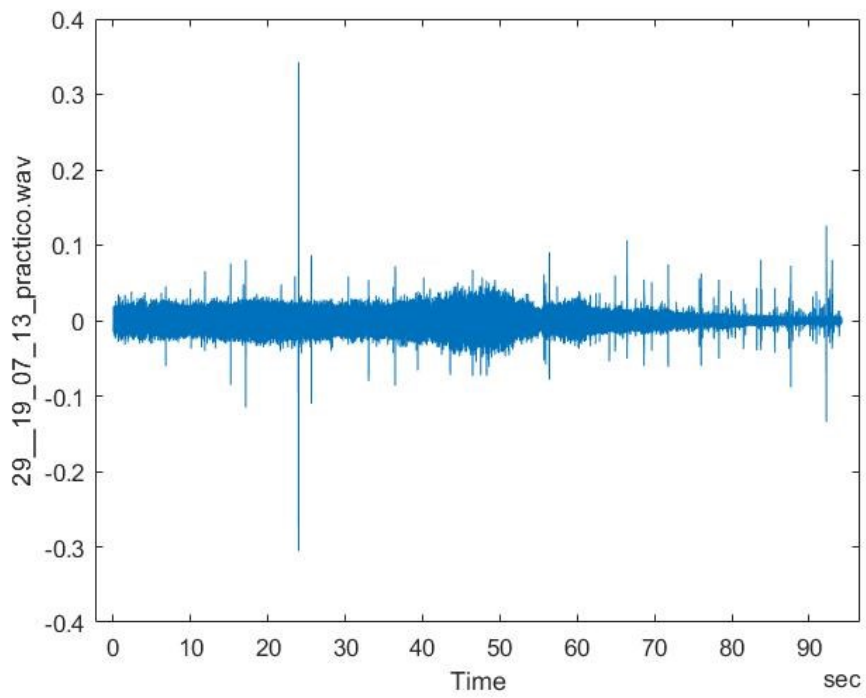


Figura 16: 29__19_07_13_practico. Tipo de embarcación: Pilot ship (Barco piloto)

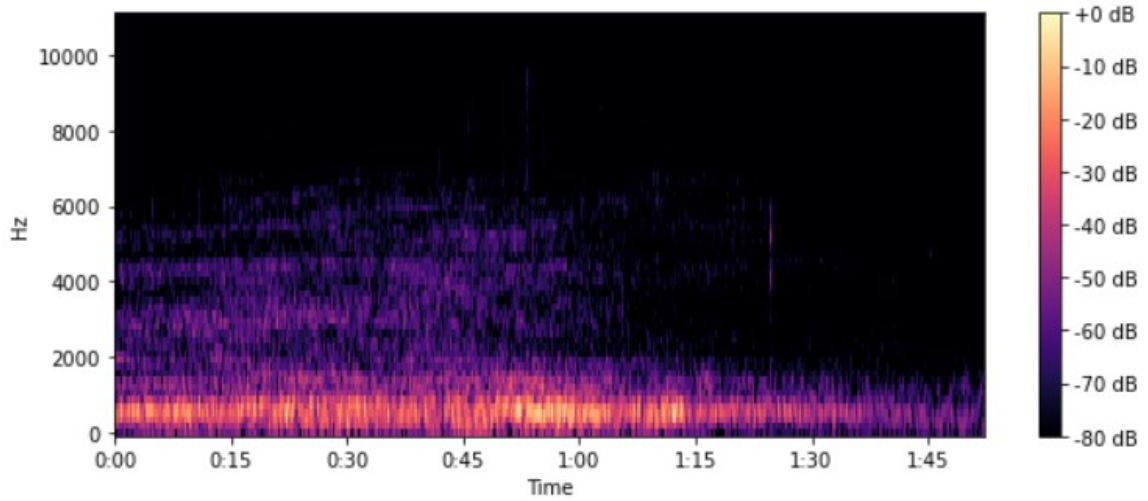


Figura 17: Espectrograma de coeficiente MFCC correspondiente al archivo: 29_19_07_13_practico

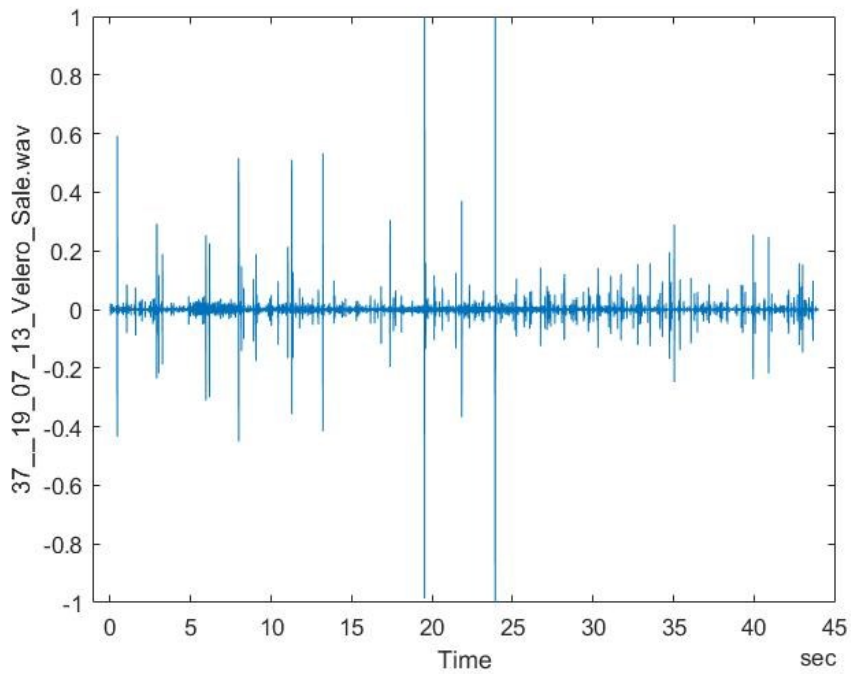


Figura 18: 37_19_07_13_Velero_Sale. Tipo de embarcación: Sailboat (Barco de vela)

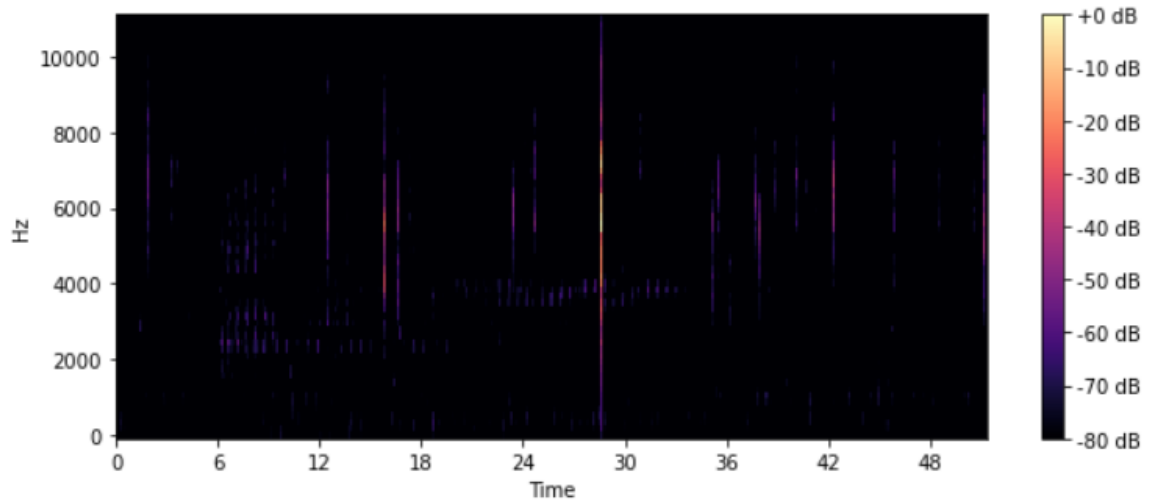


Figura 19: Espectrograma de coeficiente MFCC correspondiente al archivo: 37__19_07_13_Velero_Sale

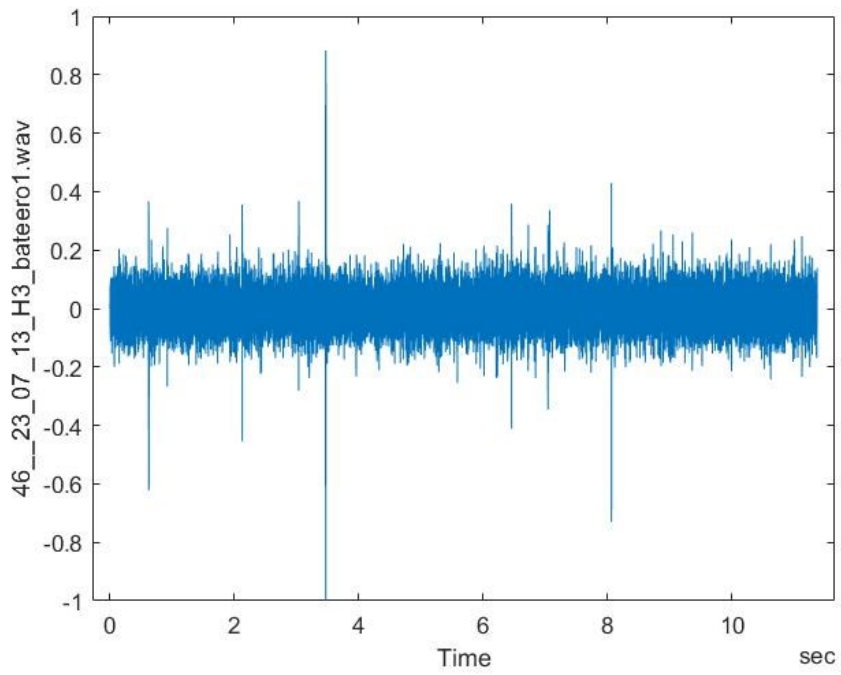


Figura 20: 46__23_07_13_H3_bateero1. Tipo de embarcación: Mussel boat (Barco mejillonero)

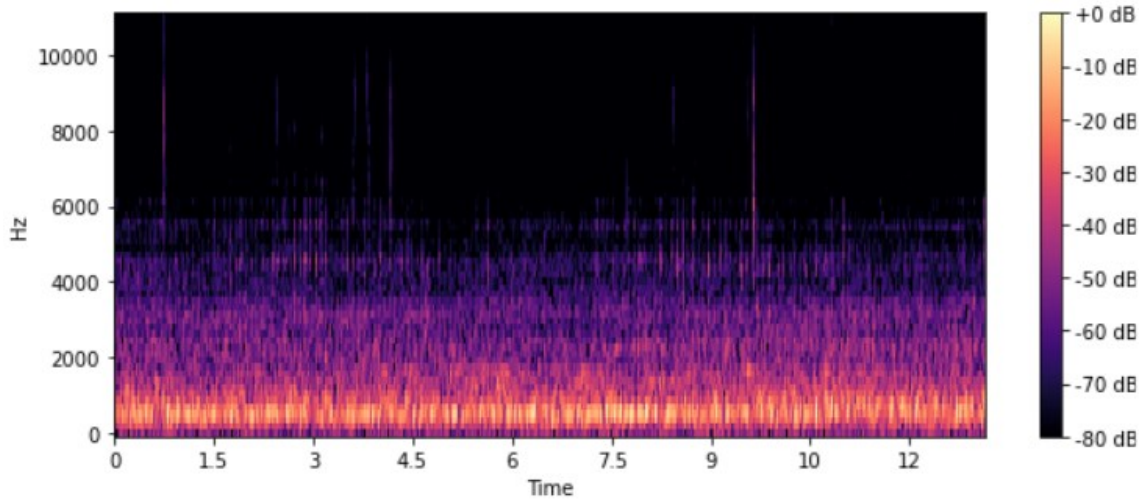


Figura 21: Espectrograma de coeficiente MFCC correspondeinte al archivo: 46__23_07_13_H3_bateero1

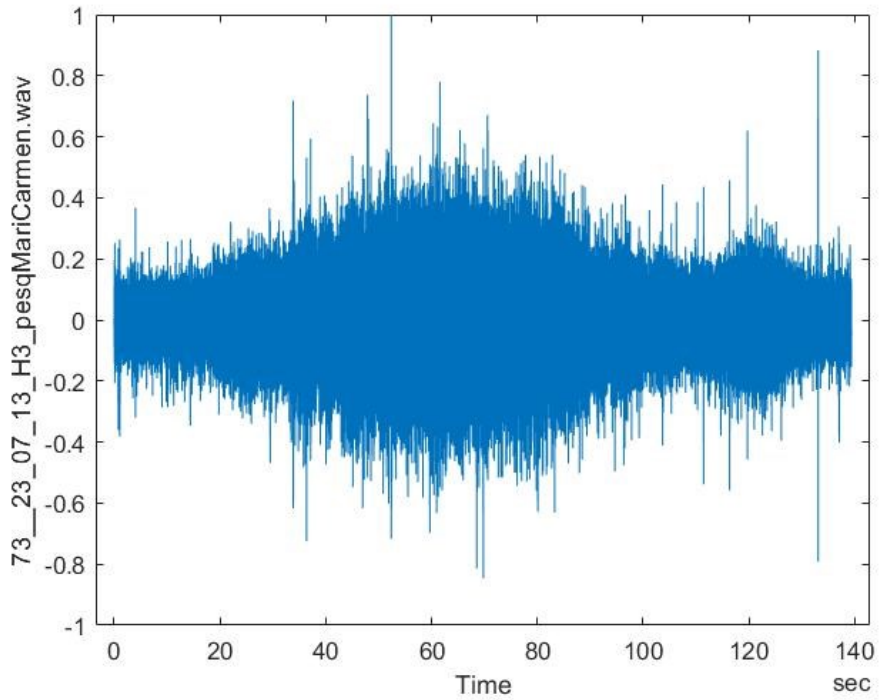


Figura 22: 73__23_07_13_H3_pesqMariCarmen. Tipo de embarcación: Fishboat (Barco de pesca)

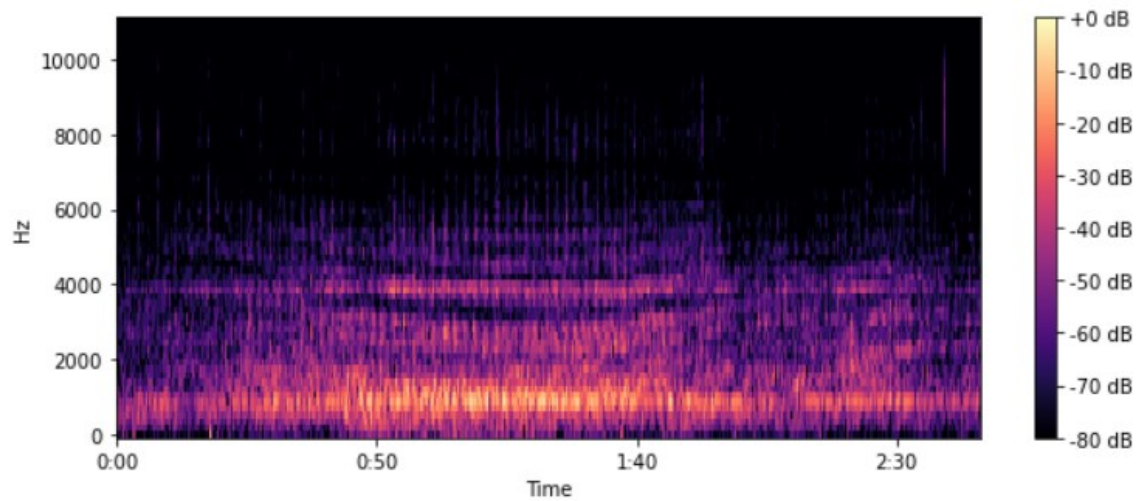


Figura 23: Espectrograma de coeficiente MFCC correspondeinte al archivo: 73__23_07_13_H3_pesqMariCarmen

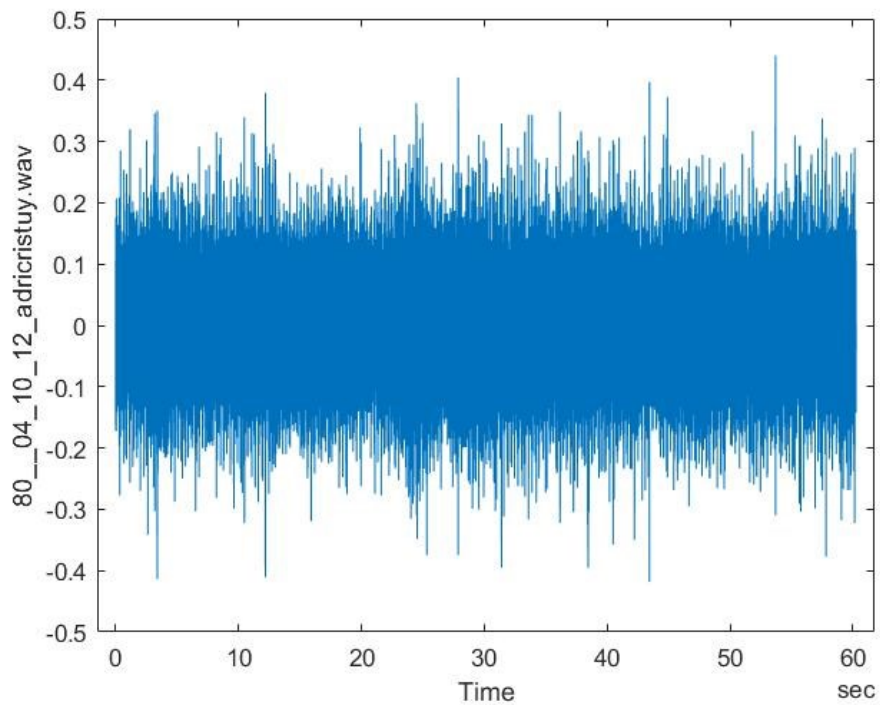


Figura 24: 80__04_10_12_adricristuy. Tipo de embarcación: Dreger (Barco dregador)

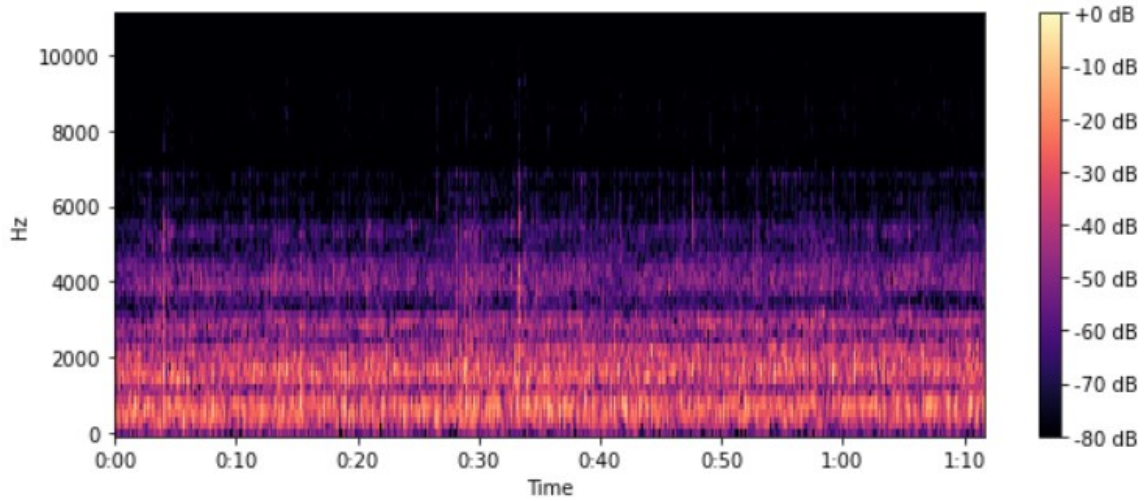


Figura 25: Espectrograma de coeficiente MFCC correspondiente al archivo: 80__04_10_12_adricristuy

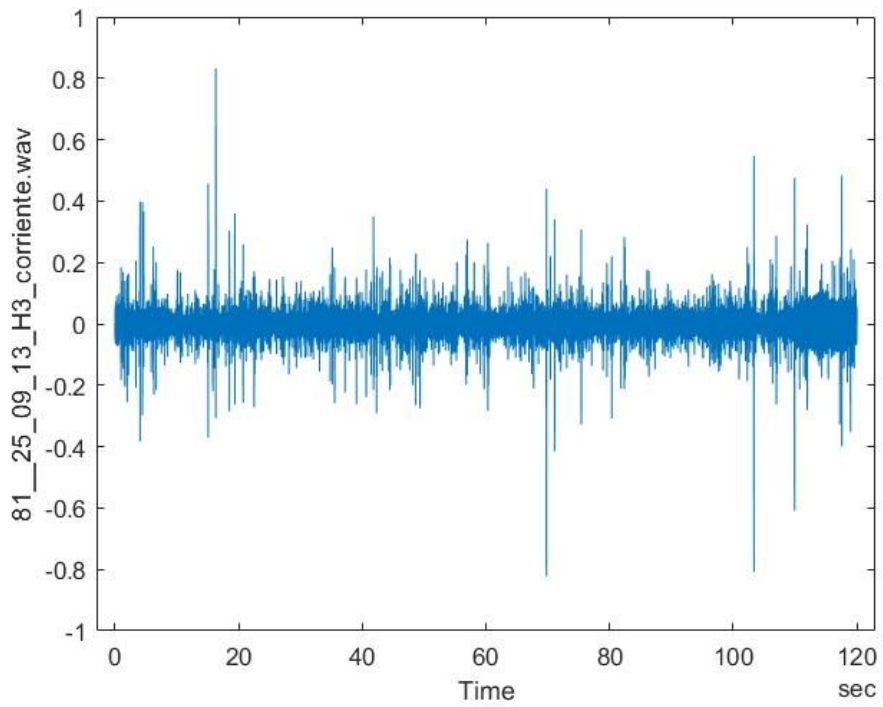


Figura 26: 81_25_09_13_H3_corriente. Tipo de embarcación: Natural Ambient (Sonido ambiental)

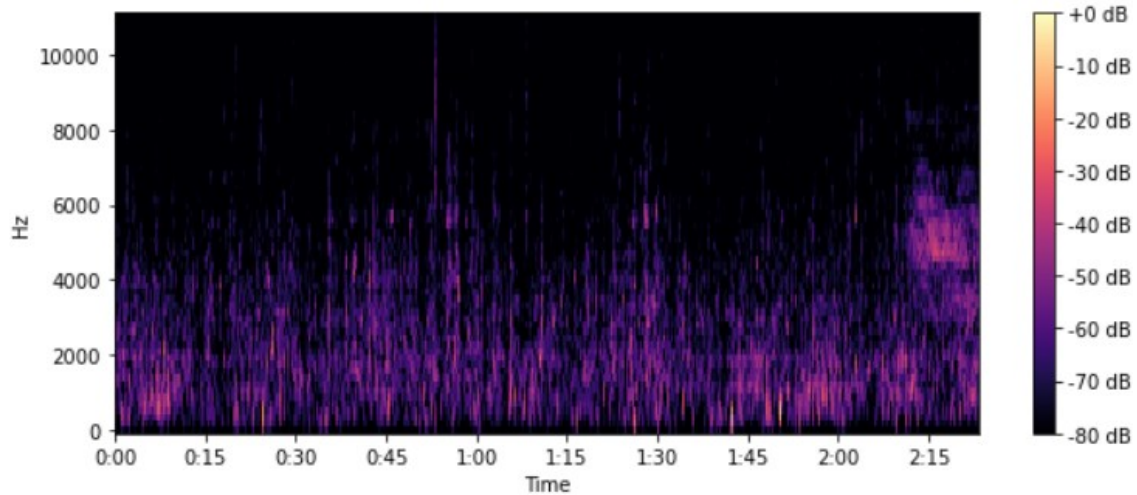


Figura 27: Espectrograma de coeficiente MFCC correspondiente al archivo: 81__25_09_13_H3_corriente.

donde el eje X muestra el intervalo de tiempo del audio, el eje Y a la derecha muestra las frecuencias que se obtienen y el color representa la intensidad de cada una de las frecuencias.

Como se puede observar en las gráficas, gracias el uso de los coeficientes MFCC podemos ver la intensidad que tiene la señal en los diferentes instantes de tiempo y a su vez en las diferentes frecuencias que se encuentran en ese instante.

4.3.2 Coeficientes cepstrales gammatones.

Para comenzar a hablar sobre los coeficientes cepstrales gammatones o también llamados gammatone cepstral coefficients, a los cuales se va a hacer referencia a ellos como GTCC, se ha de explicar primero el funcionamiento de los filtros gammatone, a los cuales nos referiremos a ellos como filtros GT.

El filtro GT es el producto de una función de distribución Gamma y un tono sinusoidal centrado en la frecuencia f_c y viene dado por la siguiente relación:

$$g(t) = K t^{(n-1)} e^{-2\pi Bt} \cos(2\pi f_c t + \varphi) \quad t > 0$$

donde B da lugar al ancho de banda del filtro, n es el orden del filtro y a su vez también nos determina la calidad de este. En comparación con el filtro Mel, la función de transferencia del filtro GT es menos abrupta respecto a la triangular del filtro Mel.

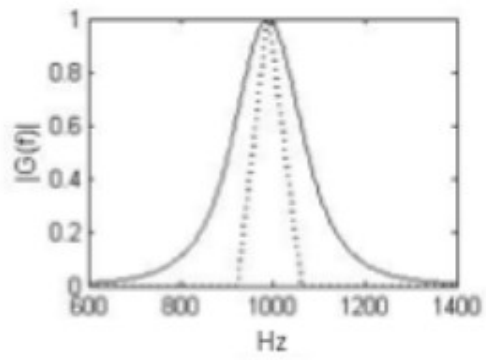


Figura 28: Respuesta en frecuencia de un filtro GT y un filtro Mel

La figura 4 representa la respuesta en frecuencia de un filtro GT (línea continua) con respecto a un filtro Mel (línea discontinua) a una misma frecuencia de 1kHz.

Ahora se procede explicar la obtención de los descriptores GTCC. Primero se realiza un ventaneo de la señal acústica y se realiza el cálculo de la DFT. Con esto obtenemos el espectro de la señal obtenida que es suavizado mediante el banco de filtros GT. Y para finalizar al igual que en la obtención de los MFCC, se procede al cálculo del logaritmo y la transformada discreta del coseno (DCT). El cálculo del logaritmo modela la sonoridad de la señal, mientras que la transformada del coseno decorrela las salidas logarítmicas del banco de filtro, haciendo que haya una mejor uniformidad de la energía.



Figura 29

Los descriptores GTCC se pueden calcular como:

$$GTCC_m = \sqrt{\frac{2}{F} \sum_{n=1}^N \log(X_n) \cos \left[\frac{\pi n}{F} \left(m - \frac{1}{2} \right) \right]} \quad 1 \leq m \leq M$$

donde X_n es la señal en el dominio espectral, F número de bandas del banco de filtros Gammatone y M es el número de GTCC.

Para poder obtener los coeficientes GTCC de nuestra base de datos mediante Python, se ha hecho uso de librerías (Spafe, 2019).

La librería Spafe tiene el objetivo de hacer más sencillo la extracción de características de los archivos de audio. En ella se incluyen varios cálculos relacionados con bancos de filtros, espectrogramas, frecuencias y características cepstrales.

Gracias a ella se ha podido usar la función `gfcc` que es la calcula los descriptores GFCC.

Una vez obtenidos los coeficientes se puede obtener el espectrograma de las señales, en relación a este tipo de coeficientes.

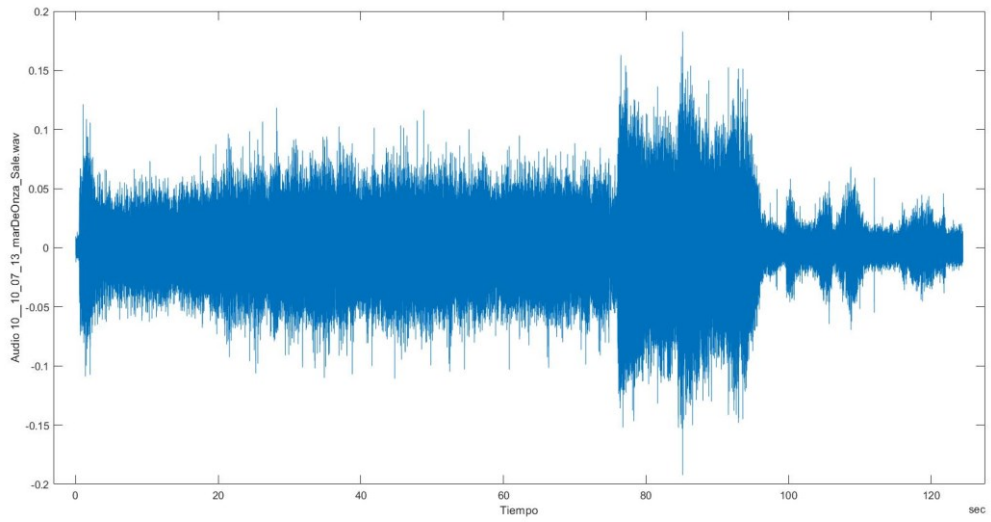


Figura 30: 10__10_07_13_marDeOnza_Sale. Tipo de embarcación: Passengers (Pasajeros)

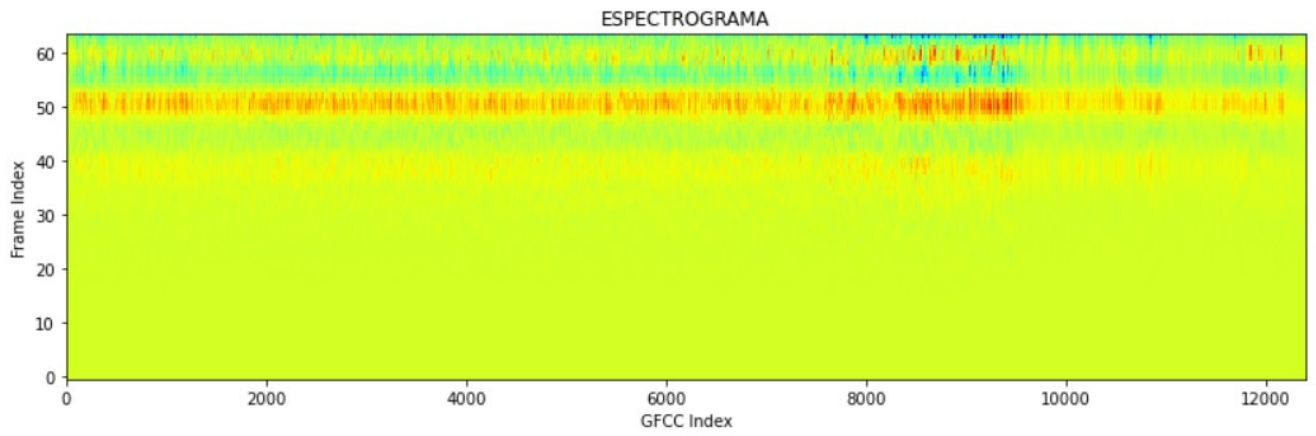


Figura 31: Espectrograma de coeficiente GTCC correspondiente al archivo: 10__10_07_13_marDeOnza_Sale

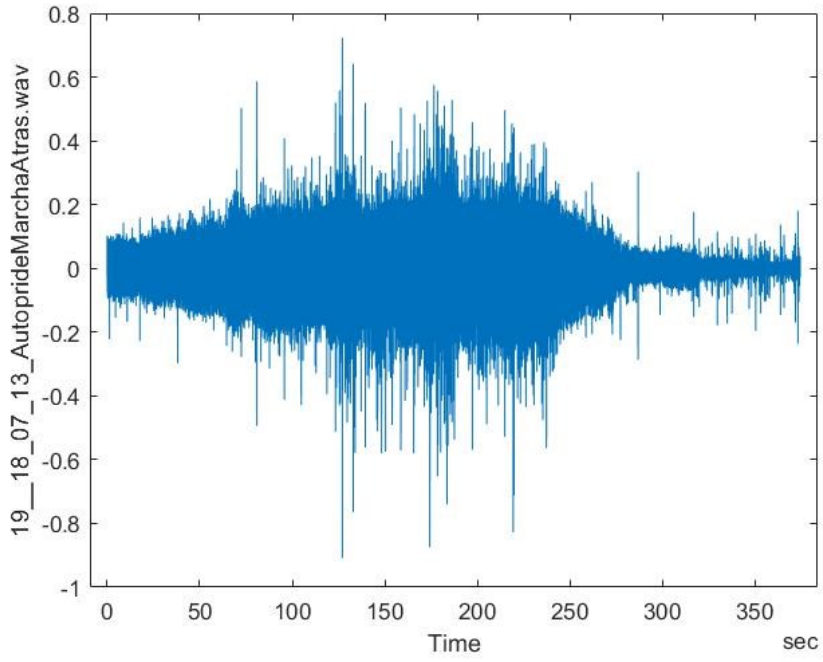


Figura 32: 19__18_07_13_AutoprideMarchaAtras. Tipo de embarcación: RORO (Mercancías)

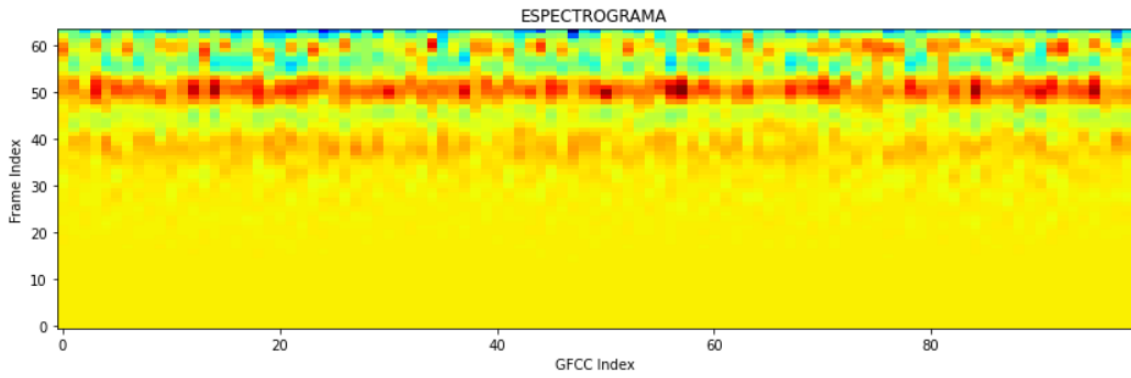


Figura 33: Espectrograma de coeficiente GTCC correspondiente al archivo: 19__18_07_13_AutoprideMarchaAtras

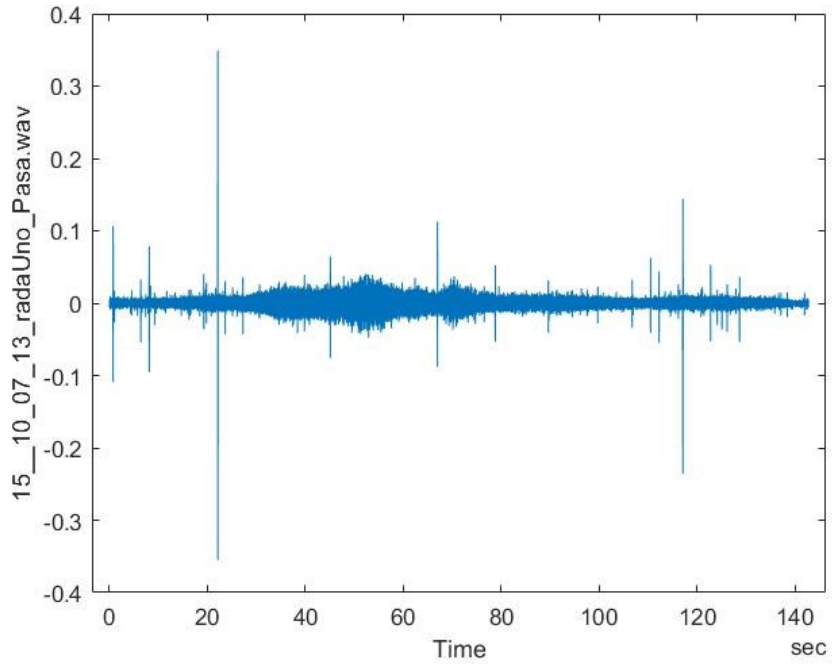


Figura 34: 15_10_07_13_radaUno_Pasa. Tipo de embarcación: Tuboat (Remolcador)

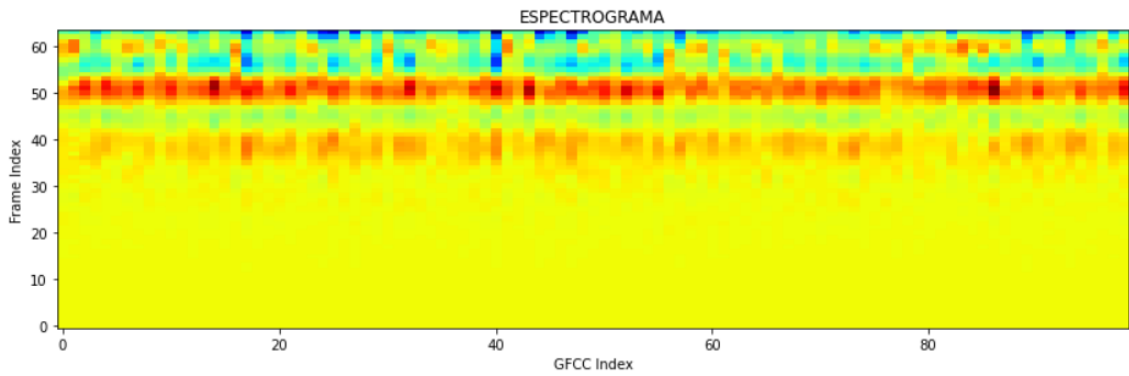


Figura 35: Espectrograma de coeficiente GTCC correspondiente al archivo: 15_10_07_13_radaUno_Pasa

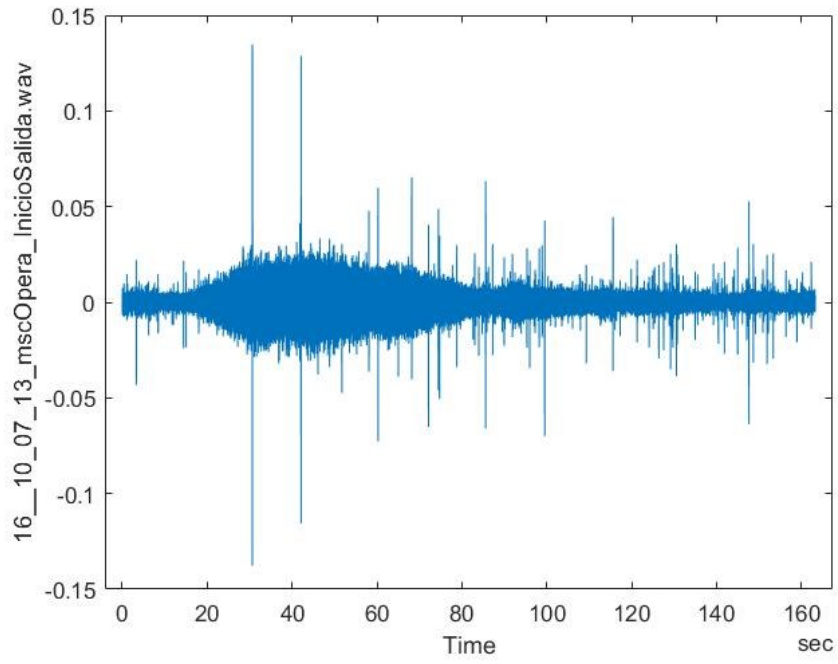


Figura 36: 16__10_07_13_mscOpera_InicioSalida. Tipo de embarcación: Ocean Liner (Trasatlántico)

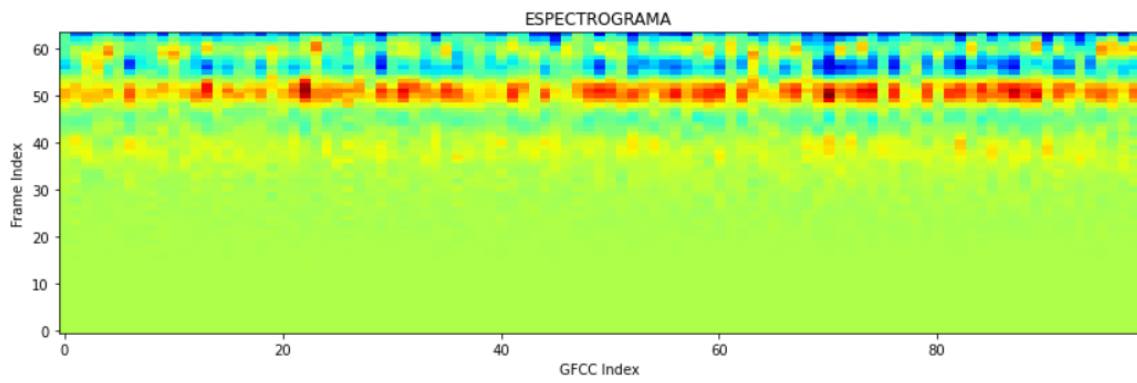


Figura 37: Espectrograma de coeficiente GTCC correspondeinte al archivo: 16__10_07_13_mscOpera_InicioSalida

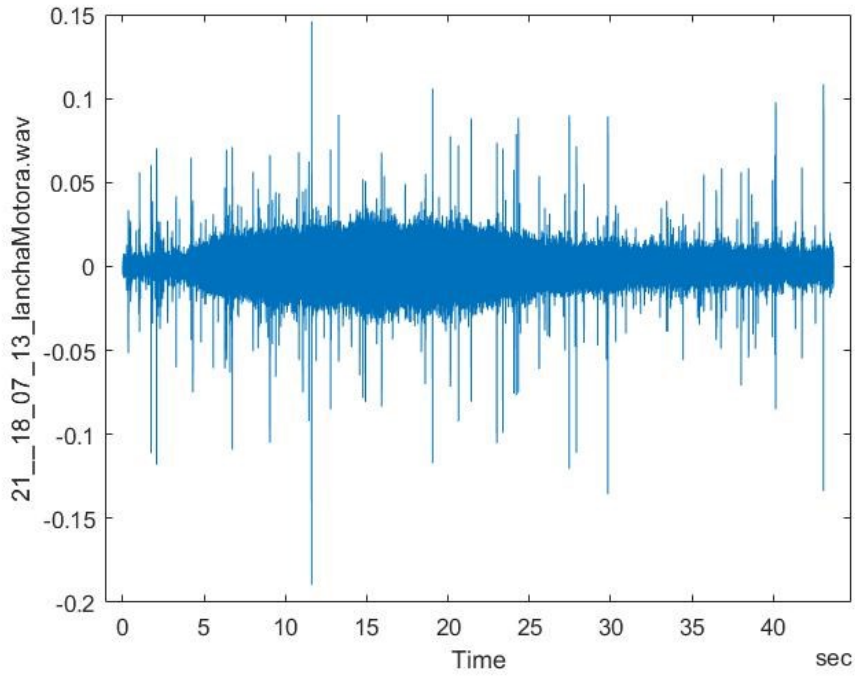


Figura 38: 21__18_07_13_lanchaMotora.Tipo de embarcación: Motorboat (Lancha motora)

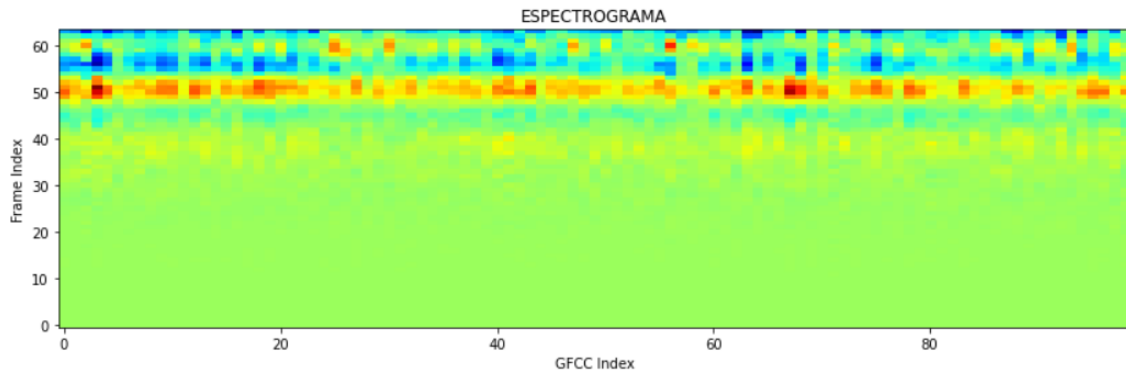


Figura 39: Espectrograma de coeficiente GTCC correspondiente al archivo: 21__18_07_13_lanchaMotora

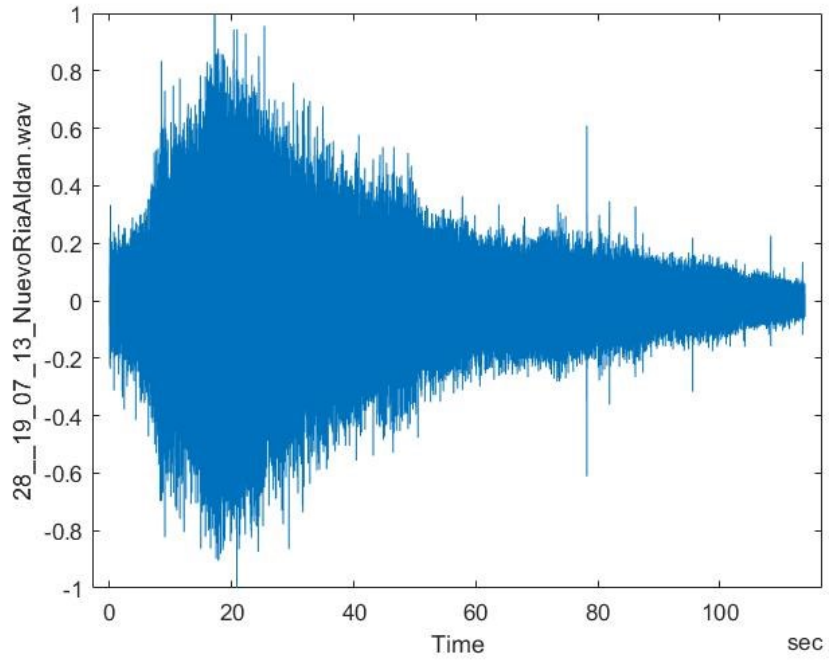


Figura 40: 28_19_07_13_NuevoRiaAldan.Tipo de embarcación: Trawler (Yate)

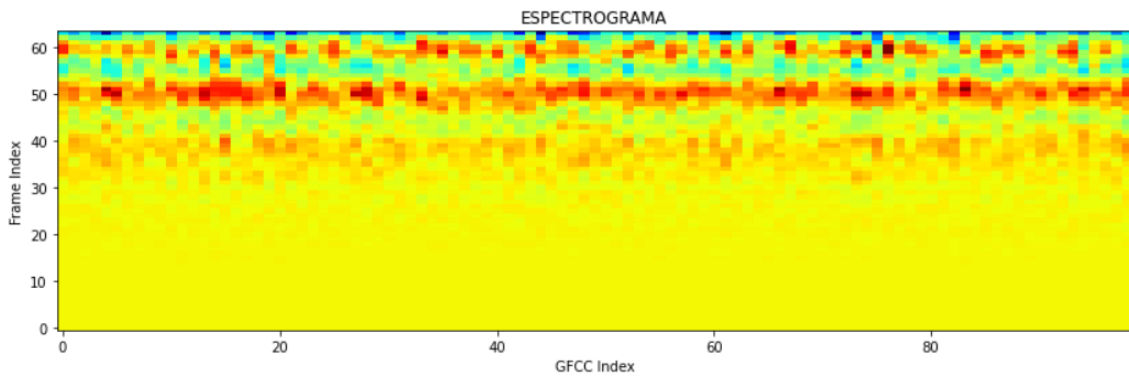


Figura 41: Espectrograma de coeficiente GTCC correspondiente al archivo: 28_19_07_13_NuevoRiaAldan

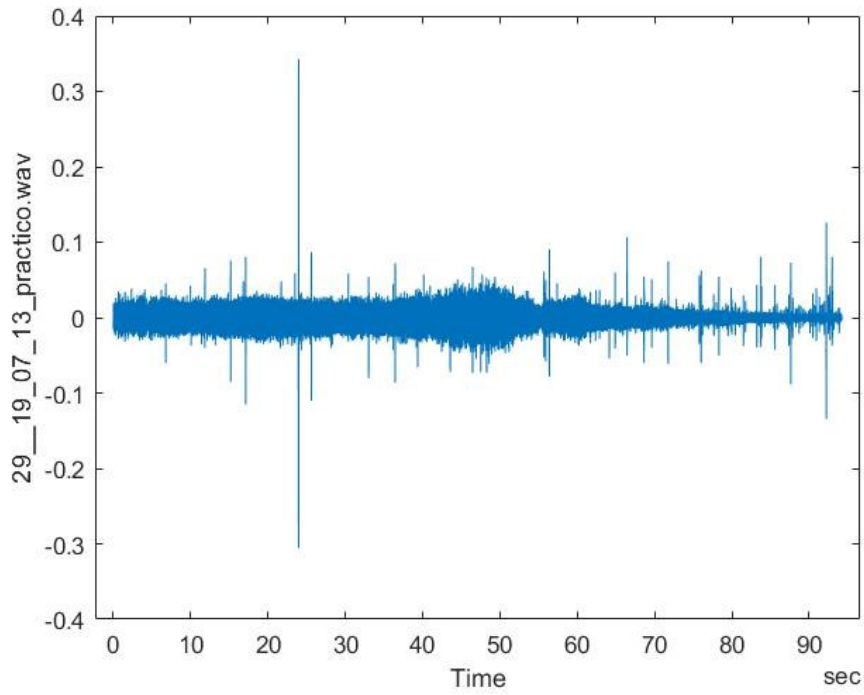


Figura 42: 29_19_07_13_practico. Tipo de embarcación: Pilot ship (Barco piloto)

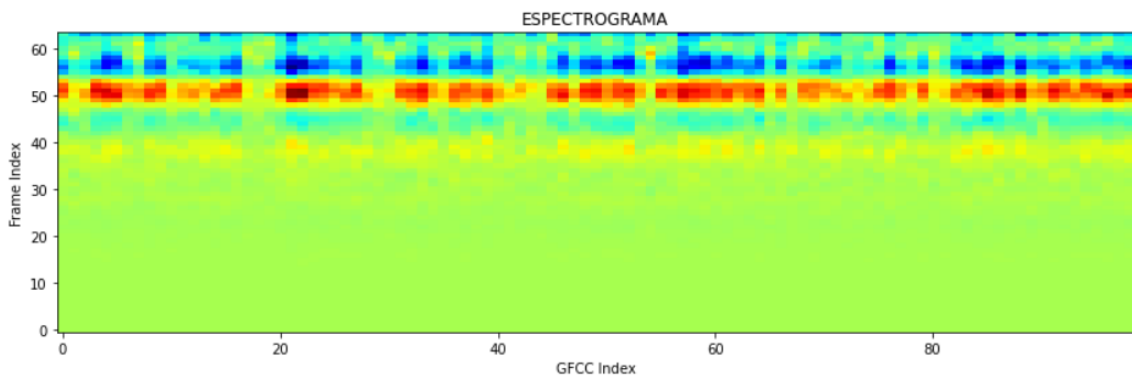


Figura 43: Espectrograma de coeficiente GTCC correspondiente al archivo: 29_19_07_13_practico

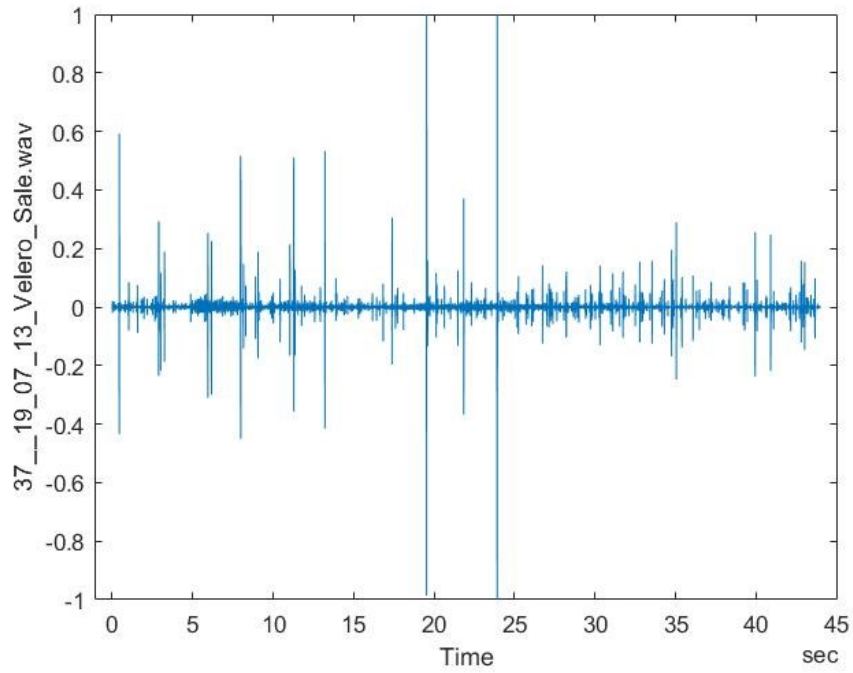


Figura 44: 37_19_07_13_Velero_Sale. Tipo de embarcación: Sailboat (Barco de vela)

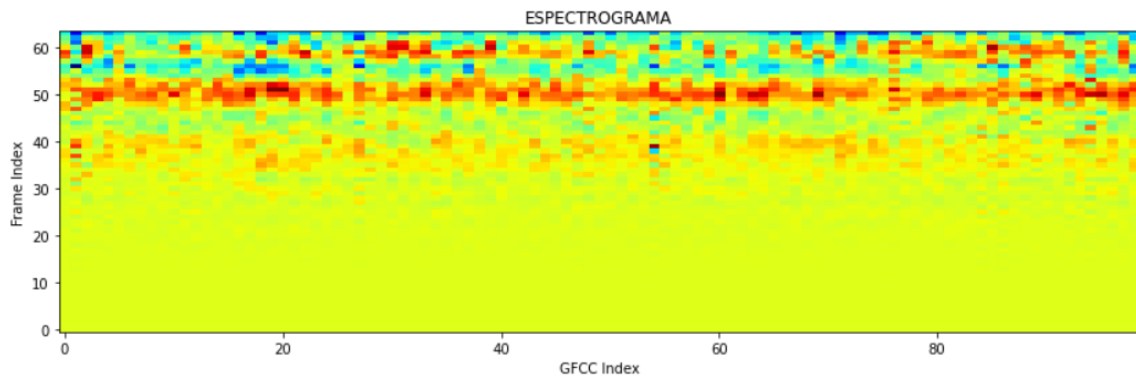


Figura 45: Espectrograma de coeficiente GTCC correspondeinte al archivo: 37_19_07_13_Velero_Sale

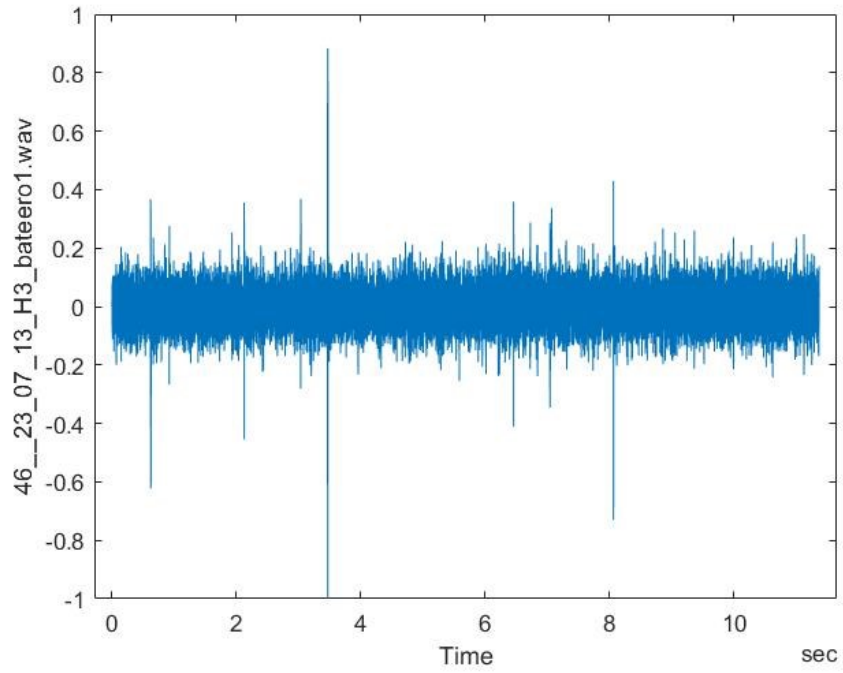


Figura 46: 46__23_07_13_H3_bateero1. Tipo de embarcación: Mussel boat (Barco mejillonero)

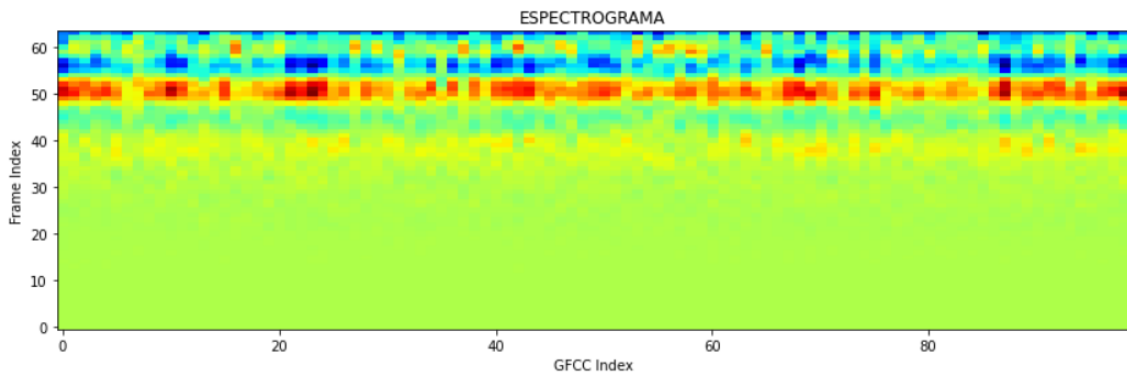


Figura 47: Espectrograma de coeficiente GTCC correspondeinte al archivo: 46__23_07_13_H3_bateero1

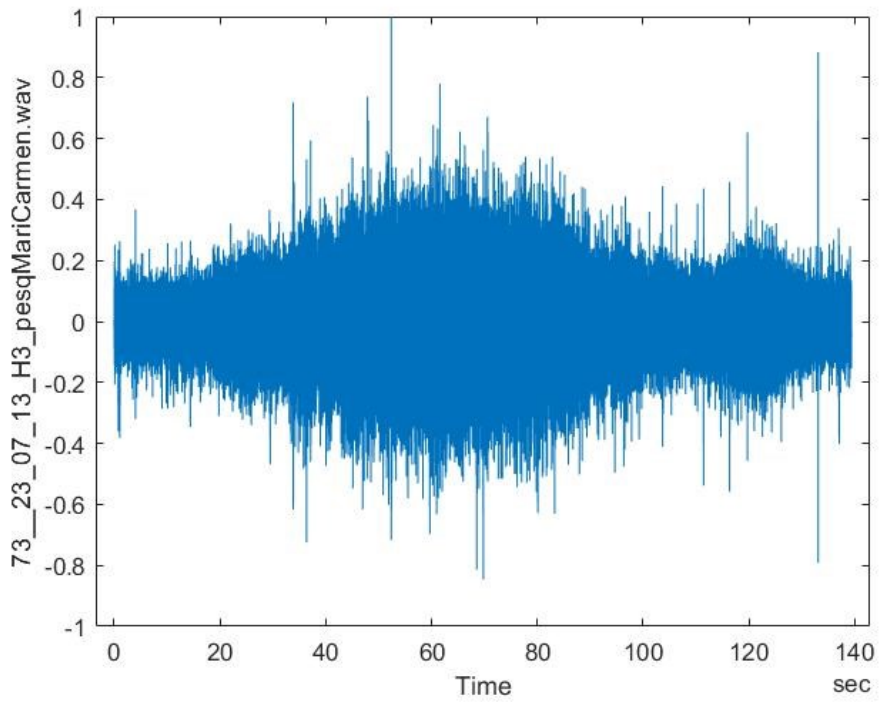


Figura 48: 73_23_07_13_H3_pesqMariCarmen. Tipo de embarcación: Fishboat (Barco de pesca)

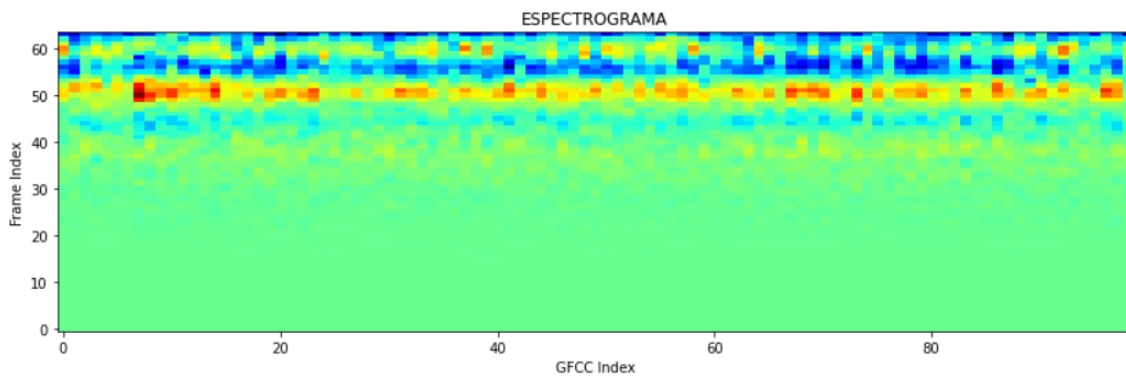


Figura 49: Espectrograma de coeficiente GTCC correspondeinte al archivo: 73_23_07_13_H3_pesqMariCarmen

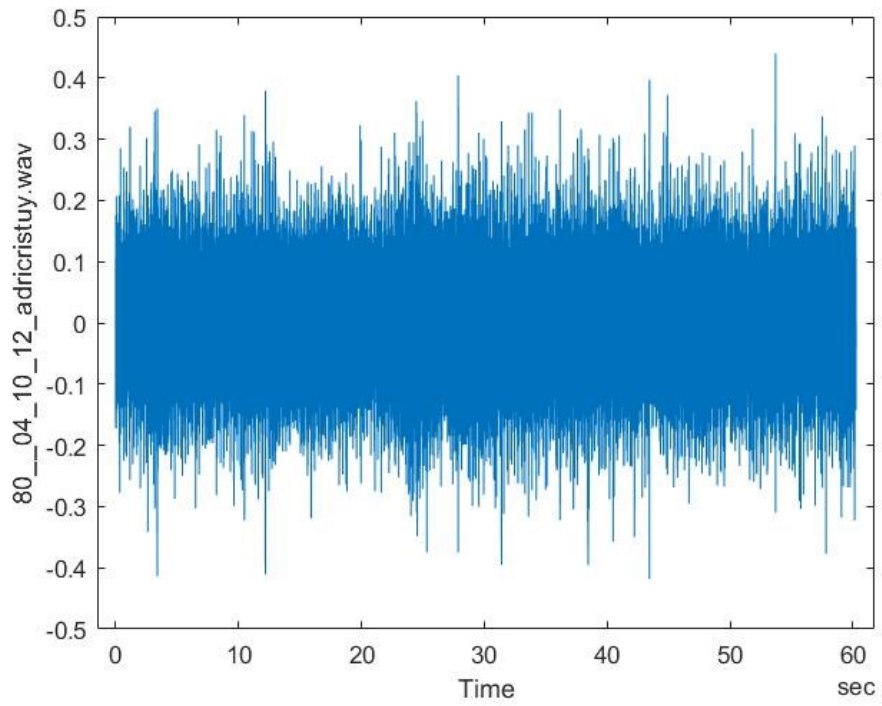


Figura 50: 80_04_10_12_adricristuy. Tipo de embarcación: Dregar (Barco dregador)

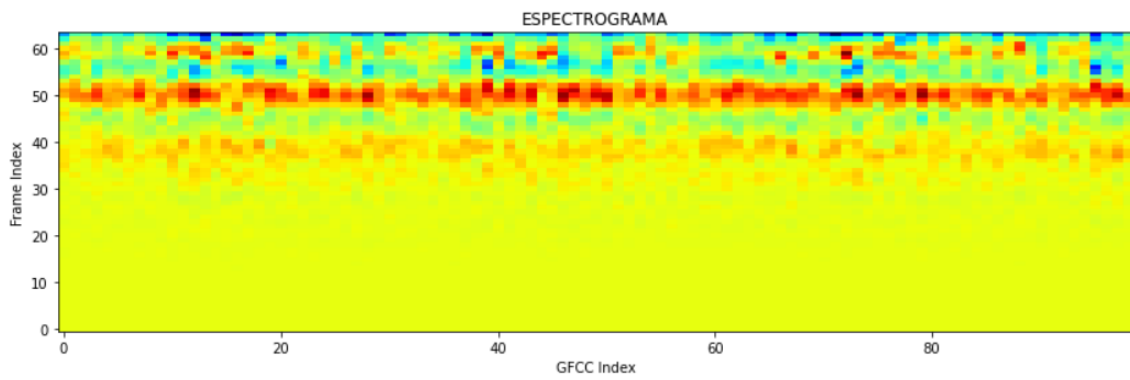


Figura 51: Espectrograma de coeficiente GTCC correspondiente al archivo: 80_04_10_12_adricristuy

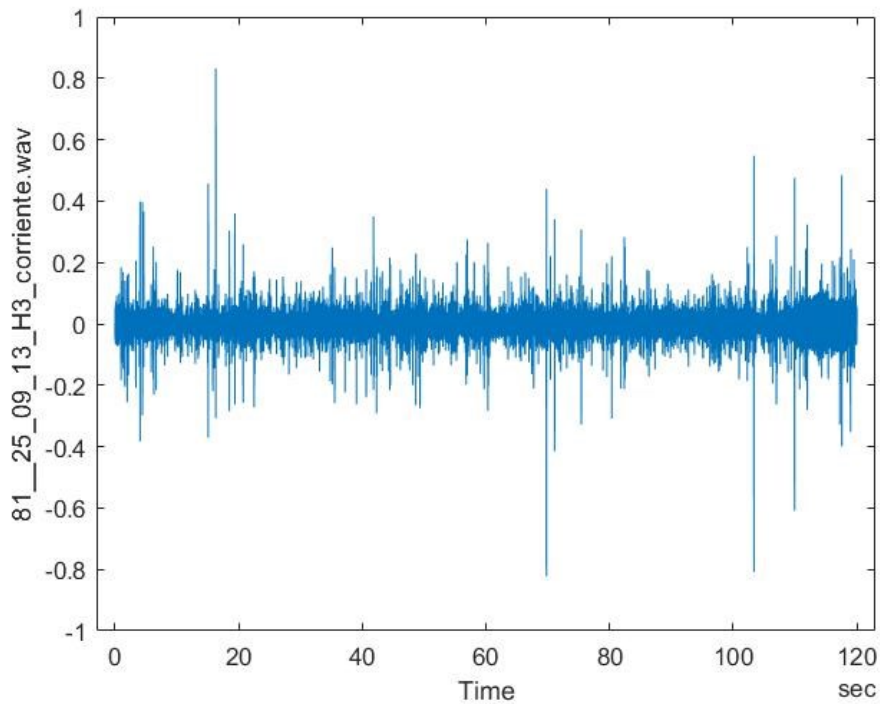


Figura 52: 81_25_09_13_H3_corriente. Tipo de embarcación: Natural Ambient (Sonido ambiental)

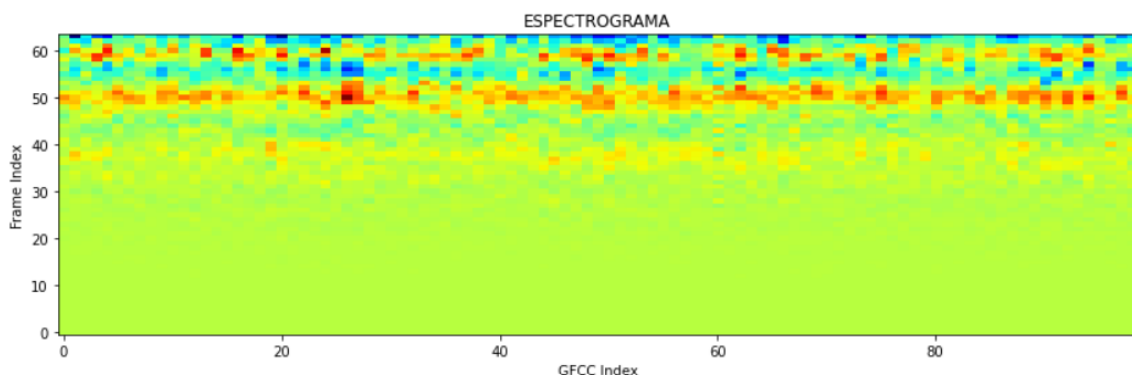


Figura 53: Espectrograma de coeficiente GTCC correspondiente al archivo: 81_25_09_13_H3_corriente.

en el eje X se ven los diferentes valores de los coeficientes y en el eje Y se observan la división de las frecuencias en un número de 64 tramas, son 64 debido a que las tramas corresponden al número de coeficientes.

En esta gráfica podemos observar el valor de los coeficientes GTCC para cada una de las frecuencias que se muestran en el audio y a su vez la mayor o menor intensidad de cada una de dichas frecuencias teniendo en cuenta que el azul sería el valor menos intenso y el tono rojo el valor de mayor intensidad.

4.3.3 Comparación de los GTCC y los MFCC

Como se ha podido observar teniendo en cuenta su función de transferencia los GTCC son unos descriptores frecuenciales muchos más robustos que son capaces de aportar una mayor información respecto a las frecuencias con las que estamos trabajando con nuestra base de datos, mientras que los MFCC están diseñados para el estudio del habla humana y son unos descriptores frecuenciales más específicos para esa labor, por lo cual al no

trabajar con el mismo tipo de frecuencias que las del habla humana deberían de ser menos efectivos a la hora de poder realizar un estudio.

También debemos de tener en cuenta que una de las grandes diferencias son el tipo de filtros que utilizan, siendo más restrictivos los filtros Mel que los filtro GT.

4.4. Machine learning o aprendizaje automático

Machine learning o aprendizaje automático es una subcategoría de la inteligencia artificial la cual estudia el proceso por el cual una máquina es capaz de desarrollar el reconocimiento de patrones o que tenga la capacidad de tener un continuo aprendizaje que le ayude a poder realizar predicciones basadas en los datos que ha podido obtener, realizando diferentes ajustes de manera automática sin que sean programados para eso.

Antes de poder hacerse uso del aprendizaje automático la manera de generar los algoritmos era de la siguiente forma:

- Se escribía el algoritmo de manera manuscrita y este se transmitía en forma de código.
- La máquina ejecutaba el código leyendo una pequeña cantidad de datos.
- Se generaban las predicciones.

A diferencia de esta manera de trabajo el aprendizaje automático da una mayor autoridad e importancia a la máquina, ya que es quien genera el algoritmo capaz de realizar las predicciones. Esto es gracias a que la máquina es capaz de aprender en base a un entrenamiento teniendo en cuenta unos datos iniciales, con los cuales es capaz de ir realizando y a su vez optimizando las predicciones que es capaz de realizar.

Por lo tanto, la labor del ser humano en este tipo de trabajo consiste en ser capaz de seleccionar y confeccionar los datos, establecer las características del entrenamiento y comprender las predicciones que va realizando el algoritmo que ha sido generado por la máquina.

Estas características del aprendizaje automático han hecho que este sea un parte fundamental del Big Data.

A partir de aquí podemos diferenciar tres tipos diferentes de aprendizaje automático

1. Aprendizaje supervisado
2. Aprendizaje no supervisado
3. Aprendizaje por refuerzo

4.4.1 Aprendizaje supervisado

El aprendizaje supervisado es aquel que para realizar su entrenamiento necesita que los datos incluyan la solución correcta en lo que denominamos como “etiquetas”.

En este tipo de aprendizaje es necesario el uso de una base de datos para poder realizar el entrenamiento, la cual, como ha dicho con anterioridad deberá de incluir los datos de entrada y a su vez las etiquetas que serán las variables de salida, de tal forma que el algoritmo sea capaz a partir de ellas de entrenarse y realizar así unas predicciones precisas.

Tras realizar el entrenamiento tendremos que hacer uso de otra base de datos para así poder testarlo y ver que funciona de manera correcta. Esta segunda base de datos al igual que la primera deberá de tener tanto los datos de entrada como las variables de salida para cada uno de ellos. La diferencia entre la base de datos de entrenamiento y la base de datos de prueba es que esta última contendrá datos nuevos que no han sido utilizados para realizar el entrenamiento del algoritmo. Por lo tanto, con esta base de datos, la maquina a partir de los datos de entrada tendrá que predecir cuales serán las variables de salida y una vez obtenidas se compararan con las variables reales que se obtienen de cada uno de los datos de entrada para ver si las predicciones han sido correctas o no.

Este proceso debe de realizarse con una base de datos diferente a la del entrenamiento ya que puede darse el caso de que al entrenar y testear el algoritmo con los mismo datos de entrada, este sea capaz de generar unas buenas predicciones pero cuando se le introducen datos nuevos no sea capaz de generarlas. A esta situación se le denomina sobreentrenamiento y es cuando el algoritmo consigue memorizar los datos de entrenamiento, pero no es capaz de realizar una buena predicción cuando los datos son nuevos. Este tipo de aprendizaje será el que utilicemos en este trabajo.

El esquema del aprendizaje supervisado sería el siguiente:

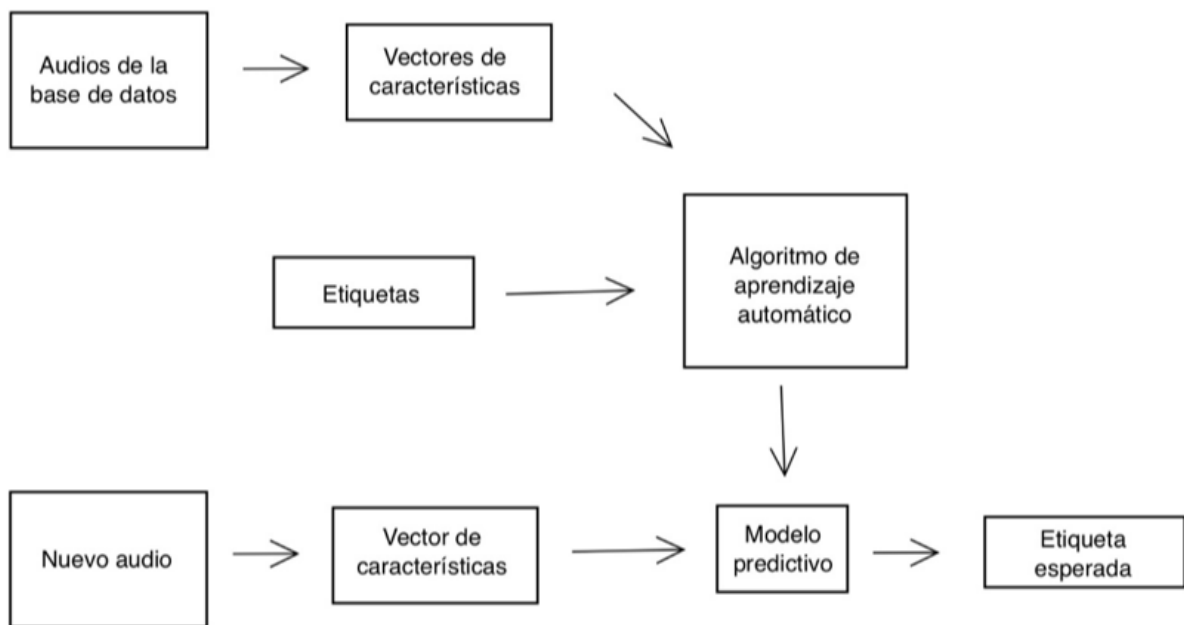


Figura 54: Esquema de como realizar aprendizaje supervisado.

4.4.2 Aprendizaje no supervisado

El aprendizaje no supervisado se diferencia del aprendizaje supervisado en que en este caso, el algoritmo no entrena teniendo unos datos en los cuales se relacionan las entradas y las salidas, sino que la es capaz de construir modelos a partir de las propias características de los datos de entrada para así poder realizar sus predicciones.

Para realizar este proceso de entrenamiento realiza los siguientes pasos:

- 1- Segmentar el conjunto de datos en relación a los atributos compartidos.
- 2- Detectar las anomalías que no son capaces de encajar en ningún conjunto de datos.

3- Simplificar los datasets agregando variables con atributos similares.

El esquema del aprendizaje no supervisado es el siguiente:

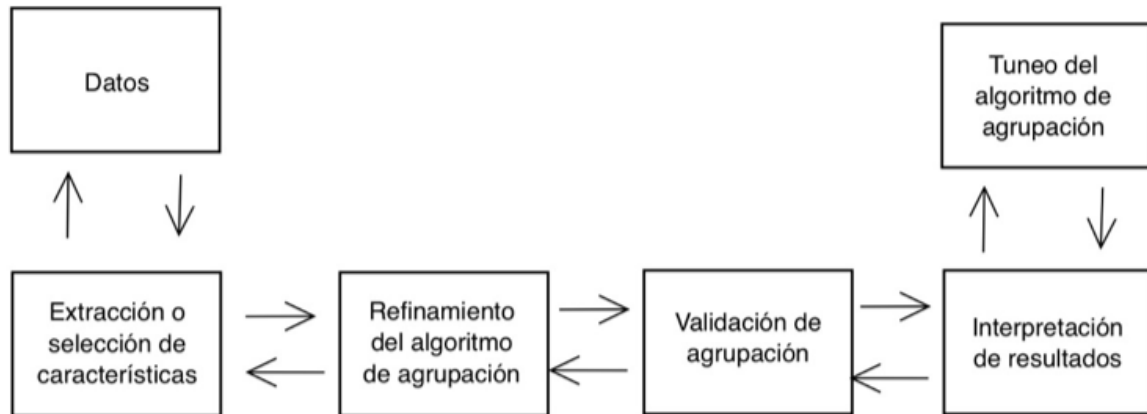


Figura 55: Esquema de como se realiza el aprendizaje no supervisado

De esta manera se consigue que la maquina estudie la estructura intrínseca de los datos.

Este tipo de aprendizaje se utiliza para resolver problemas de tanto de agrupación como de reducción de la dimensionalidad.

4.4.3 Aprendizaje por refuerzo

Este tipo de aprendizaje automático se caracteriza por ser capaz de funcionar con una cantidad pequeña de datos de entrenamiento. Este tipo de aprendizaje es capaz de realizar predicciones siguiendo la estrategia de prueba y error que consiste en utilizar “recompensas” para reforzar el comportamiento deseado del algoritmo.

Este sistema es capaz de aprender por sí solo a partir de las recompensas y las penalizaciones, por lo cual lo que busca a la larga es maximizar su nivel de recompensa. Por ello, es muy útil cuando se conoce cuál es el paso para lograr un resultado y cuál es el resultado, pero se desconoce el camino para lograrlo. Este proceso requiere mucha iteración.

Los pasos por seguir para que nuestro algoritmo aprenda serán los siguientes:

- 1- Observación del entorno.
- 2- Decidir cómo actuar.
- 3- Actuar de acuerdo con esa decisión, lo cual puede modificar el entorno.
- 4- Recibir una recompensa o una penalización.
- 5- Aprender de las experiencias anteriores y así poder depurar la estrategia.
- 6- Iterar hasta que sea capaz de obtener la estrategia óptima.

4.4.4 Métodos de clasificación que se van a usar.

4.4.4.1 Clasificación mediante árboles de decisión.

Los árboles de decisión son un método de aprendizaje automático supervisado que es usado para la clasificación. Es capaz de manejar problemas de salida múltiple y el coste computacional para la realización de este algoritmo es bajo ya que tiene el costo tiene una

relación logarítmica con respecto del número de puntos de datos que se usan para entrenar el árbol.

La figura 24 nos muestra como es un árbol de decisión. Se ha usado para generarlo el estudio 3, el caso general, de los descriptores MFCC.

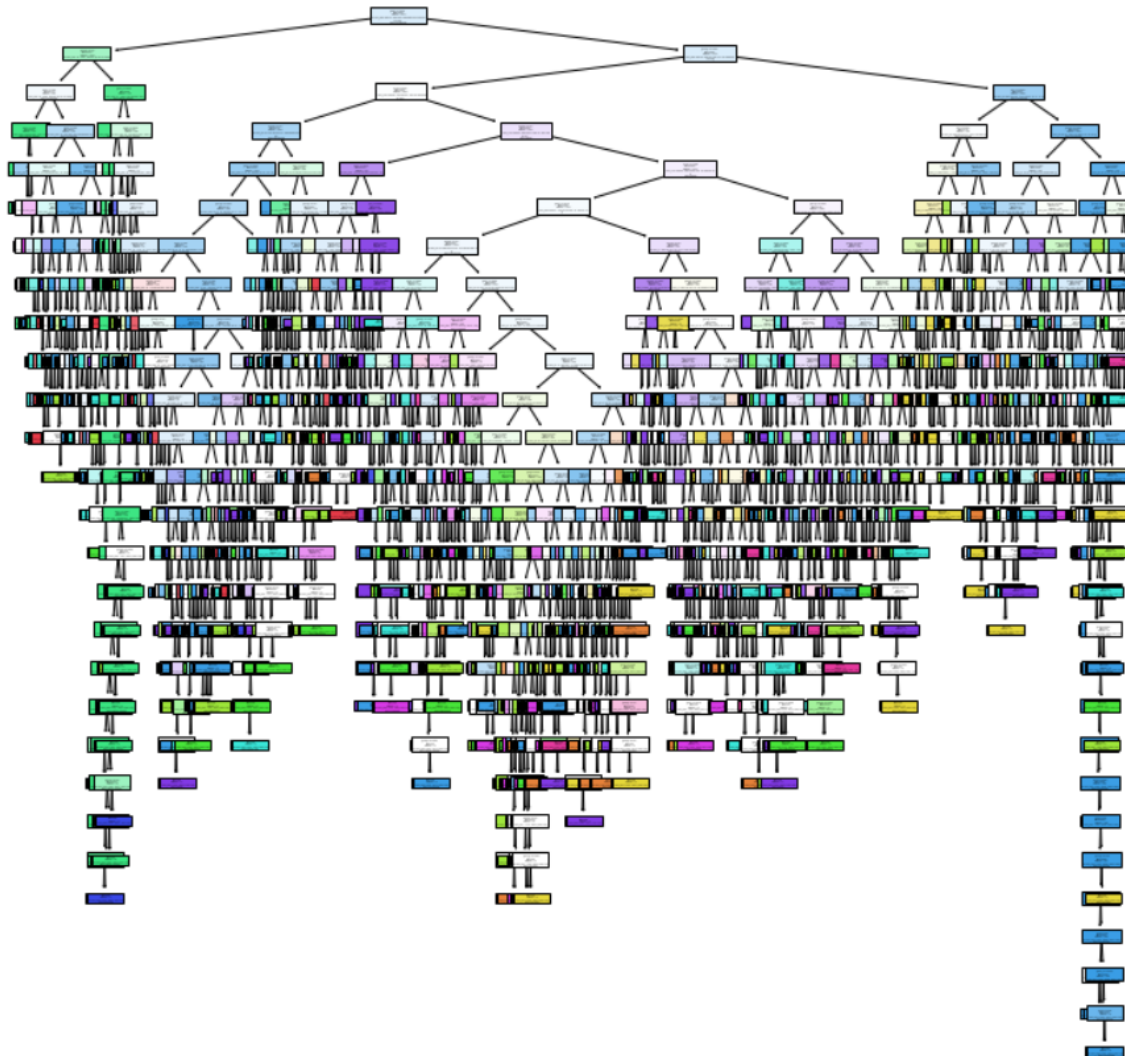


Figura 56: Árbol de decisión del caso de estudio 3

Para realizar este tipo de entrenamiento haremos uso de la librería sklearn para Python la cual se caracteriza en el uso de herramientas simples y a su vez con una gran eficacia para poder realizar análisis predictivos de datos.

4.4.4.2 Clasificación mediante redes neuronales convolucionales, CNN.

Antes de comenzar a hablar de las redes neuronales convolucionales, se ha de explicar que es una red neuronal y en este caso una red neuronal multicapa.

Una red neuronal consiste en un conjunto de algoritmos computacionales que están diseñados según la estructura cerebral humana. Tienen una gran capacidad de aprender a partir de lo que se denomina datos de entrada.

La red neuronal simple, es un modelo matemático el cual está inspirado en una única neurona. Se utiliza para clasificadores binarios. Este tipo de redes está formado:

- Una entrada, por donde se introducirán los datos de entrada.
- Los pesos los cuales ofrecen un valor predeterminado para el inicio del aprendizaje, estos pesos se actualizan con el paso del entrenamiento.
- Una función de activación o no lineal, las cuales toman un único número y realizan una operación determinada sobre él. Las más comunes son la Sigmoide, la ReLu o la Tanh.
- La suma ponderada, la cual consiste en suma cada valor de entrada con su peso correspondiente.
- La salida, muestra nuestra salida predicha.

La figura muestra cómo sería la estructura de una red neuronal simple.

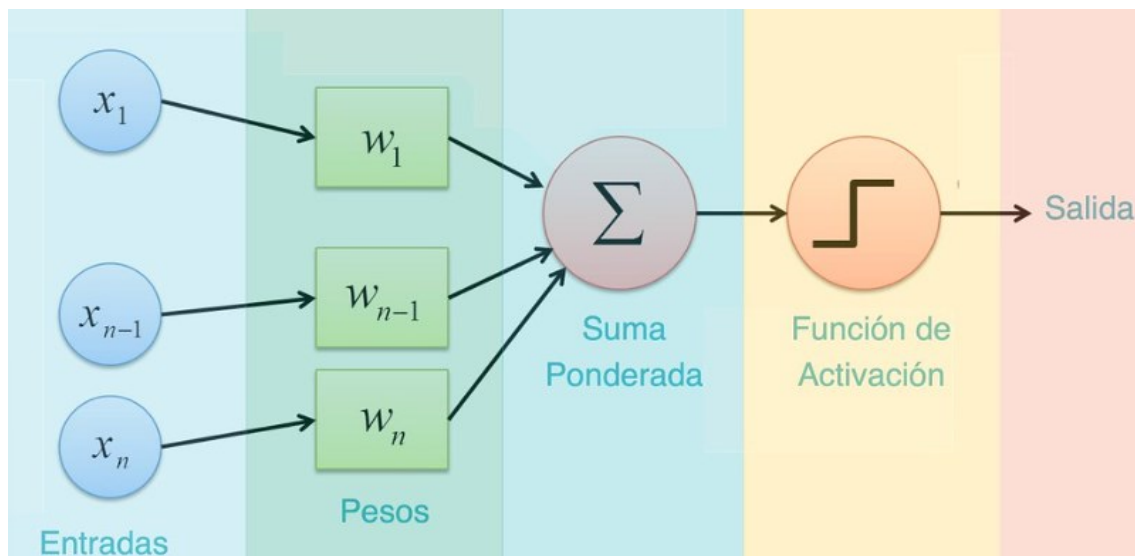


Figura 57: Estructura red neuronal simple

El modelo de red neuronal multicapa consiste básicamente en lo mismo que el modelo neuronal simple con el añadido de que encontramos un número mayor de capas, las cuales se denominan capas ocultas. Este tipo de algoritmo también puede ser llamado como algoritmo de retropropagación y está formado por dos etapas, una hacia delante y otra hacia atrás.

Una manera sencilla de entender el funcionamiento de estas etapas sería la siguiente: las funciones de activación se propagan desde la capa de entrada hasta la capa de salida (etapa hacia delante) y el error entre el valor que se obtiene y el valor que se debería de obtener, tiene lugar en la capa de salida para después poder así modificar los pesos (etapa hacia atrás).

Este tipo de red neuronal también está caracterizada para los casos en los que la decisión se basa en un modelo binario.

La figura muestra un esquema de cómo sería una red neuronal multicapa:

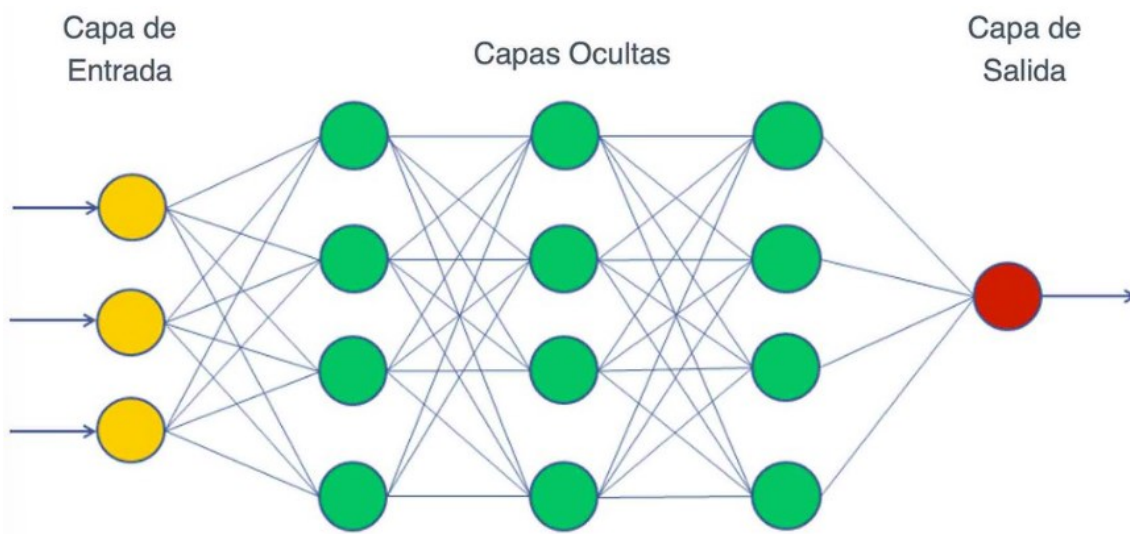


Figura 58: Esquema red neuronal multicapa

Por último, se va a explicar en qué consiste la red neuronal convolucional. Estas redes son parecidas a las redes neuronales multicapa. La gran diferencia consiste en que, en las redes convolucionales, cada una de las capas se entrena específicamente para una única tarea, con lo cual, podemos reducir de forma significativa el número de capas ocultas con respecto a las redes neuronales multicapa.

Este tipo de redes neuronales son muy utilizadas para el análisis de imágenes, ya que son capaces de detectar características simples y realizar composiciones más complejas para detectar.

Estas redes neuronales están compuestas por:

- Una capa de partida, que es equivalente a la capa de entrada.
- Capas de convolución, en las cuales se realizan operaciones de sumas y productos entre la capa de partida y los n filtros (las neuronas), con los cuales se genera un mapa de características. Después de que se le aplique la convolución, el mapa de características se pasa por la función de activación.
- Capas de reducción, en que lo que ocurre es que se disminuye el número de parámetros para quedarnos únicamente con las características más comunes. En este proceso el método pierde precisión, pero mejora la compatibilidad.
- Capa clasificadora, la cual ha de tener el mismo número neuronas como número de clases se desea clasificar. Es lo equivalente a la red salida.

La figura nos muestra un esquema de como podría ser una red convolucional.

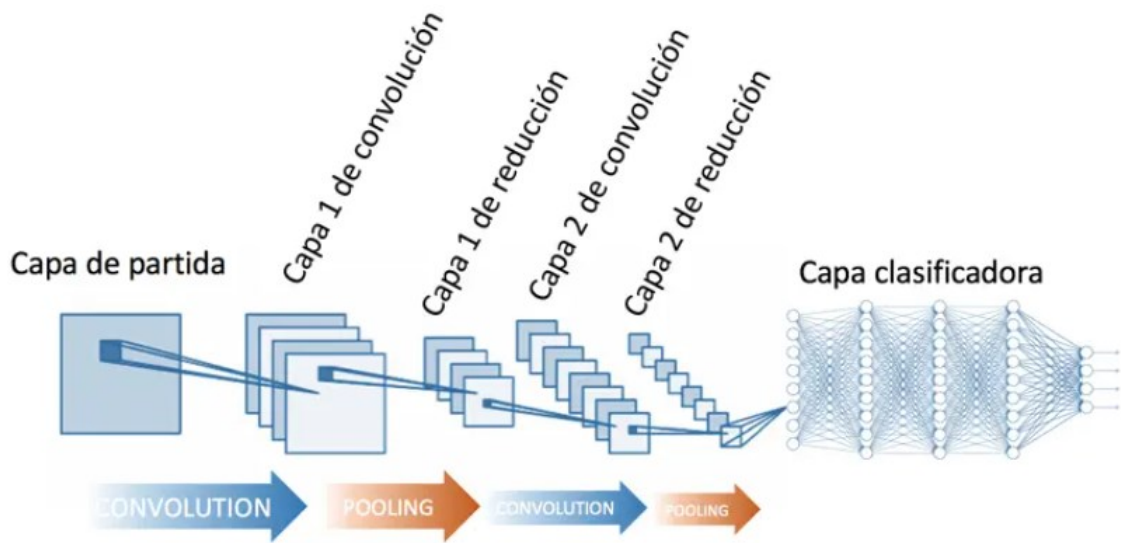


Figura 59: Esquema de una red convolucional

4.5 Definiciones de métricas

Se van a explicar algunos conceptos básicos para entender los datos que vamos a obtener tras el entrenamiento del algoritmo.

- Matriz de confusión:
Es un instrumento, el cual, es capaz de visualizar el comportamiento del algoritmo generado por nuestra máquina. Cada una de las columnas va a ilustrar el número de predicciones de cada clase, mientras que las filas nos muestran las instancias en la clase real. En resumen, ayuda a mostrar que aciertos y errores se están dando en nuestro modelo.

- Precisión:
Es una métrica la cual mide una relación entre todo lo que se ha predicho de manera correcta con respecto a todo lo que se ha predicho como positivo.

$$Precision = \frac{Todo\ predicción\ positiva\ correcta}{Predicción\ positiva\ correcta + Predicción\ positiva\ falsa}$$

- Recall:
Esta métrica se puede entender como teniendo en cuenta todas las predicciones positivas, cuáles son las que logró encontrar nuestro algoritmo.

$$Recall = \frac{Todo\ predicción\ positiva\ correcta}{Positivos\ correctos + Falsos\ negativos}$$

- F1 score:
Es la media armónica entre la precisión y el recall.

$$F1\ score = 2 * \frac{Precisión * Recall}{Precisión + Recall}$$

4.6 Construcción de la base de datos.

Una vez sabido como obtener los descriptores frecuenciales MFCC y GTCC y cómo funciona el aprendizaje automático supervisado, se deberá de confeccionar la base de datos para así realizar el estudio para poder determinar si hay embarcaciones o no hay embarcaciones y la posible identificación de dichos barcos en el caso de que los haya.

Para ello lo primero que se debe de tener en cuenta es que en el aprendizaje automático supervisado los datos de entrada deben de ir en forma de vector o de imagen. Por lo tanto, se debe tener en cuenta que al obtener los descriptores frecuenciales MFCC y GTCC, estos nos devuelven una matriz, la cual deberemos de ordenar en forma de vector para poder así usarlo como dato de entrada, en el caso de que se necesite tenerlos en forma de vector.

Para este estudio se ha considerado que los datos de entrada deben de ser el vector o la 44imagen correspondiente a una matriz en la cual se encuentre la información de un segundo de los audios obtenidos de la base de datos propia. Con respecto al espectrograma ocurrirá lo siguiente en relación a cada uno de los descriptores a la hora de dividir la señal.

- Descriptores MFCC:

Tenemos el espectro total de la señal, en este caso hemos usado la señal correspondiente al audio “10__10_07_13_marDeOnza_Sale.wav”, representado en la figura 60.

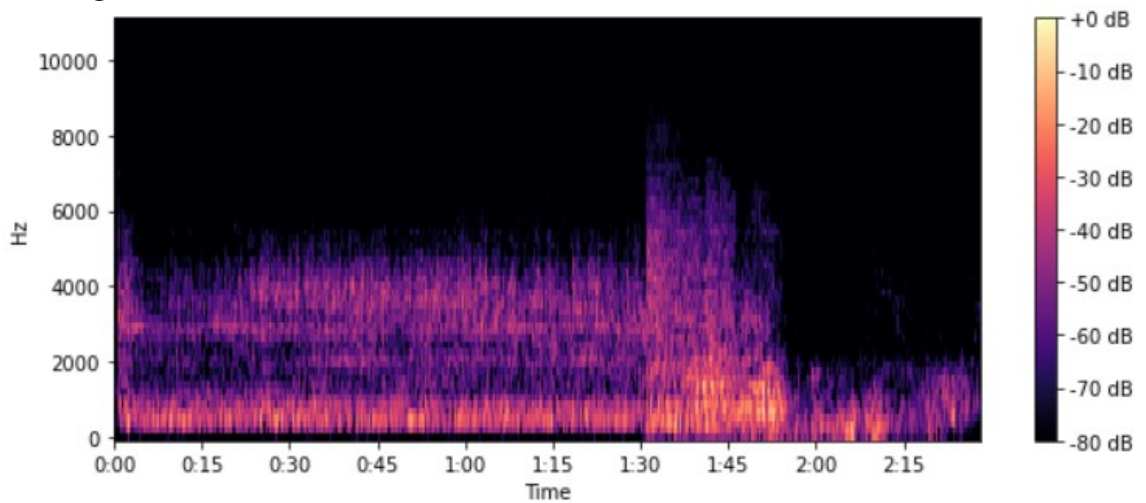


Figura 60: Espectrograma de descriptores MFCC para el audio 10__10_07_13_marDeOnza_Sale

Lo que se desea hacer es “cortar” este espectrograma en segundos.

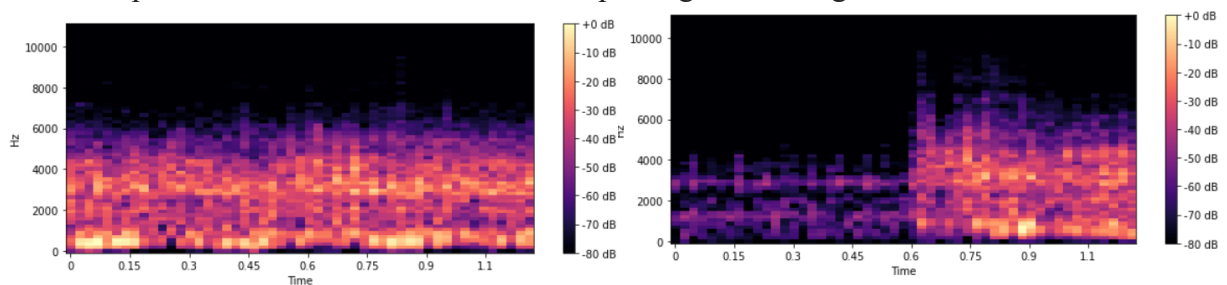


Figura 61: Corte del espectrograma anterior en el segundo 1 Figura 62: Corte del espectrograma anterior en el segundo 2

La figura 61 y 62 corresponderían a los dos primeros segundos del espectrograma representado por la figura 60. Una vez obtenidos los datos de los espectrogramas que se obtienen al dividir la señal original en segundos, se procede a la realización de la base de datos conforme se ha descrito con anterioridad.

- Descriptores GTCC:

Lo mismo ocurre en el caso de los descriptores GTCC, usando el espectrograma de la señal original, en este caso será otra vez la señal obtenida del audio “10_10_07_13_marDeOnza_Sale.wav”, el cual se encuentra representado en la figura 63.

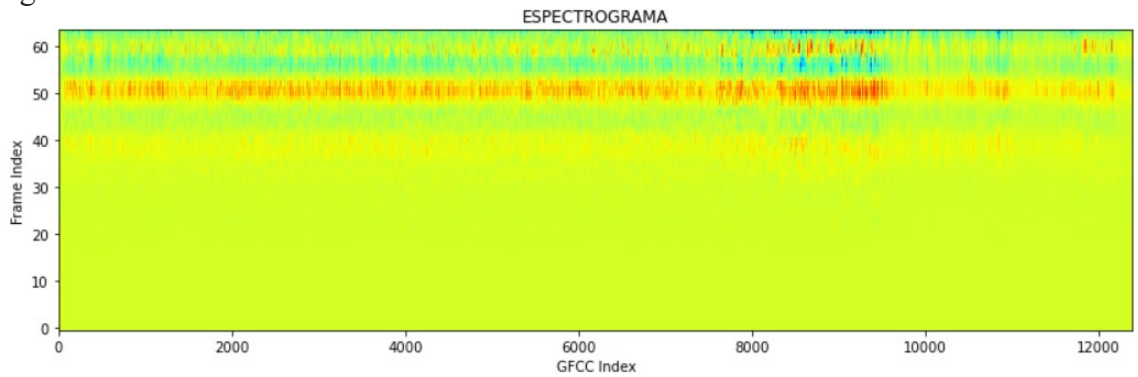


Figura 63: Espectrograma de los descriptores GTCC del audio 10_10_07_13_marDeOnza_Sale

Se obtendrán los datos necesarios para realizar los espectrogramas de un segundo de duración, como se puede ver en las figuras 64 y 65, que representan los segundos uno y dos del espectrograma completo.

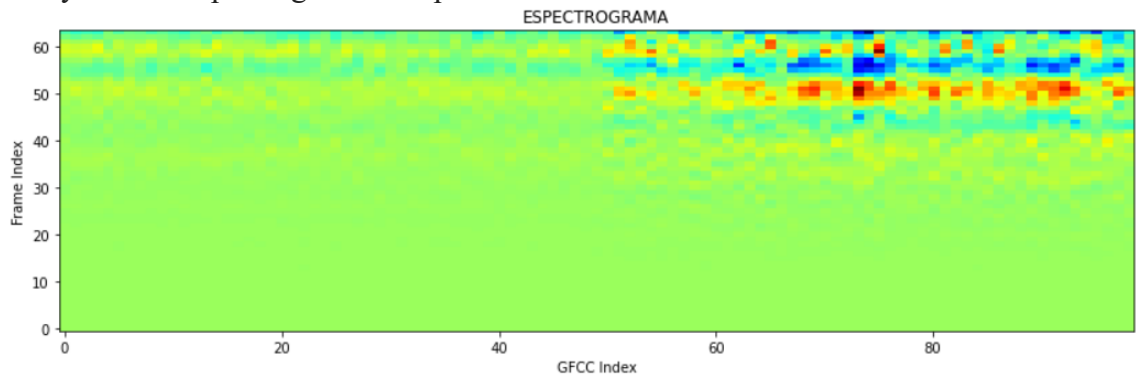


Figura 64: Corte del espectrograma anterior en el segundo 1

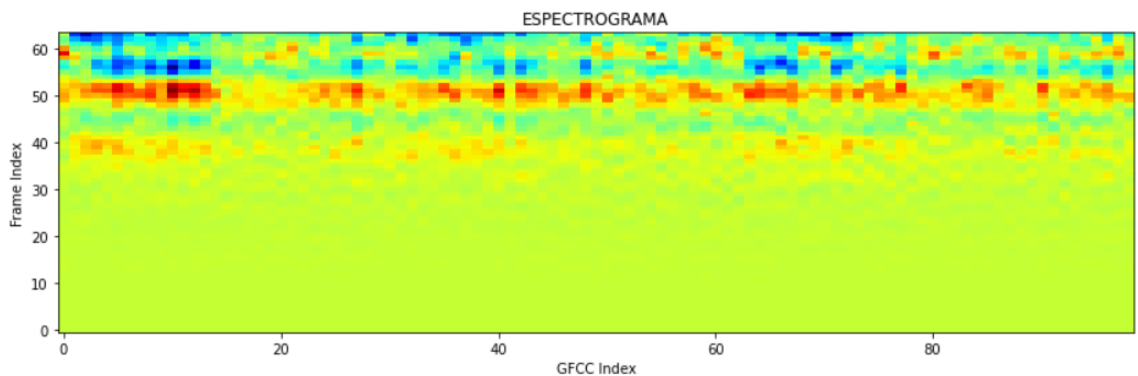


Figura 65: Corte del espectrograma anterior en el segundo 2

Estos espectrogramas obtenidos en la Figura 64 y 65 serán los usados para el posterior análisis y construcción de base de datos.

4.5.1 Construcción de la base de datos para descriptores frecuenciales MFCC para árbol de decisión

Para la construcción de esta base de datos se han seguido los siguientes pasos:

- 1- Generar una función para obtener los datos de entrada cumpliendo los requisitos descritos con anterioridad.
- 2- En la misma función generar las etiquetas correspondientes para cada uno de los datos de entrada.
- 3- A partir de esa función generar los datos de entrada y etiquetas para cada uno de los audios de nuestra base de datos propia.
- 4- Dividir la base de datos que hemos realizado en base de datos de entrenamiento y base de datos de testeo.

El paso 1 y 2 se ha realizado de la siguiente manera:

- Primero se muestrea la señal de audio deseada.
- A la señal muestreada se le han calculado los coeficientes MFCC obteniendo así una matriz de coeficientes.
- Pasar de matriz de coeficientes a vector: pasar de la matriz que contiene los coeficientes de la señal completa, a una matriz que contenga un vector (una fila) por cada segundo de la señal con los coeficientes correspondientes de cada segundo, se ha tenido en cuenta que lo correspondiente a un segundo dentro de la matriz de la señal completa es igual a todas las filas de la matriz delimitadas por un numero específico de columnas. El cálculo para saber el número de columnas es el siguiente:

$$Columnas_{para\ un\ seg} = \left\lceil \left\lfloor \frac{sr}{WIN_SIZE} \right\rfloor * y, \left\lfloor \frac{sr}{WIN_SIZE} \right\rfloor * (y + 1) \right\rceil$$

donde sr es la frecuencia de muestreo, WIN_SIZE el tamaño de ventana e y es el valor del segundo en el que nos encontramos.

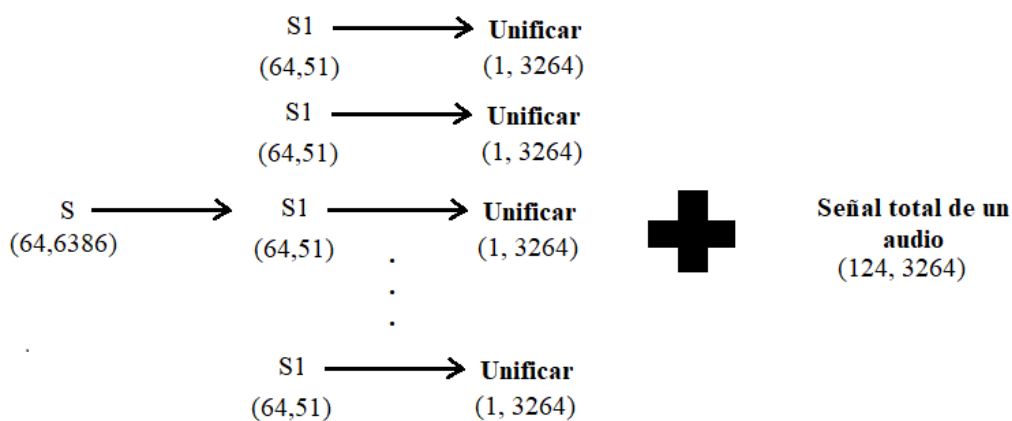


Figura 66: Esquema de como se obtienen los datos de entrada de un audio para árboles de decisión

La figura 66 muestra un esquema de lo descrito anteriormente con respecto a la señal de audio “10_10_07_13_marDeOnza_Sale.wav”. Donde S es el espectrograma de toda la señal, S1 corresponde a cada uno de los segundos en los cuales dividimos la señal, Unificar se refiere al vector que se forma por cada espectrograma de S1 y por último Señal de audio total corresponde a todas las entradas de un segundo dispuestas para poder entrenar el algoritmo. Los valores dentro de los paréntesis hacen referencia al tamaño de las matrices.

Una vez obtenido la matriz correspondiente a un segundo, lo que hacemos es pasar todas las filas a una única fila formada por la unión de todas las filas correspondientes. Mediante un bucle se realiza este proceso para cada uno de los segundos de la señal y se forma una matriz que tiene como filas la información de cada uno de los segundos de la señal que podrán ser usados como señales de entrada para el aprendizaje automático.

Para generar las etiquetas lo que se ha hecho es leer un documento en el cual se encuentra la relación del nombre de los audios con un número que simboliza que tipo de barco es. Entonces mediante un bucle for se ha buscado en la lista de nombres cual es la que coincide con el audio, y una vez se ha encontrado encontrado mediante un segundo bucle for se crea una matriz de una única columna con el número de filas igual a los segundos que dura el audio.

Para el paso 3, lo que hay que hacer es unificar en una misma matriz todos los datos de entrada obtenidos por cada uno de los audios, este paso se puede ver ilustrado en la figura 71.

Y por último en el paso cuatro debemos de separar los datos de entrenamiento y los datos de testeo.

4.5.2 Construcción de la base de datos para descriptores frecuenciales GTCC

Para la construcción de la base de datos con respecto a estos descriptores frecuenciales se han de seguir los mismos pasos que se han seguido en el apartado anterior, la única variación consiste en el paso 1 y 2

Para los descriptores GTCC, al dividir la matriz obtenida al sacar los coeficientes en una matriz que contenga un único segundo, hay que tener en cuenta que cada segundo equivale dentro de la matriz a todas las columnas y un número delimitado de filas. El cálculo para saber el número equivalente de filas es el siguiente:

$$Filas_{para\ un\ seg} = \left[\left[\frac{n^{\circ}\ filas}{tiempo} \right] * y, \left[\frac{n^{\circ}\ filas}{tiempo} \right] * (y + 1) \right]$$

donde tiempo es el número de segundos que contiene el audio del cual estamos obteniendo los GTCC.

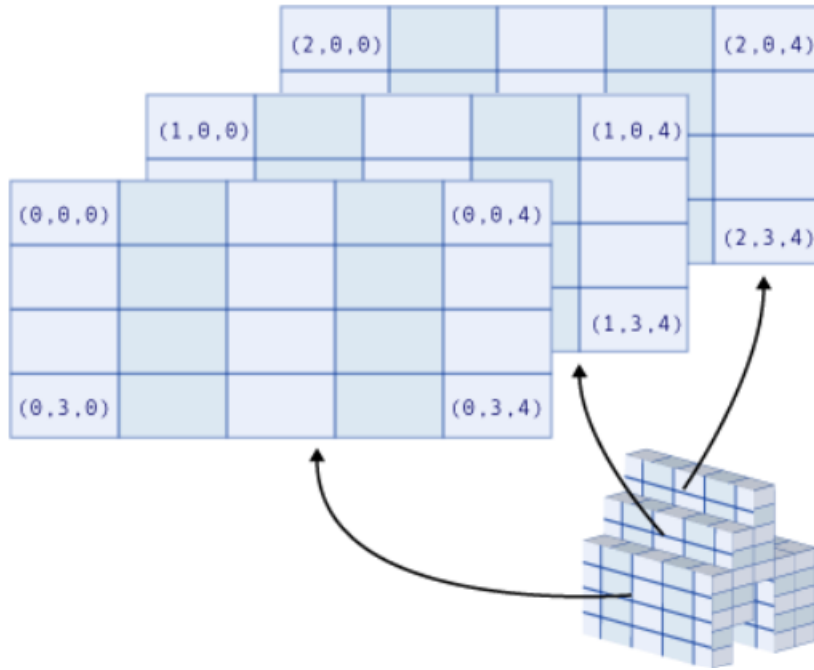


Figura 68: Representación de una matriz tridimensional

Teniendo en cuenta que la matriz vendrá dada por tres coordenadas (x,y,z), se ha usado la coordenada “x” para referirse a que número de imagen es y las coordenadas “y”, “z”, se han utilizado para guardar los valores correspondientes a la imagen.

- Descriptores frecuenciales MFCC:

El paso 1 y 2 se ha realizado de la siguiente manera:

1. Se ha obtenido el espectrograma de la señal completa.
2. La imagen obtenidase ha dividido en imágenes correspondientes a los diferentes segundos y se han guardado en una matriz de tres dimensiones.

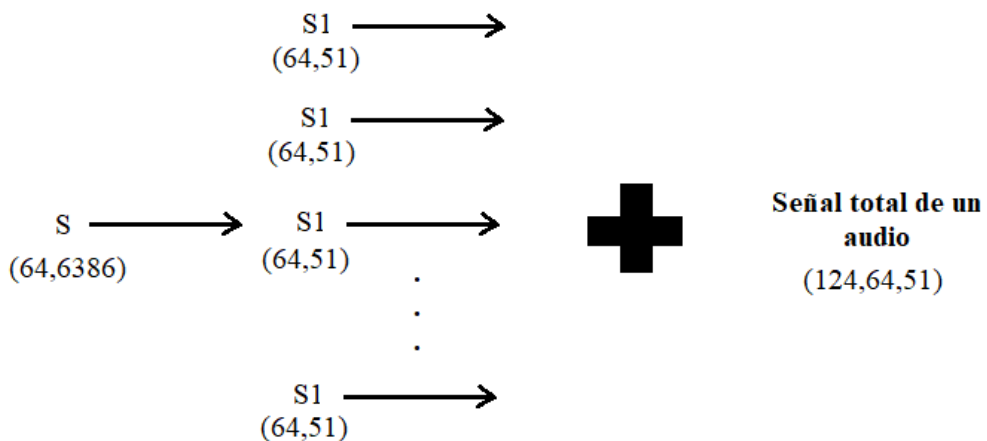


Figura 69: Esquema de como se obtienen los datos de entrada de un audio para red convolucional

La figura 69 muestra un esquema de lo descrito anteriormente con respecto a la señal de audio “10_10_07_13_marDeOnza_Sale.wav”. Donde S es el espectrograma de toda la señal, S1 corresponde a cada uno de los segundos en los cuales dividimos la señal, y por último Señal de audio total corresponde a todas las entradas de un segundo dispuestas para poder

entrenar el algoritmo. Los valores dentro de los paréntesis hacen referencia al tamaño de las matrices.

Se obtiene de la misma manera que en los casos anteriores las etiquetas de los datos de entrada.

3. Después se ha unificado en dos matrices las señales de entrada y las etiquetas de todos los audios, como puede verse en la figura 71.
4. Se ha seleccionado un conjunto de audios de los diferentes tipos de embarcaciones para poder realizar un testeo del aprendizaje que realice el algoritmo.

- **Descriptores frecuenciales GTCC:**

En este caso se ha seguido el mismo esquema para la construcción de la base de datos que como para los descriptores frecuenciales MFCC.

1. Obtener espectrograma de la señal completa
2. Dividir el espectrograma en imágenes de un segundo siguiendo el siguiente esquema:

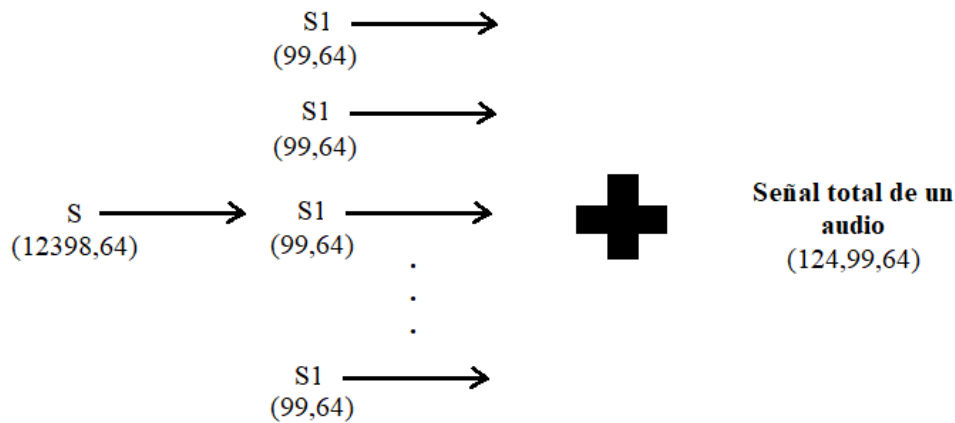


Figura 70: Esquema de como se obtienen los datos de entrada de un audio para red convolucional

Donde S es el espectrograma de toda la señal, S1 corresponde a cada uno de los segundos en los cuales dividimos la señal, y por último Señal de audio total corresponde a todas las entradas de un segundo dispuestas para poder entrenar el algoritmo. Los valores dentro de los paréntesis hacen referencia al tamaño de las matrices.

3. Unificación en dos matrices de las etiquetas y los datos de entrada, como se puede ver en la figura 71.
4. Seleccionar un conjunto de audios, donde haya una variedad de audios en la cual se encuentren todas las embarcaciones y reservarla para realizar el testeo de la red neuronal convolucional.

4.5.4 Esquema resumen referido a la construcción de la base de datos.

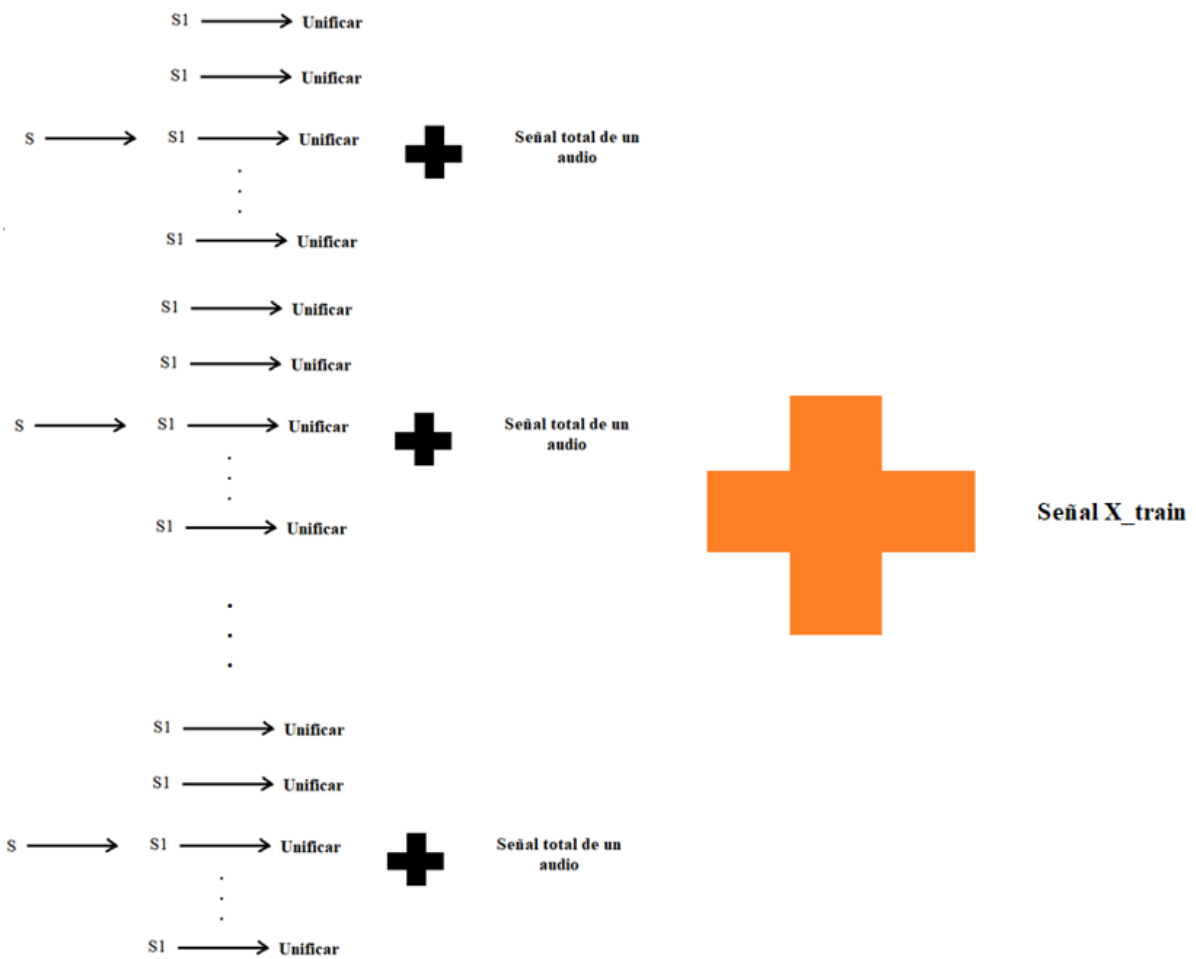


Figura 71: Representación de como se generan los datos de entrada tanto de redes neuronales convolucionales como para árboles de decisión

La figura 71 muestra el esquema que se ha seguido para construir la base de datos necesaria para entrenar tanto la red de aprendizaje automático basada en árboles de decisión como la red neuronal de convolución.

Primero se generan los datos necesarios para obtener el espectrograma de un audio. Después se divide el espectrograma en espectrogramas con la duración correspondiente a un segundo. Se unifica cada uno de esos espectrogramas en un único vector, creando los datos de entrada. Dichos datos de entrada se guardan en una matriz para cada uno de los audios y por último se unen las matrices de todos los audios generando la matriz Señal X_train.

En el caso de la construcción para la red neuronal convolucional, será exactamente igual, pero obviando el paso de unificar.

5. Entrenamiento y resultados.

Tras obtener la base de datos para poder realizar el proceso de aprendizaje automático se ha procedido a realizar los diferentes entrenamientos para ver cómo se comporta el algoritmo con respecto a la base de datos.

5.1 Ejercicios de entrenamiento con árboles de decisión

5.1.1 Caso de estudio 1. Caso binario. Hay barco, no hay barco.

Para este caso hemos clasificado la base de datos de la siguiente manera:

Tipo de Embarcación	Si hay, no hay	Etiquetas
Dredger - dragador	Si	0
Fishboat - pesquero	Si	0
Motorboat - lancha motora	Si	0
Musselboat - mejillonera	Si	0
Natural ambient – ruido ambiental	No	1
Ocean Liner – Transatlántico	Si	0
Passenger – barco de pasajeros, ferry	Si	0
Pilotship – barco piloto	Si	0
RORO- mercancías	Si	0
Sailboat – barco velero	Si	0
Trawler - yate	Si	0
Tugboat - remolcador	Si	0

Tabla 4: Relación de las embarcaciones respecto a su clasificación

5.1.1.1 Resultados obtenidos para los descriptores MFCC.

Métricas de precisión, recall y F1 score.

Accuracy: 97%	Precisión	Recall	F1-score
Barco	98%	99%	99%
No barco	90%	87%	88%

Tabla 5: Tabla de resultados de descriptores MFCC

Matriz de confusión.

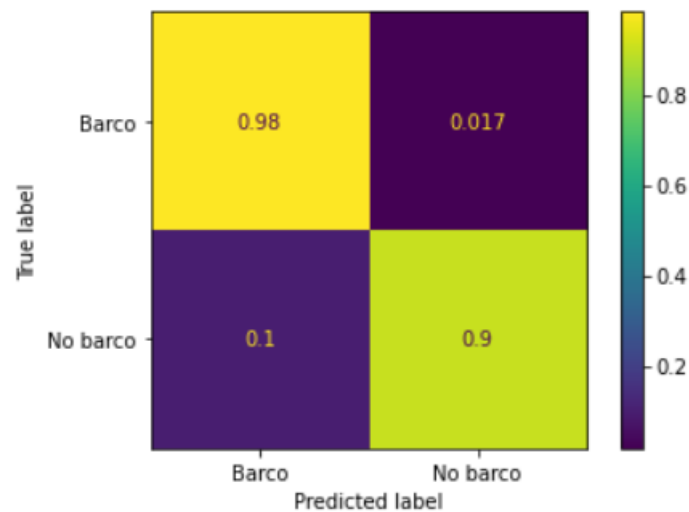


Figura 72: Matriz de confusión para caso binario descriptores MFCC

Como se puede observar para el caso binario para predecir si se encuentran o no se encuentran embarcaciones dentro de los audios, nos encontramos con una media de aciertos del 97%.

A su vez, a través de la precisión de cada uno de los casos, se ha podido medir la calidad del modelo para poder clasificarlos. Encontrando que tenemos una precisión del 98% cuando se encuentran embarcaciones y una precisión del 90% cuando no.

A través del recall se puede medir el ratio de positivos detectados, esto significa que ha representado los positivos reales que hemos obtenido. Es decir, para el caso de que hay embarcaciones sería el porcentaje obtenido entre la cantidad de veces que el algoritmo ha identificado que haya embarcaciones con respecto a todos los casos entre todos los casos. Los datos obtenidos del recall han sido los siguientes, un 99% para el caso de que se encuentre una embarcación y un 87% en el caso de que no lo haya.

Con el F1-score se puede observar el rendimiento que está teniendo el algoritmo para cada uno de los casos, mostrando los valores de 99% y 88%.

Los porcentajes obtenidos para este tipo de clasificación han sido muy satisfactorios. En el caso de que no haya embarcaciones, es verdad que tanto la precisión como el recall son menores, pero esto se debe a que estamos ante un desbalanceo de clase, donde hay más datos de entrada del caso “hay barcos”, que del caso contrario. Esto ha hecho que el algoritmo en caso de duda tienda a elegir entre el caso que sería más probable.

5.1.1.2 Resultados obtenidos para los descriptores GTCC.

Métricas de precisión, recall y F1 score.

Accuracy: 98%	Precisión	Recall	F1-score
Barco	99%	98%	99%
No barco	89%	90%	90%

Tabla 6. Tabla de resultados de descriptores GTCC

Matriz de confusión.

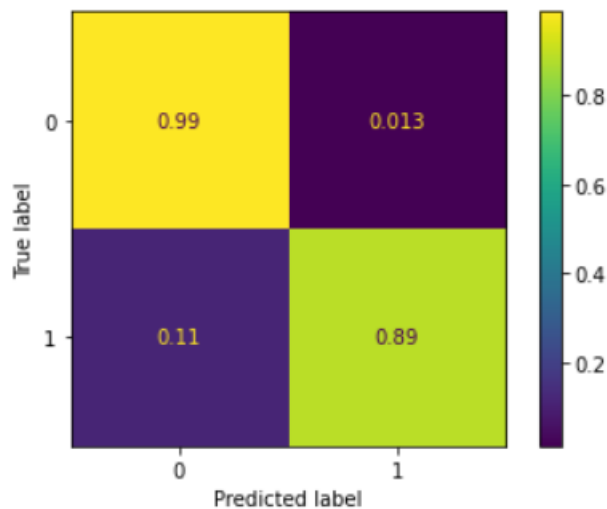


Figura 73: Matriz de confusión para caso binario descriptores GTCC

En este caso como en el anterior volvemos a tener una media de acierto elevada del 98% a la hora de haber realizado la clasificación con este modelo.

Mirando los datos de la precisión, de nuevo se ha observado que el algoritmo muestra una buena calidad a la hora de clasificar las embarcaciones, teniendo una precisión del 99% para el caso en el que hay embarcaciones y del 89% para el caso contrario.

Respecto al recall se ha obtenido también unos datos bastante favorables. Podemos ver que teniendo en cuenta el computo global de las señales de entrada. Nuestro algoritmo ha sido capaz de identificar de manera correcta un 98% para el caso en que hay embarcación y un 90% para el caso contrario. Como se observa en el caso de que no haya embarcación, este porcentaje es menor debido a que como se ha dicho antes es debido al desbalanceo de clases.

Se ha podido confirmar que para este tipo de clasificación y con el uso de estos descriptores, el algoritmo se ha comportado de manera satisfactoria.

5.1.1.3 Comparación de resultados para el caso binario.

MFCC		Accuracy → 97%	
		Hay barco	No hay barco
Precisión		98%	90%
Recall		99%	87%
F1-Score		99%	88%
GTCC		Accuracy → 98%	
Precisión		99%	89%
Recall		98%	90%
F1-Score		99%	90%

Tabla 7: Tabla de comparación de resultados de ambos descriptores

Esta tabla muestra los porcentajes que se han obtenido para cada uno de los casos, en la parte superior se encuentran los datos de los descriptores frecuencias MFCC y en la parte inferior los datos de los descriptores frecuencias GTCC.

Observando detenidamente, se puede observar que los datos obtenidos por ambos descriptores frecuenciales han sido bastantes similares. Pero el uso de los GTCC supone una leve mejora a la hora de identificar cuando no hay embarcaciones, haciendo a estos descriptores frecuenciales un poco menos vulnerables frente al desequilibrio de clases. Por lo tanto, hemos podido afirmar que para este caso el uso de GTCC es un poco mejor con respecto al uso de los descriptores frecuenciales MFCC.

5.1.2 Caso de estudio 2. No hay barco, es barco pequeño o es barco grande.

Para este caso hemos clasificado la base de datos de la siguiente manera:

Tipo de Embarcación	Grande, pequeño o no hay	Etiquetas
Dredger - dragador	Grande	0
Fishboat - pesquero	Pequeño	1
Motorboat - lancha motora	Pequeño	1
Musselboat - mejillonera	Pequeño	1
Natural ambient – ruido ambiental	No hay barco	2
Ocean Liner – Transatlántico	Grande	0
Passenger – barco de pasajeros, ferry	Pequeño	1
Pilotship – barco piloto	Pequeño	1
RORO- mercancías	Grande	0
Sailboat – barco velero	Pequeño	1
Trawler - yate	Pequeño	1
Tugboat - remolcador	Grande	0

Tabla 8: Relación de las embarcaciones con respecto a su clasificación.

5.1.2.1 Resultados obtenidos para los descriptores MFCC.

Métricas de precisión, recall y F1 score.

Accuracy: 85%	Precisión	Recall	F1-score
Barco grande	78%	77%	77%
Barco pequeño	87%	87%	87%
No barco	91%	94%	92%

Tabla 9: Tabla de resultados de descriptores MFCC

Matriz de confusión.

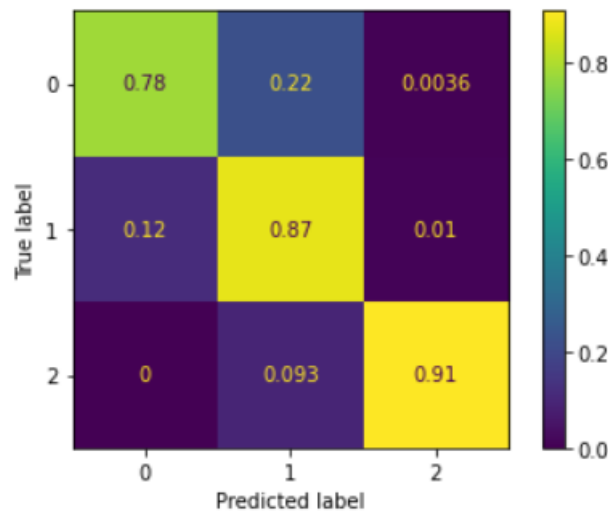


Figura 74: Matriz de confusión para caso de estudio 2, descriptores MFCC

En este caso de estudio se ha obtenido un valor de media de acierto de un 85% a la hora de haber realizado la clasificación en las diferentes clases.

Con respecto a la calidad del algoritmo podemos decir que ha obtenido una precisión del 78% para el caso de que hay una embarcación grande, 87% para el caso de que sea pequeña y un 91% para el caso de que no haya embarcación. Al fijarnos en la matriz de confusión se ha podido apreciar que tanto para el caso de embarcación grande como el de embarcación pequeña, en términos generales, se ha observado que cuando el algoritmo es erróneo en sus predicciones, suele confundir estas dos clases entre ellas mismas. Mientras que en el caso de que no haya barcos, suele confundirse con respecto a la clase donde los barcos son de menor tamaño. Esto se debe, en el primer caso, entre los dos tipos de embarcaciones, a que la diferencia de las frecuencias entre la señal que transmite una embarcación y otra no sean demasiado distintas, habiendo embarcaciones pequeñas cuyo motor genere frecuencias un poco más graves o viceversa, produciendo así el error. En el caso de que no haya embarcaciones, se debe a que el sonido producido por la marea y el oleaje tiene unas frecuencias altas, al igual que las embarcaciones pequeñas, las cuales, puede haber algún caso en el que sean excesivamente altas y se produzca ahí el error.

Teniendo en cuenta el recall, se ha observado que el programa es capaz de identificar el 77% de los datos para el caso en el que es barco es grande, un 87% para el caso en el que

el barco es pequeño y un 94% para el caso en el que no hay barcos. De momento se han seguido manteniendo valores elevados a la hora de contabilizar la cantidad de veces que se ha podido identificar cada uno de los casos, lo cual sigue siendo algo favorable para la validez del algoritmo.

5.1.2.2 Resultados obtenidos para los descriptores GTCC.

Métricas de precisión, recall y F1 score.

Accuracy: 85%	Precisión	Recall	F1-score
Barco grande	79%	79%	79%
Barco pequeño	88%	88%	88%
No barco	89%	89%	89%

Tabla 10: Tabla 11: Tabla de resultados de descriptores GTCC

Matriz de confusión.

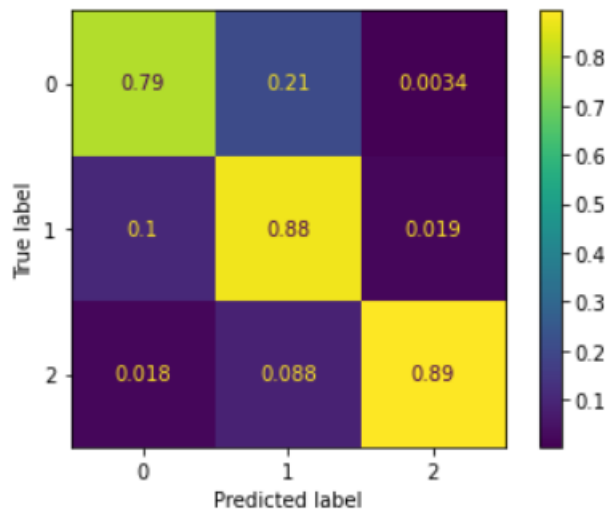


Figura 75: Matriz de confusión para caso de estudio 2, descriptores GTCC

Para este caso se ha obtenido una media de acierto del 85%, a la hora de realizar la clasificación con el uso de este algoritmo.

Analizando la calidad del algoritmo, se han observado los datos de precisión de los tres casos que ocupan este caso de estudio, obteniendo unos valores de 79% de acierto para la situación donde la embarcación es de mayor tamaño, un 88% para cuando la embarcación es de menor tamaño y un 89% en el caso de que no haya barco.

Como se puede observar se han obtenido unos valores de precisión elevada que, teniendo en cuenta, su porcentaje de identificación o recall, 79%, 88% y 89% respectivamente, dan fe de que el algoritmo en este caso de estudio sigue funcionando de una manera eficiente.

Al igual que ha pasado en el caso de los descriptores MFCC, viendo la matriz de confusión podemos ver que, por norma general, el algoritmo en caso de error confunde embarcaciones pequeñas con embarcaciones grandes y viceversa y en caso de que el error se produzca con respecto a que no haya embarcaciones, por norma, suele confundirse con

embarcaciones pequeñas. Esto se debe a lo explicado en el epígrafe anterior con respecto a los límites que delimitan una embarcación grande con una pequeña y una embarcación pequeña con el ruido marino.

5.1.2.3 Comparación de resultados para el caso de estudio 2.

MFCC	Accuracy 85%	→	Barco grande	Barco pequeño	No hay barco
Precisión			78%	87%	91%
Recall			77%	87%	94%
F1-Score			77%	87%	92%
GTCC	Accuracy 85%	→			
Precisión			79%	88%	89%
Recall			79%	88%	89%
F1-Score			79%	88%	89%

Tabla 12: Tabla de comparación de resultados de ambos descriptores

En la tabla anterior, se puede ver que los valores obtenidos para los descriptores MFCC y GTCC, son prácticamente los mismos en este caso de estudio.

Por lo tanto, en este caso, se ha podido observar que el uso de los coeficientes MFCC o GTCC, ha de deberse al uso que quiera darse.

En el caso en el que se valore más, el hecho de saber si no hay barco, entonces, se ha de usar los descriptores MFCC que tienen un mayor índice de acierto en ese caso que los GTCC. En el hipotético caso de que lo que se valore más sea saber que tipo de embarcación es la que se ha detectado, entonces ha de usarse los descriptores frecuenciales GTCC.

En general, ambos descriptores frecuenciales obtienen un comportamiento prácticamente igual con el algoritmo que se está utilizando.

5.1.3 Caso de estudio 3. Caso multisalida. Todos los tipos de barcos

Para este caso hemos clasificado la base de datos de la siguiente manera:

Tipo de Embarcación		Etiquetas
Dredger	Barco dragador	1
Fishboat	Barco de pesca	2
Motorboat	Lancha motora	3
Musselboat	Barco mejillonero	4
Natural ambient	Sonido ambiente	5
Ocean Liner	Trasatlántico	6
Passenger	Crucero, ferry	7
Pilotship	Barco piloto	8
RORO	Mercancías	9
Sailboat	Barco de vela	10
Trawler	Yate	11
Tugboat	Remolcadora	12

Tabla 13: Relación de las embarcaciones con respecto a su clasificación.

5.1.3.1 Resultados obtenidos para los descriptores MFCC.

Métricas de precisión, recall y F1 score.

Accuracy: 65%	Precisión	Recall	F1-score
Dredger	41%	35%	38%
Fishboat	58%	51%	54%
Motorboat	49%	51%	50%
Musselboat	50%	46%	48%
Natural ambient	93%	81%	86%
Ocean Liner	58%	64%	61%
Passenger	71%	76%	73%
Pilotship	64%	50%	56%
RORO	68%	65%	67%
Sailboat	47%	49%	48%
Trawler	42%	45%	43%
Tugboat	50%	52%	51%

Tabla 14: Tabla de resultados de descriptores MFCC

Matriz de confusión.

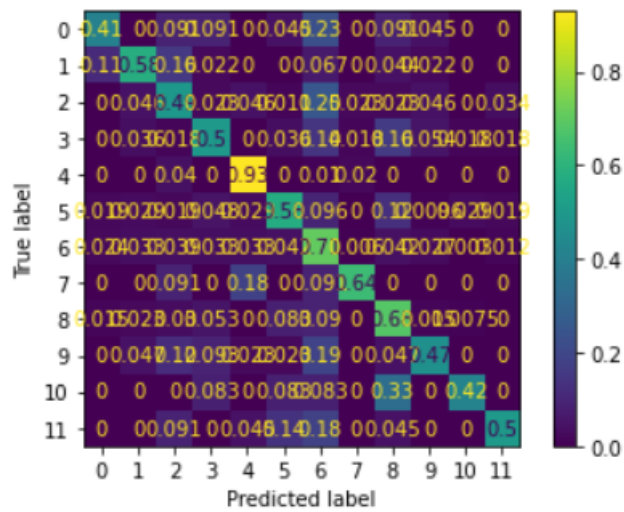


Figura 76: Matriz de confusión para caso de estudio 3, descriptores MFCC

En este caso de estudio se obtiene una media de acierto del 65%, teniendo en cuenta que se ha realizado una clasificación con doces salidas mediante el uso de árboles de decisión es un valor satisfactorio, aunque a la hora de clasificar no sea lo suficientemente efectivo. Aunque hay que tener en cuenta que este valor puede ser un poco más elevado de debido a que al no haber homogeneidad en los resultados, casos de aciertos como puedan ser el de los sonidos ambientales, como los barcos de pasajeros, que muestran un pico de acierto más superior. Hacén que la media de acierto aumente, aunque el algoritmo no clasifique correctamente.

Teniendo en cuenta los valores de precisión que se han dado se puede ver que sigue diferenciando muy bien el ejemplo de que hay o no hay barcos, esto se afirma al ver la matriz de decisión en la fila cuatro observamos que la mayoría de los valores de la columna de esa fila son cero.

También se puede apreciar que en el caso de que no haya barcos, el algoritmo los confunde con barcos pequeños como pueden ser las lanchas motoras (motorboat) o con los barcos de pasajeros como puedan ser el ferry (passengers).

Viendo el resultado de las precisiones, también se ve que en este caso al igual que en el caso dos tiene una mayor facilidad para reconocer a los barcos que son pequeños con respecto a los barcos grandes.

Con respecto al ratio de positivos detectados, es decir, el recall, se puede observar que los valores obtenidos para la mitad de los casos no llegan a más del 50% y salvo alguna excepción no alcanza más del 60%. Esto junto a la poca precisión que hemos obtenido en la gran mayoría de las clases hace el algoritmo en este caso de estudio no sea lo suficientemente eficaz a la hora de clasificar.

Por último al observar la matriz de confusión, se puede ver que el barco que con más frecuencia confunde el algoritmo es el Passenger (pasajeros), esto se debe a lo que ya se ha explicado con anterioridad con respecto a que es la embarcación de la cual tenemos una mayor cantidad de datos de entrada.

5.1.3.2 Resultados obtenidos para los descriptores GTCC.

Métricas de precisión, recall y F1 score.

Accuracy: 69%	Precisión	Recall	F1-score
Dredger	63%	73%	68%
Fishboat	60%	63%	62%
Motorboat	55%	50%	52%
Musselboat	54%	54%	54%
Natural ambient	88%	88%	88%
Ocean Liner	68%	61%	64%
Passenger	74%	78%	76%
Pilotship	80%	57%	67%
RORO	66%	69%	68%
Sailboat	57%	41%	48%
Trawler	40%	36%	38%
Tugboat	63%	81%	71%

Tabla 15: Tabla de resultados de descriptores GTCC

Matriz de confusión.

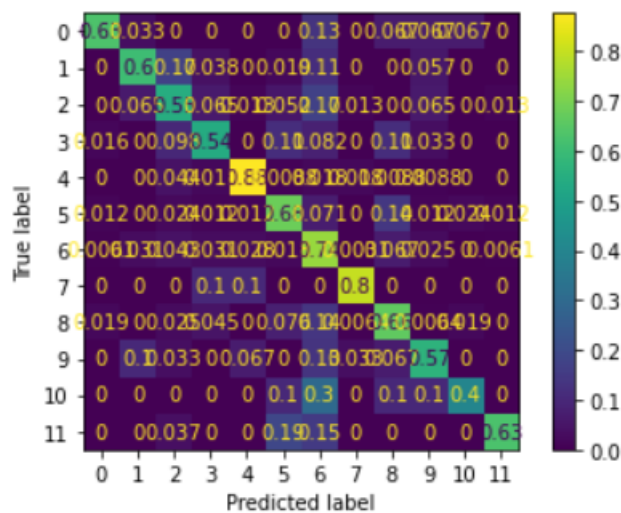


Figura 77: Matriz de confusión para caso de estudio 3, descriptores GTCC

Para estos descriptores obtenemos un porcentaje de acierto medio del 69%. Al igual que con los descriptores MFCC es un valor bastante bueno teniendo en cuenta el método de aprendizaje utilizado, pero sigue sin ser lo suficientemente eficaz como para considerar que realiza una buena clasificación.

En este caso se ha podido observar que los resultados son más homogéneos, aunque se ha seguido encontrando picos de acierto que hacen que aumente la media de acierto “camuflando” así el que el algoritmo no es lo suficientemente preciso para realizar este tipo de clasificación.

Al medir la calidad del modelo, se ha apreciado que la mitad de las embarcaciones, no llegan a más de un 60% de acierto, dos se encuentran entre el 60 y 70% y otras dos, las que se han considerado como los picos de acierto, superan el 70%. Estos valores son bajos para realizar una clasificación competente.

En el caso del recall ocurre lo mismo, lo que hace entender que a la hora de clasificar hay muchos errores ya que mientras más errores haya a la hora de clasificar una clase, menor será el recall de las demás. En este caso 8 clases se encuentran por debajo del 60%, dos clases se encuentran entre el 60% y el 70% y únicamente dos clases alcanzan un porcentaje mayor que el 70%

En este caso seguimos observando que las embarcaciones del tipo passenger (pasajeros) siguen siendo las que más confunde el algoritmo debido a lo que se ha explicado en apartados anteriores.

5.1.3.3 Comparación de resultados para el caso de estudio 3.

MFCC	Accuracy 65%	→	Precisión	Recall	F1-score
Dredger dragador			41%	35%	38%
Fishboat pesquero			58%	51%	54%
Motorboat lancha motora			49%	51%	50%
Musselboat mejillonera			50%	46%	48%
Natural ambient – ruido ambiental			93%	81%	86%
Ocean Liner Transatlántico			58%	64%	61%
Passenger barco de pasajeros, ferry			71%	76%	73%
Pilotship – barco piloto			64%	50%	56%
RORO- mercancías			68%	65%	67%
Sailboat – barco velero			47%	49%	48%
Trawler – yate			42%	45%	43%
Tugboat – remolcador			50%	52%	51%

GTCC	Accuracy 69%	→	Precisión	Recall	F1-score
Dredger dragador	–		62%	69%	65%
Fishboat pesquero	–		65%	55%	60%
Motorboat lancha motora	–		54%	55%	54%
Musselboat mejillonera	–		49%	52%	51%
Natural ambient – ruido ambiental			89%	89%	89%
Ocean Liner Transatlántico	–		71%	62%	67%
Passenger barco de pasajeros, ferry	–		75%	74%	75%
Pilotship – barco piloto			100%	57%	73%
RORO-mercancías			66%	70%	68%
Sailboat – barco velero			46%	46%	46%
Trawler – yate			40%	55%	46%
Tugboat remolcador	–		59%	76%	67%

Tabla 16: Tabla de comparación de resultados de ambos descriptores

Viendo los resultados que obtenemos en la tabla para este caso, cual es un claro ejemplo de salida múltiple, se puede ver que en este caso los resultados de los descriptores GTCC tienen unas mayores prestaciones que los MFCC. Primero ya que obtiene unos resultados más homogéneos al igual que unos valores de recall más elevados, haciendo que el algoritmo obtenga mejores resultados con los descriptores GTCC, que con los MFCC.

Aun así, ninguno de los descriptores utilizando este algoritmo da valores para obtener una clasificación fiable.

5.1.4 Uso de la herramienta SMOTE y resultados para los tres casos de estudio.

La herramienta SMOTE se utiliza para compensar la base de datos, es decir, es una herramienta que se utiliza para hacer que todas las clases de la base tenga la misma cantidad de datos las unas respecto de las otras.

Para ello lo que hace esta función es analizar cada una de las clases y generar de manera automática diferentes datos para dichas clases, y así hacer que todas las clases de nuestra base de datos tengan la misma cantidad de entradas de datos.

El código utilizado para conseguir esta homogeneidad en los datos es el siguiente:

Una vez conseguida dicha homogeneidad y realizado el proceso de entrenamiento y testeo obtenemos estos datos para cada uno de los casos de estudio

- Caso de estudio 1.
 - a) Descriptores frecuenciales MFCC.
 - Métricas de precisión, recall y F1 score.

Accuracy: 97%	Precisión	Recall	F1-score
Hay barco	98%	99%	99%
No hay barco	90%	87%	88%

Tabla 17: Tabla de resultados de descriptores MFCC

Matriz de confusión.

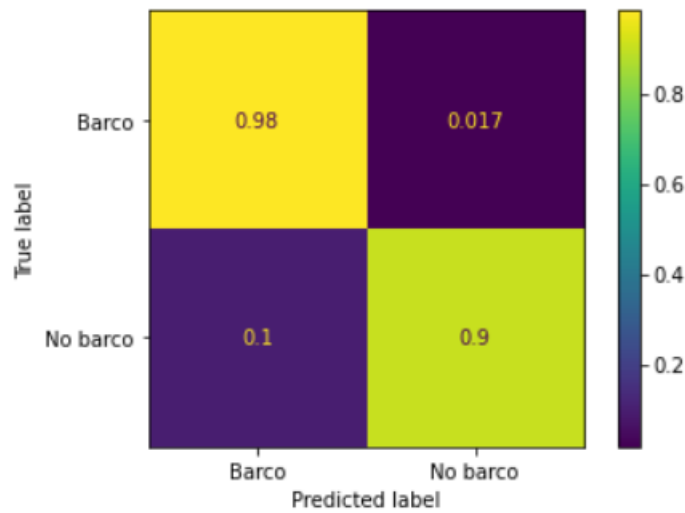


Figura 78: Matriz de confusión para caso binario, descriptores MFCC

- b) Descriptores frecuenciales GTCC.
 - Métricas de precisión, recall y F1 score.

Accuracy: 97%	Precisión	Recall	F1-score
Hay barco	98%	98%	98%
No hay barco	96%	84%	85%

Tabla 18: Tabla de resultados de descriptores GTCC

Matriz de confusión.

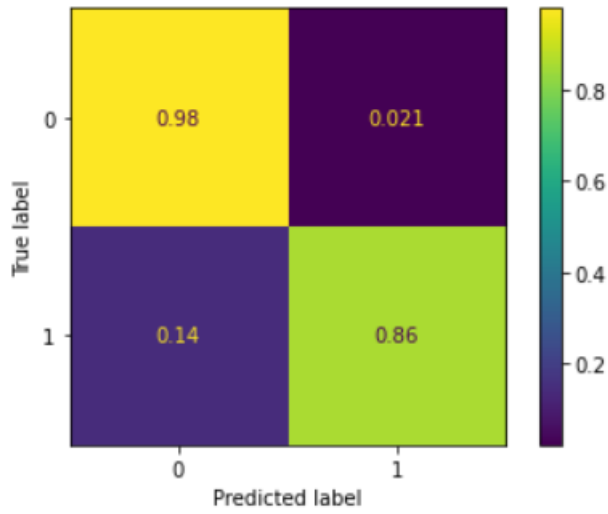


Figura 79: Matriz de confusión para caso binario, descriptores GTCC

- Caso de estudio 2.
 - a) Descriptores frecuenciales MFCC.
Métricas de precisión, recall y F1 score.

Accuracy: 81%	Precisión	Recall	F1-score
Barco grande	71%	75%	73%
Barco pequeño	85%	83%	84%
No barco	88%	89%	89%

Tabla 19: Tabla de resultados de descriptores MFCC

Matriz de confusión.

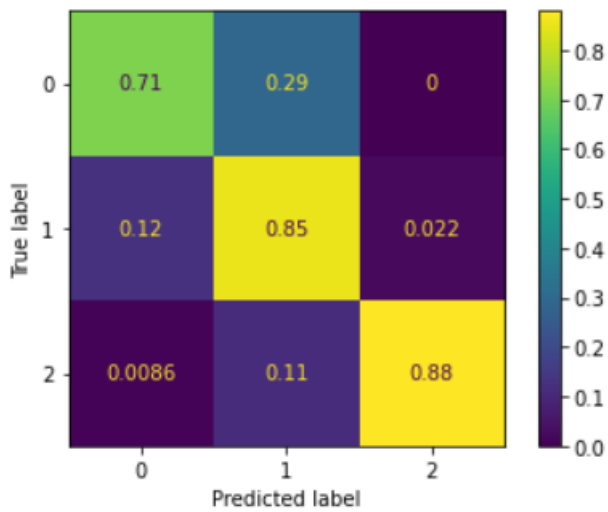


Figura 80: Matriz de confusión para caso de estudio 2, descriptores MFCC

b) Descriptores frecuenciales GTCC.
Métricas de precisión, recall y F1 score.

Accuracy: 81%	Precisión	Recall	F1-score
Barco grande	73%	77%	75%
Barco pequeño	85%	84%	85%
No barco	83%	78%	81%

Tabla 20: Tabla de resultados de descriptores GTCC

Matriz de confusión.

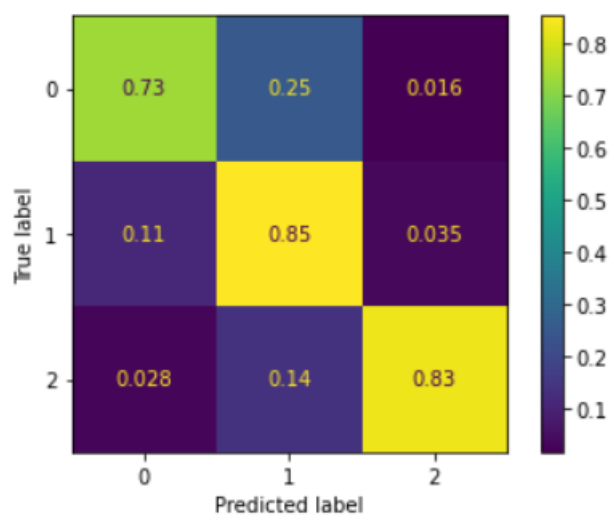


Figura 81: Matriz de confusión para caso de estudio 2, descriptores GTCC

- Caso de estudio 3.
 - a) Descriptores frecuenciales MFCC.
Métricas de precisión, recall y F1 score.

Accuracy: 63%	Precisión	Recall	F1-score
Dredger	48%	50%	49%
Fishboat	53%	61%	56%
Motorboat	53%	68%	59%
Musselboat	37%	44%	40%
Natural ambient	93%	87%	90%
Ocean Liner	63%	63%	63%
Passenger	76%	68%	72%
Pilotship	38%	36%	37%
RORO	66%	54%	59%
Sailboat	38%	46%	42%
Trawler	35%	55%	43%
Tugboat	41%	57%	48%

Tabla 21: Tabla de resultados de descriptores MFCC

Matriz de confusión.

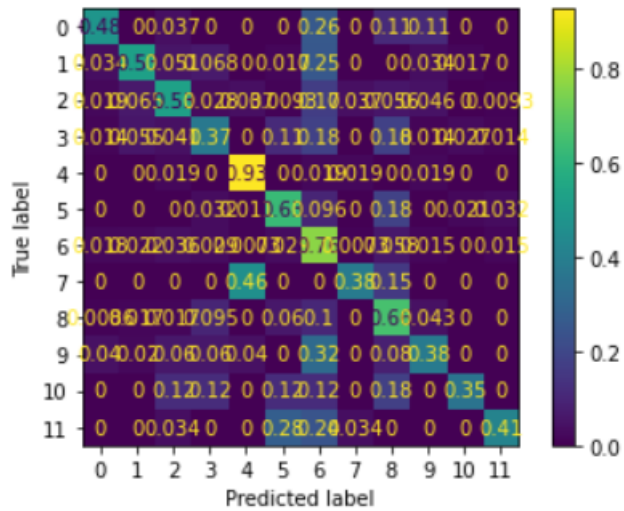


Figura 82: Matriz de confusión para caso de estudio 3, descriptores MFCC

b) Descriptores frecuenciales GTCC.
 Metricas de precisión, recall, F1 score.

Accuracy: 57%	Precisión	Recall	F1-score
Dredger	37%	42%	39%
Fishboat	50%	51%	50%
Motorboat	46%	51%	49%
Musselboat	37%	44%	40%
Natural ambient	89%	89%	89%
Ocean Liner	43%	50%	46%
Passenger	69%	59%	64%
Pilotship	73%	57%	64%
RORO	56%	55%	55%
Sailboat	41%	44%	42%
Trawler	36%	45%	40%
Tugboat	35%	38%	36%

Tabla 22: Tabla de resultados de descriptores GTCC

Matriz de confusión.

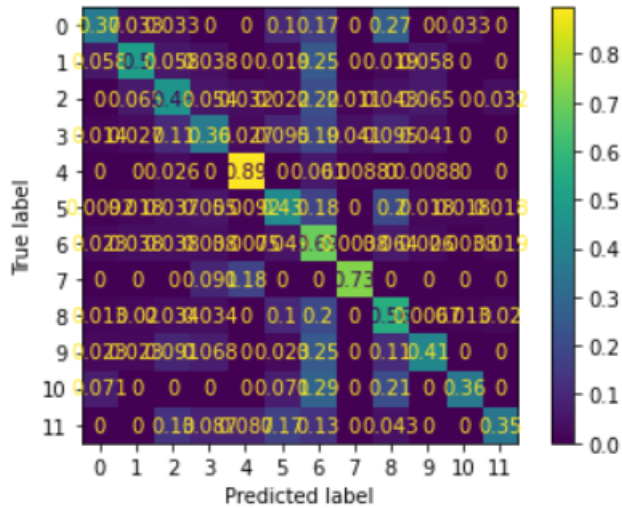


Figura 83: Matriz de confusión para caso de estudio 3, descriptores GTCC

Comparando cada uno de los tres casos de estudio con su homólogo con el uso de la herramienta SMOTE se han podido obtener soluciones las cuales son generales para todos los casos.

El porcentaje de media de acierto es más bajo usando la herramienta, esto ha podido deberse a que la herramienta al recrear datos de entrada buscando relaciones entre un mismo grupo, ha hecho que el algoritmo baje un poco en su porcentaje medio de acierto, ya que no se han usado señales reales y ha podido provocar un mayor error. Sin embargo, encontramos que la precisión de acierto entre clases ahora sea más pareja que cuando no hacemos uso de esta herramienta, provocando que los porcentajes de precisión sean más homogéneos.

5.2 Ejercicios de entrenamiento con red neuronal convolucional

La clasificación de los barcos ha sido exactamente iguales que para el entrenamiento con árboles de decisión, para cada uno de los tres casos de estudio.

5.2.1 Caso de estudio 1. Caso binario. Hay barco no hay barco

5.2.1.1 Resultados obtenidos para los descriptores MFCC

Métricas de precisión, recall y F1 score.

Accuracy: 96%	Precisión	Recall	F1-score
Hay barco	99%	97%	98%
No hay barco	70%	88%	78%

Tabla 23: Tabla de resultados de descriptores MFCC

Matriz de confusión.

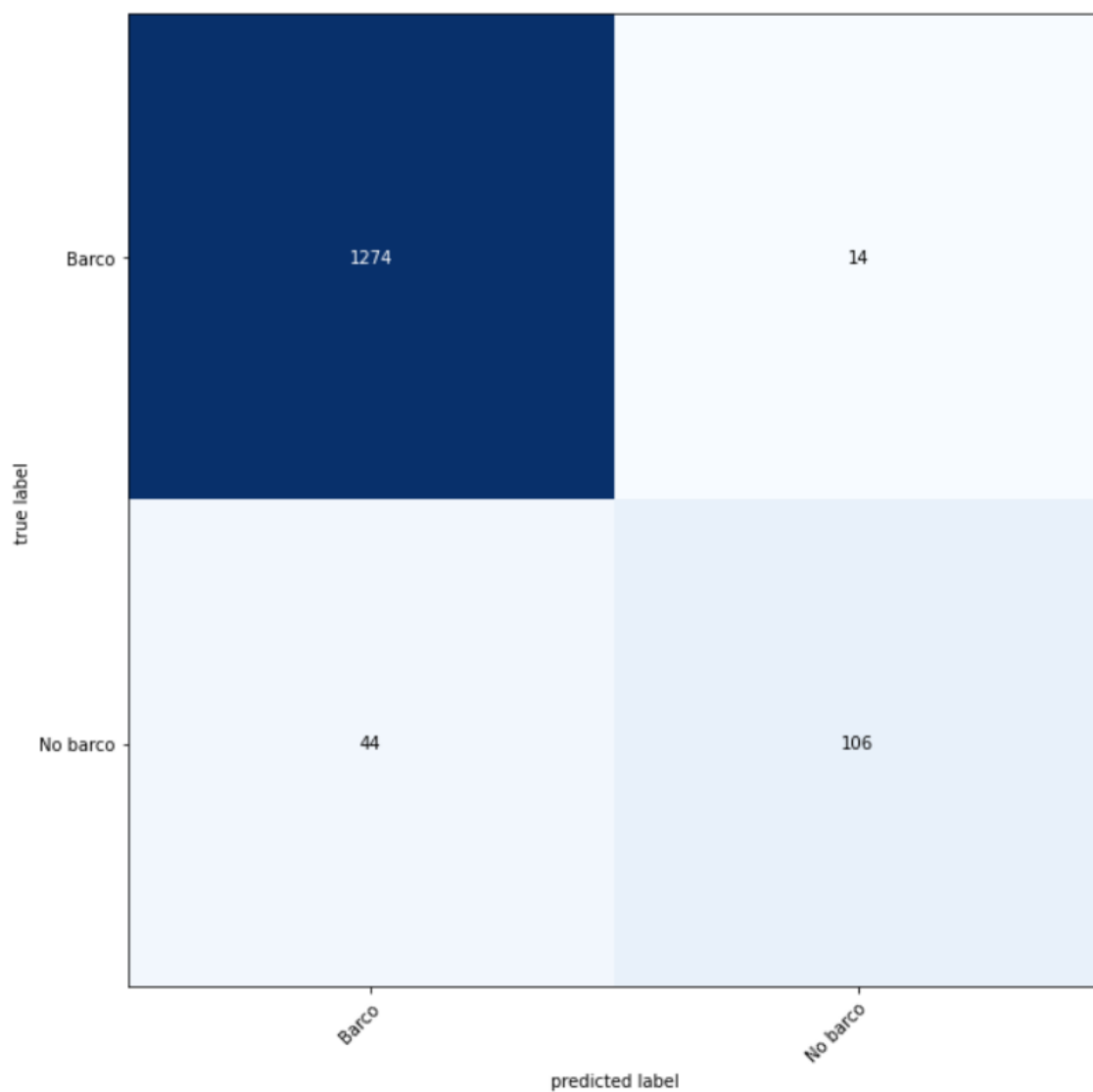


Figura 84: Matriz de confusión para caso binario, descriptores MFCC

Gráfica de entrenamiento del modelo.

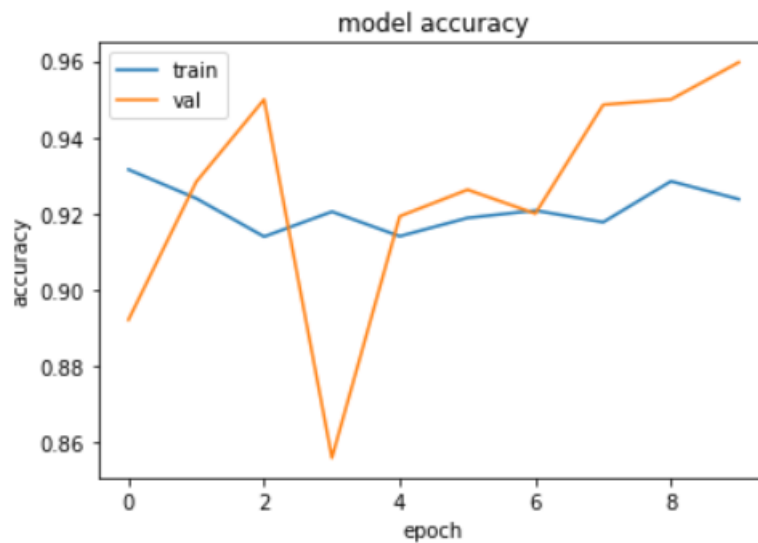


Figura 85: Gráfica de entrenamiento del modelo

Observando primero la figura 85, se ha podido observar que no se ha llegado a lo que se denomina sobre entrenamiento, debido a que los valores de valores de validación no disminuyen y se encuentran creciendo al igual que los valores de entrenamiento.

Se ha observado la calidad del algoritmo y se ha podido ver que los valores de precisión son elevados, pero hay una gran diferencia entre la precisión de el caso en el que hay barco, con el caso en el que no lo hay. Esto ha podido deberse a que el algoritmo tiende a elegir el caso con más número de datos de entrada.

Con respecto al recall, se han podido obtener unos datos favorables. Como era de esperarse se ha dado el caso en el que el recall del caso no hay embarcaciones es mayor al del caso que si las hay. Esto se debe a que el algoritmo ha tenido una tendencia a favorecer al caso más numeroso y esto también ha podido verse mostrado en el recall.

Por último, podemos decir que ha tenido una media de acierto elevada, 96%. Teniendo en cuenta los otros casos, ha sido el valor más pequeño, pero aun así es un resultado satisfactorio para este tipo de clasificación.

5.2.1.2 Resultados obtenidos para los descriptores GTCC

Métricas de precisión, recall y F1 score.

Accuracy: 99%	Precisión	Recall	F1-score
Hay barco	99%	99%	99%
No hay barco	99%	99%	99%

Tabla 24: Tabla de resultados de descriptores GTCC

Matriz de confusión.

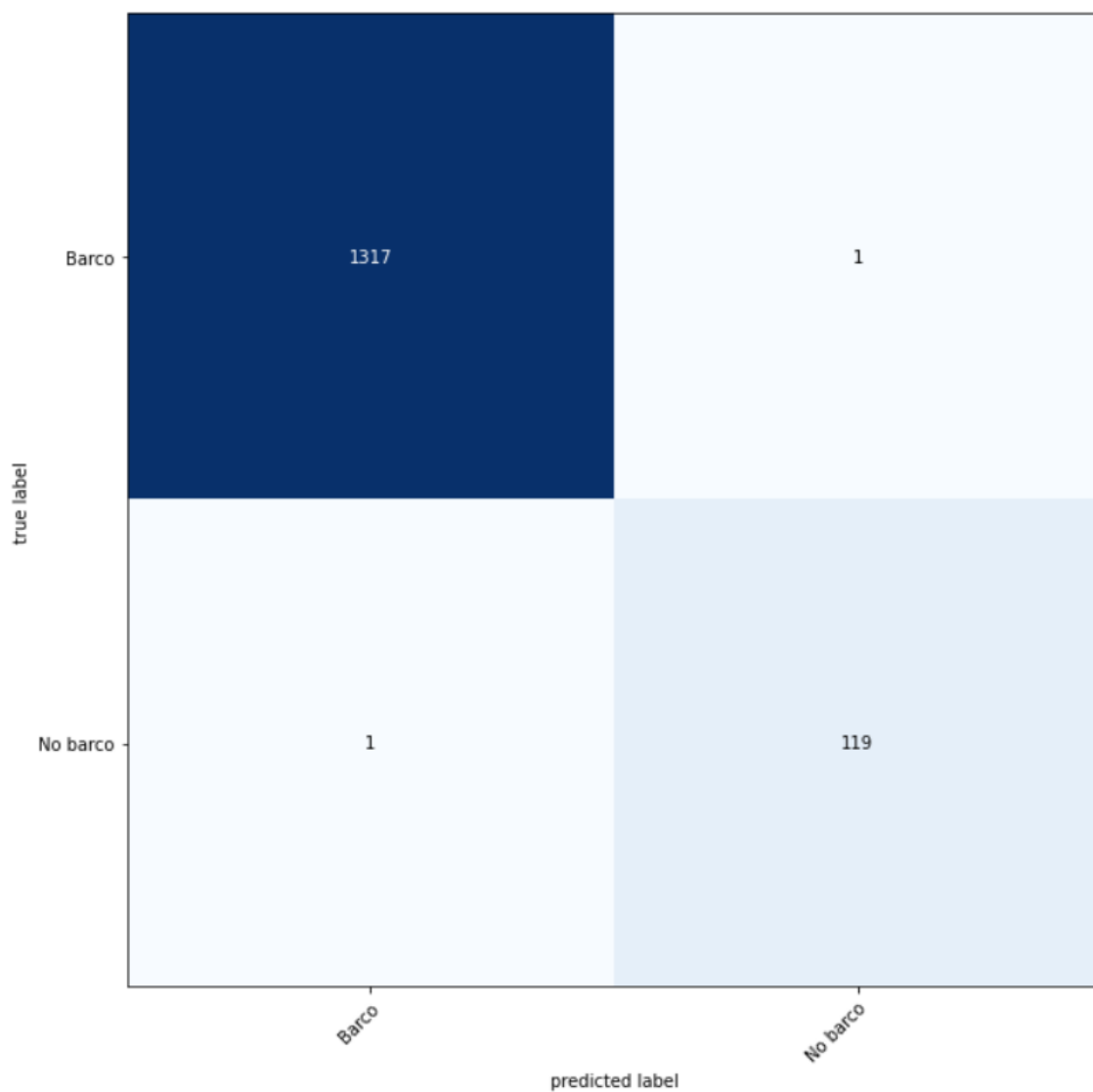


Figura 86: Matriz de confusión para caso binario, descriptores GTCC

Gráfica de entrenamiento del modelo.

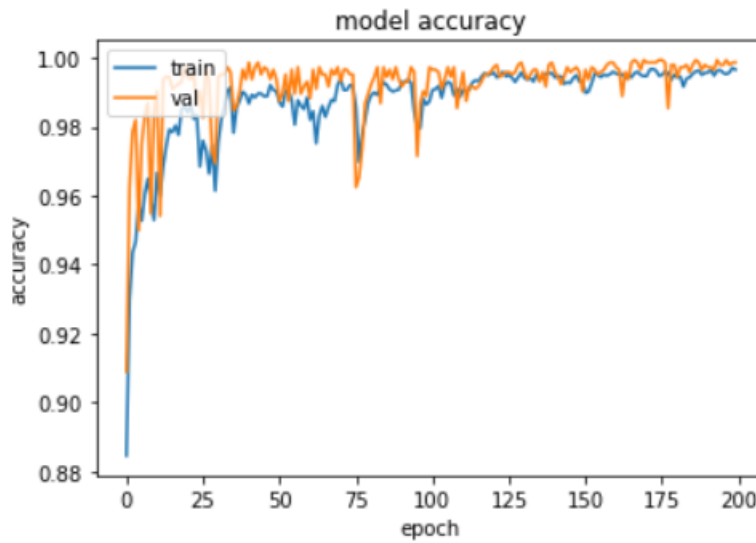


Figura 87: Gráfica de entrenamiento del modelo

En este modelo, gracias a la gráfica de entrenamiento, se ha podido observar como tampoco se ha llegado al sobre entrenamiento, ya que los valores que hemos obtenido de entrenamiento y de validación se mantienen constantes. Para que haya sobre entrenamiento, los valores de validación deberían de disminuir. Esto se debe a que el algoritmo, después de tantas épocas de entrenamiento, empieza a aprender de memoria a clasificar los datos de entrenamiento y deja de ser fiable cuando pasa por el testeo y se proponen los datos de validación.

Para los descriptores GTCC y usando una red CNN, se han podido obtener los datos más elevados de media de acierto, un 99%. En este caso binario, el algoritmo ha sido capaz de clasificar casi sin error todos los valores del testeo. Esto ha hecho de este caso el más eficiente para clasificación binaria.

Observando el recall, solo podemos decir que el algoritmo ha sido capaz de contabilizar correctamente los casos totales de verdaderos positivos que se han encontrado en el test de validación. Esto supone el modelo funciona correctamente y respalda los datos obtenidos por la precisión.

La calidad de este modelo ha sido excelente y sus datos de clasificación son los mayores obtenidos en todo el estudio.

5.2.1.3 Comparación de resultados para el caso binario

MFCC		Accuracy → 96%	Hay barco	No hay barco
Precisión			99%	70%
Recall			97%	88%
F1-Score			98%	78%
GTCC		Accuracy → 99%		
Precisión			99%	99%
Recall			99%	99%
F1-Score			99%	99%

Tabla 25: Tabla de comparación de resultados de ambos descriptores

Comparando ambos resultados, podemos decir que nos hemos encontrado con el valor más elevado para el caso de estudio binario y a su vez con el caso con menos porcentaje. Por lo tanto, se ha podido afirmar que, para el caso de estudio binario, con el uso de red neuronal convolucionales, el uso de los descriptores frecuenciales GTCC es mejor con respecto el uso de los descriptores MFCC.

Ambos descriptores han sido validos para una implementación ya que en ambos casos los datos que se han obtenido han sido satisfactorios, pero sin duda los datos de los descriptores GTCC, son mejores con respecto a los MFCC

5.2.2 Caso de estudio 2. No hay barco, hay barco grande o hay barco pequeño

5.2.2.1 Resultados obtenidos para los descriptores MFCC

Métricas de precisión, recall y F1 score.

Accuracy: 73%	Precisión	Recall	F1-score
Barco grande	75%	80%	78%
Barco pequeño	78%	67%	72%
No barco	55%	79%	65%

Tabla 26: Tabla de resultados de descriptores MFCC

Matriz de confusión.

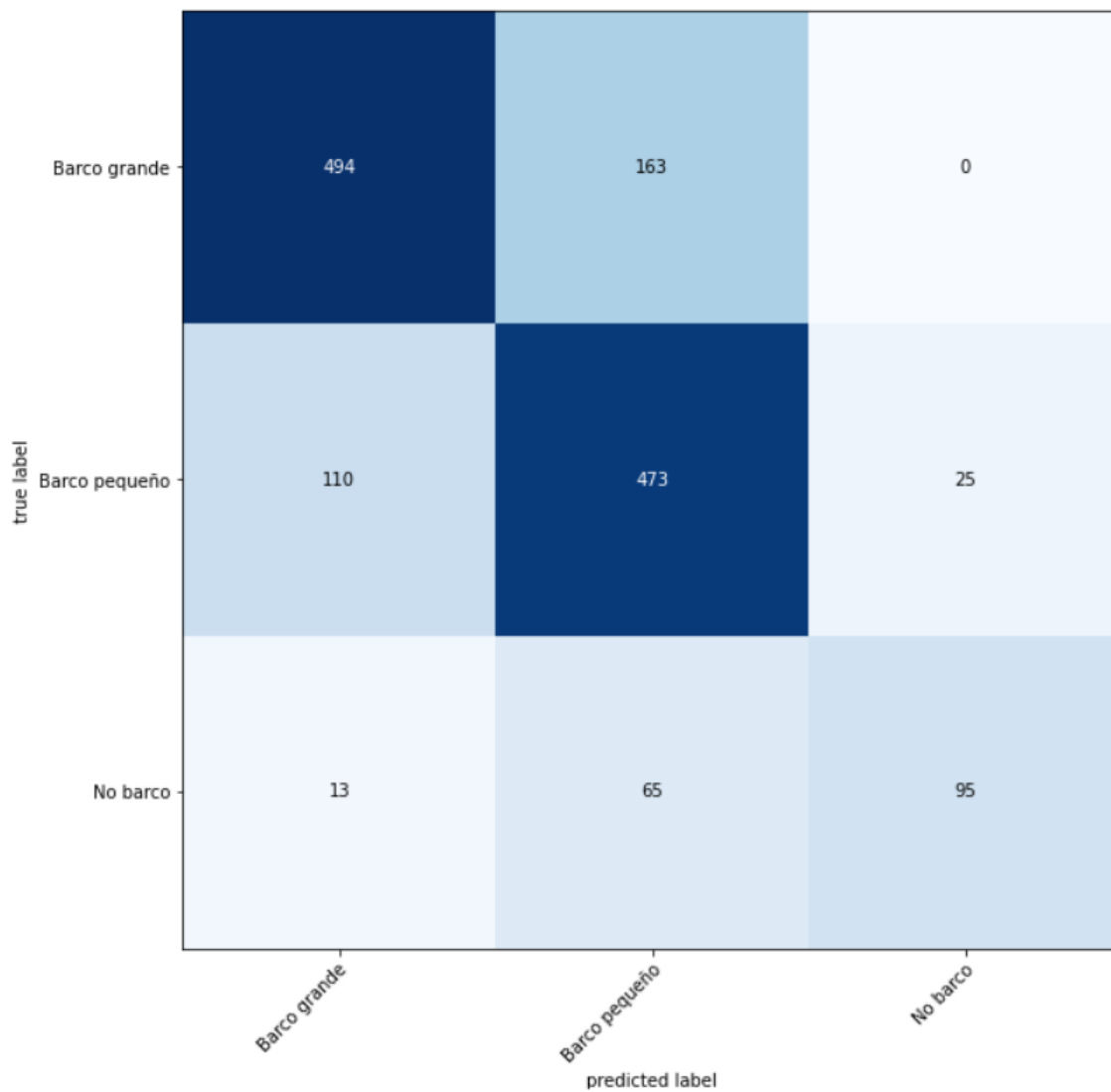


Figura 88: Matriz de confusión para caso de estudio 2, descriptores MFCC

Gráfica de entrenamiento del modelo.

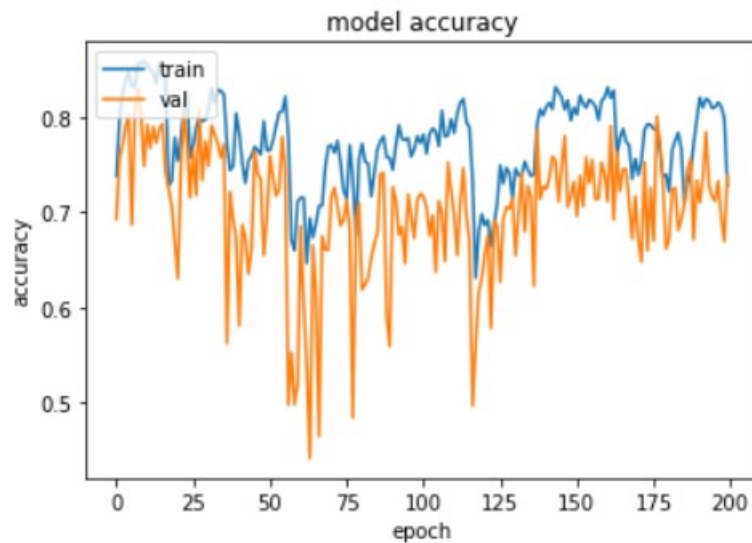


Figura 89: Gráfica de entrenamiento del modelo

Respecto a los datos que hemos obtenido con el uso de los MFCC en este caso de estudio, se ha podido observar que los datos han sido favorables, pero no han sido los esperados, y aun así el algoritmo no ha llegado al caso de sobreentrenamiento.

Como se puede observar en la calidad del algoritmo, se ha observado que es capaz de clasificar con relativa facilidad los casos en los que hay barcos, ya sean pequeños o grandes. Sin embargo, en caso en que no haya barcos, ha tenido una precisión muy pequeña. Observando la matriz de confusión, se ha podido observar que esto se debe a que, en su gran mayoría, los datos de entrada correspondientes a la clase no hay barco, ha sido confundida con la clase de barcos pequeños.

Esto se ha hecho más notorio debido a que se han podido dar dos circunstancias a la vez.

- Algunos tipos de embarcaciones pequeñas pueden verse confundidas con los sonidos ambientales debido a su baja frecuencia.
- El caso de que no haya barcos es el caso más desfavorecido de los tres por lo tanto en caso de duda el algoritmo tenderá a una de las otras dos opciones y no acertará.

Se ha seguido observando la matriz de confusión y como era de esperar, en aspectos generales, los casos de hay y no hay barco se han confundido entre sí.

Este estudio ha conseguido obtener unos valores de acierto medio del 73%, siendo relativamente bajos teniendo en cuenta que se trata de una red CNN, ya que los valores obtenidos han sido más bajos que los obtenidos por la red de clasificación de árboles de decisión.

5.2.2.2 Resultados obtenidos para los descriptores GTCC

Accuracy: 89%	Precisión	Recall	F1-score
Barco grande	91%	83%	87%
Barco pequeño	86%	92%	89%
No barco	97%	97%	97%

Tabla 27: Tabla de resultados de descriptores GTCC

Matriz de confusión.

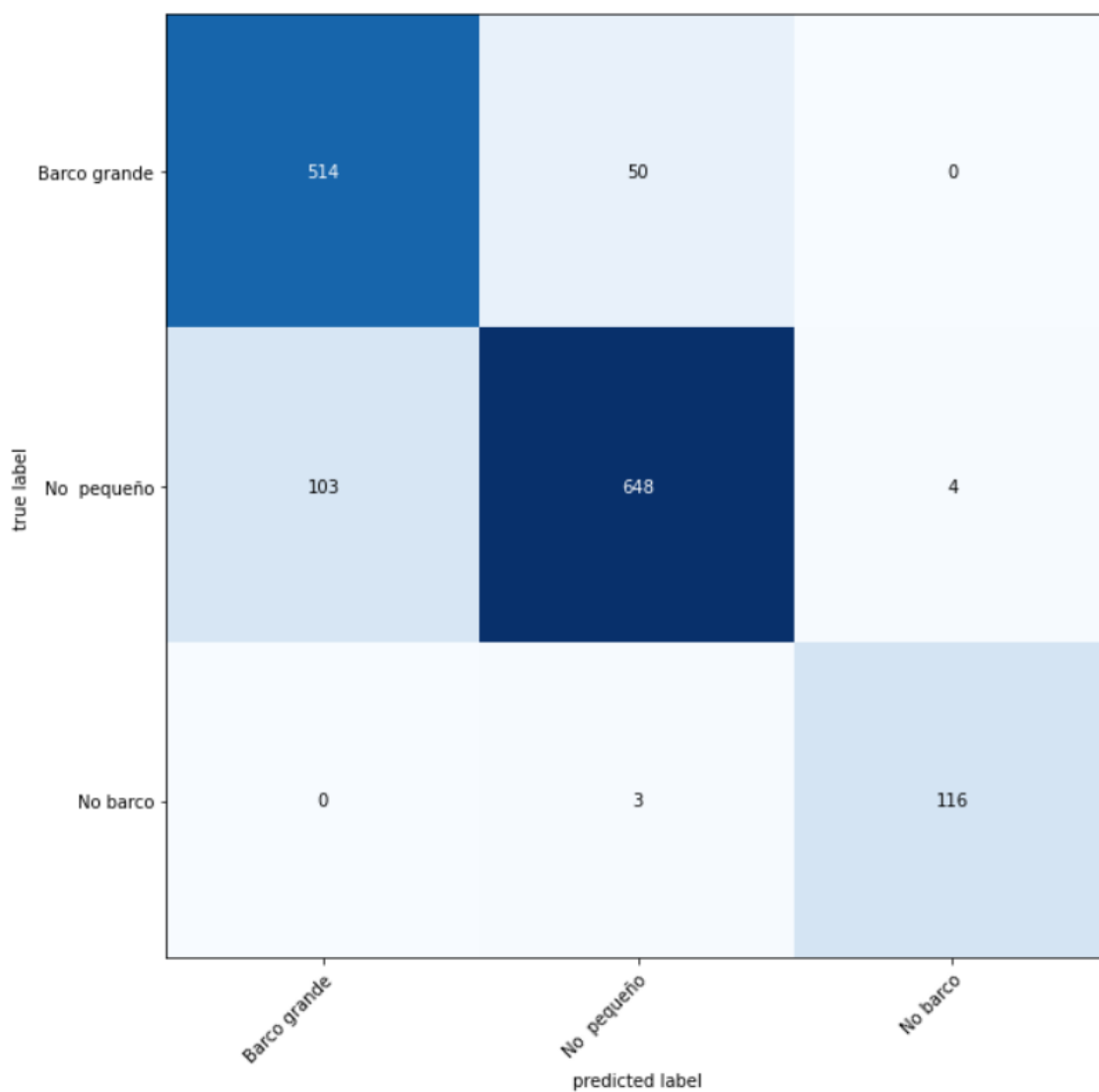


Figura 90: Matriz de confusión para caso de estudio 2, descriptores GTCC

Gráfica de entrenamiento del modelo.

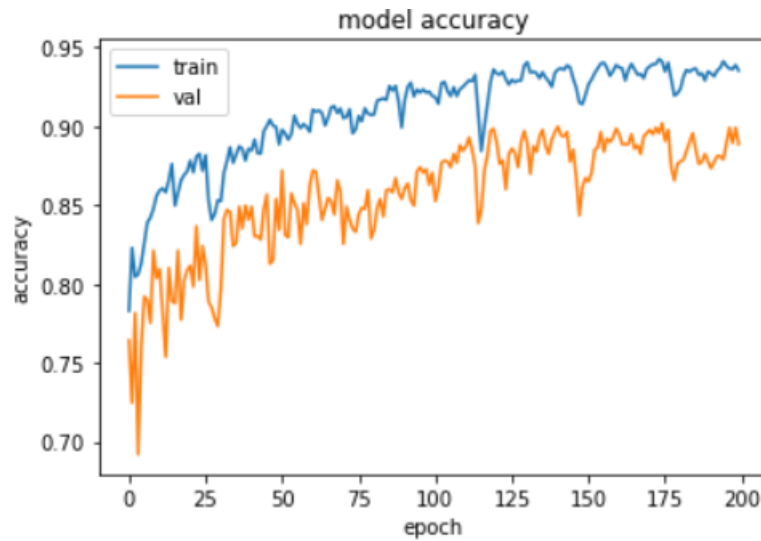


Figura 91: Gráfica de entrenamiento del modelo

Si se observa la gráfica de entrenamiento del modelo, vemos que no hemos llegado todavía al caso de sobre entrenamiento. Esto es importante a tener en cuenta a la hora de valorar el algoritmo.

Con el uso de los descriptores GTCC, se ha podido observar como la calidad del algoritmo ha mejorado bastante, obteniendo valores del 91% para embarcaciones grande, 86 para embarcaciones pequeñas y de un 97% para el caso de que no haya barcos. Esto hace del algoritmo el más fiable a la hora de clasificar en este caso de estudio,

Algo para tener en cuenta en el uso de los descriptores GTCC en este caso, es el elevado rendimiento que tienen clasificando la clase que corresponde a que no haya barcos, la cual es la más elevada con respecto las demás aun siendo la que menos datos de entrada tiene.

Observando la matriz de confusión, hemos reconocido lo que es el patrón que ha estado siguiendo este caso de estudio, el cual consiste en que los barcos tanto grande como pequeños se han equivocado entre ellos mismos, y en el caso de que no haya barcos, por lo general ha seguido dándose el caso de que haya sido confundido con el caso de que sean barcos pequeños.

Aun con todo este caso de estudio ha podido considerarse como satisfactorio ya que tiene una media de acierto del 89%, la más elevada para el caso de estudio 2 que se ha obtenido.

5.2.2.3 Comparación de resultados para el caso de estudio 2

MFCC	Accuracy 73%	→	Barco grande	Barco pequeño	No hay barco
Precisión			75%	78%	55%
Recall			80%	67%	79%
F1-Score			78%	72%	65%
GTCC	Accuracy 89%	→			
Precisión			91%	86%	97%
Recall			83%	92%	97%
F1-Score			87%	89%	97%

Tabla 28: Tabla de comparación de resultados de ambos descriptores

En comparación, sin duda los mejores descriptores frecuenciales a la hora de clasificar este caso de estudio han sido los descriptores GTCC.

Se puede observar que gran parte de que así haya sido se debe a que los descriptores GTCC, han sido capaces de clasificar con una mayor facilidad y precisión, el caso en el que no hay barcos, donde han sido capaces de clasificar con un 22% más de precisión.

Aun así, tanto a la hora de clasificar tanto barcos grandes como pequeños, los descriptores GTCC han podido realizar dicha tarea con una mayor calidad. Teniendo el algoritmo una mayor calidad para este caso.

Se ha vuelto a dar con que para la red CNN, en el caso de los coeficientes MFCC se ha obtenido el peor dato medio de clasificación mientras que con el uso de los GTCC, hemos obtenido los mejores valores de clasificación con respecto a este caso de estudio.

5.2.3 Caso de estudio 3. Caso multisalida. Todos los barcos

5.2.3.1 Resultados obtenidos para los descriptores MFCC

Métricas de precisión, recall y F1 score.

Accuracy: 88%	Precisión	Recall	F1-score
Dredger	96%	68%	80%
Fishboat	98%	80%	88%
Motorboat	81%	93%	87%
Musselboat	91%	86%	88%
Natural ambient	77%	100%	87%
Ocean Liner	83%	96%	90%
Passenger	87%	80%	83%
Pilotship	80%	61%	70%
RORO	92%	98%	95%
Sailboat	84%	76%	80%
Trawler	99%	90%	94%
Tugboat	84%	60%	70%

Tabla 29: Tabla de resultados de descriptores MFCC

Matriz de confusión.



Figura 92: Matriz de confusión para caso de estudio 3, descriptores MFCC

Gráfica de entrenamiento del modelo.

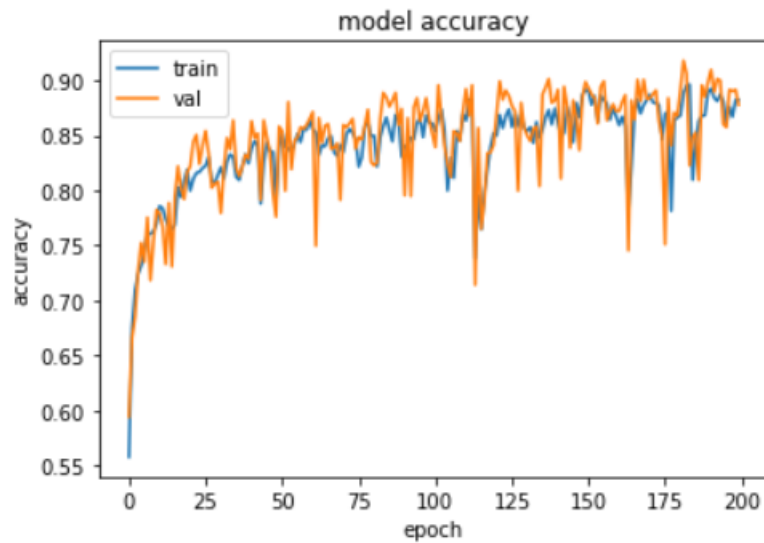


Figura 93: Gráfica de entrenamiento del modelo

Teniendo en cuenta el valor medio de acierto, se ha podido ver que estamos ante un caso de clasificación multisalida muy favorable, ya que ha sido capaz de clasificar 12 clases con una media de acierto del 88%.

Observando los datos de calidad del algoritmo se ha podido observar que obtenemos unos datos bastante homogéneos. Esto ha podido verse ya que la clase con menor precisión de clasificación ha sido la clase natural ambient (sonido ambiental) con un 77% mientras que la más alta ha sido Trawler (Yate) con un 99%.

Al haber obtenido valores de precisión tan altos, es normal que los valores de recall, también hayan sido elevados, como es el caso, ya que ha habido pocos casos en los cuales el algoritmo haya podido confundirse. Sin embargo, encontramos el caso de la clase Pilot ship (Barcos piloto) la cual tiene un porcentaje de recall más bajo con respecto a los demás. Esto se debe a que ha sido confundido una mayor cantidad de veces con relación a su porcentaje total de precisión.

Viendo la gráfica de entrenamiento del modelo, se puede observar que en este caso con un entrenamiento de 200 épocas, no hemos obtenido sobre entrenamiento y hemos sido capaces de obtener unos valores de acierto bastantes altos.

5.2.3.2 Resultados obtenidos para los descriptores GTCC

Métricas de precisión, recall y F1 score.

Accuracy: 80%	Precisión	Recall	F1-score
Dredger	87%	65%	74%
Fishboat	64%	27%	38%
Motorboat	74%	84%	79%
Musselboat	77%	96%	85%
Natural ambient	90%	95%	92%
Ocean Liner	74%	90%	80%
Passenger	77%	75%	76%
Pilotship	88%	63%	74%
RORO	80%	95%	87%
Sailboat	83%	31%	45%
Trawler	94%	86%	90%
Tugboat	88%	44%	59%

Tabla 30: Tabla de resultados de descriptores GTCC

Matriz de confusión

true label	Dredger	Fishboat	Motorboat	Musselboat	Natural ambient	Ocean liner	Passengers	Pilot ship	RORO	Sail boat	Trawler	Tugboat
Dredger	20	0	0	0	0	0	2	1	0	0	0	0
Fishboat	0	14	0	0	1	0	5	0	2	0	0	0
Motorboat	3	4	104	0	0	0	18	3	0	1	4	3
Musselboat	5	3	0	80	0	0	3	0	0	11	0	2
Natural ambient	0	2	0	0	114	0	7	1	0	3	0	0
Ocean liner	0	0	3	1	0	85	4	0	10	0	3	9
Passengers	3	9	5	2	5	2	215	7	6	8	1	15
Pilot ship	0	0	1	0	0	0	3	28	0	0	0	0
RORO	0	19	10	0	0	4	30	2	355	9	8	7
Sail boat	0	0	1	0	0	0	0	2	0	15	0	0
Trawler	0	2	0	0	0	2	0	0	1	1	97	0
Tugboat	0	0	0	0	0	2	1	0	0	1	0	28

Figura 94: Matriz de confusión para caso de estudio 3, descriptores GTCC

Gráfica de entrenamiento del modelo.

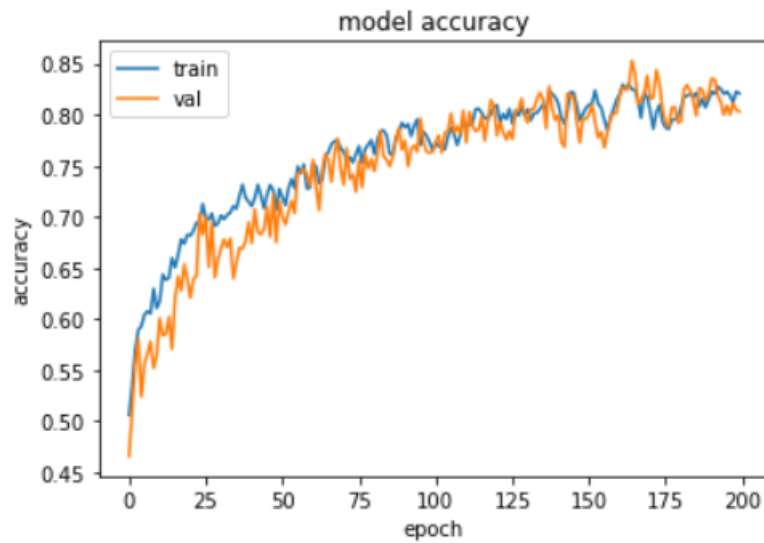


Figura 95: Gráfica de entrenamiento del modelo

Observando los datos obtenidos, se ha podido afirmar que nos encontramos también ante un caso favorable de clasificación multisalida, ya que se ha conseguido obtener unos valores del 80% de acierto medio, lo cual para un caso de clasificación de 12 clases ha sido un resultado relativamente bueno.

Fijándonos en la calidad del algoritmo se puede observar que se han vuelto a obtener unos datos de precisión homogéneos que varían entorno al 74% y 94%, menos el caso de la clase fishboat (barcos pesqueros) que ha obtenido una precisión del 64%.

En esta clasificación podemos observar gracias al recall, que las clases con las que más han sido confundidas las demás clases, con respecto a su precisión han sido las clases sailboat (barco velero) y fishboat (barco pesquero), las cuales tienen unos valores de recall del 31 y del 27%, aunque estos porcentajes no han sido significativos para la caída de la precisión en otras clases, ya que los datos de entrada de estas clases han sido una menor cantidad con respecto a las demás clases, como puede observarse en la matriz de confusión.

Con respecto a si ha habido sobreentrenamiento o no, se ha podido observar que en este caso tampoco se produce sobreentrenamiento ya que los datos de validación no están disminuyendo después de 200 etapas de entrenamiento.

5.2.3.3 Comparación de resultados para el caso de estudio 3

MFCC	Accuracy 88%	→	Precisión	Recall	F1-score
Dredger – dragador			96%	68%	80%
Fishboat – pesquero			98%	80%	88%
Motorboat – lancha motora			81%	93%	87%
Musselboat – mejillonera			91%	86%	88%
Natural ambient – ruido ambiental			77%	100%	87%
Ocean Liner – Transatlántico			83%	96%	90%
Passenger – barco de pasajeros, ferry			87%	80%	83%
Pilotship – barco piloto			80%	61%	70%
RORO-mercancías			92%	98%	95%
Sailboat – barco velero			84%	76%	80%
Trawler – yate			99%	90%	94%
Tugboat – remolcador			84%	60%	70%
GTCC	Accuracy 80%	→	Precisión	Recall	F1-score
Dredger – dragador			87%	65%	74%
Fishboat – pesquero			64%	27%	38%
Motorboat – lancha motora			74%	84%	79%
Musselboat – mejillonera			77%	96%	85%
Natural ambient – ruido ambiental			90%	95%	92%
Ocean Liner – Transatlántico			74%	90%	80%
Passenger – barco de pasajeros, ferry			77%	75%	76%
Pilotship – barco piloto			88%	63%	74%
RORO-mercancías			80%	95%	87%

Sailboat – barco velero	83%	31%	45%
Trawler – yate	94%	86%	90%
Tugboat – remolcador	88%	44%	59%

Tabla 31: Tabla de comparación de resultados de ambos descriptores

Comparando el uso de los descriptores MFCC y GTCC para el caso de estudio 3, el cual consiste en la clasificación general de las clases, se ha podido ver que ambos tienen unos resultados satisfactorios. Ya que obtenemos porcentajes de aciertos altos y una cierta homogeneidad en los resultados.

Aun así, se ha podido ver que el uso de descriptores frecuenciales MFCC ha sido más efectivo para este tipo de clasificación multisalida, ya que se han obtenido valores más elevados para este algoritmo.

De todas maneras, ambos descriptores han superado con los valores obtenidos en el mismo caso de estudio por la red de árboles de decisión.

6.Conclusiones.

En este TFG se ha abarcado el uso de dos descriptores frecuenciales, los MFCC y los GTCC, el uso de estos descriptores se ha debido a que los MFCC pueden considerarse como los más “conocidos”, debido a sus múltiples aplicaciones en el uso del habla y los GFCC ya han sido usados con anterioridad para el estudio de embarcaciones, tal y como el caso que abarca el TFG.

Se ha dividido el estudio de dichos descriptores en tres casos, el caso binario para cuando detectar si hay embarcaciones o no, el caso de estudio 2 para diferenciar embarcaciones grandes y pequeñas además de si hay o no embarcaciones, y el caso de estudio 3 para poder clasificar las embarcaciones en sus diferentes tipos. Además, se han realizado estos dos casos de estudio tanto para una red simple, la red basada en árboles de decisión, como para una red compleja, la cual ha sido una red neuronal convolucional. Se han elegido estos tres casos de estudio, para poder ver el comportamiento del algoritmo de entrenamiento a medida que se iba aumentando la dificultad dentro de la clasificación.

Los resultados obtenidos con respecto a estos experimentos tanto en el uso de la red formada por árboles de decisión como para el uso de la red neuronal convolucional han sido los siguientes:

Para la red de árboles de decisión se han obtenido los siguientes valores medios de acierto para los descriptores MFCC y GTCC: para el caso binario se ha obtenido un 97% y 98%, respectivamente; para el caso de estudio 2 se ha obtenido un 85% y 85%, respectivamente y para el caso de estudio 3 se ha obtenido un 65% y 69%, respectivamente.

Para la red neuronal convolucional se han obtenido los siguientes valores medios de acierto para el uso de los descriptores MFCC y GTCC: para el caso binario se ha obtenido un 96% y 99%, respectivamente; para el caso de estudio 2 un 73% y 89% respectivamente y para el caso de estudio 3 un 88% y 80%, respectivamente.

Tras haber analizado los datos obtenidos se ha podido llegar a la conclusión de que tanto los descriptores MFCC como los GTCC han sido validos para poder realizar una identificación y posterior clasificación de las embarcaciones.

Analizando los datos de cada modelo por separado, es decir del uso árboles de decisión y del uso de la red neuronal convolucional, se ha llegado a las siguientes conclusiones.

En el caso del uso de árboles de decisión, se ha podido observar que son eficaces a la hora de poder realizar la clasificación en el caso de que haya o no haya barco y también en el caso de diferenciar que no haya barcos o que los barcos sean grandes o pequeños. Sin embargo, para el uso de una clasificación más específica como es la clasificación de las embarcaciones en las diferentes doce clases, se puede ver que este modelo resulta ineficiente. Además, usando este modelo, se puede afirmar que el uso de los descriptores GTCC es un poco más eficiente que el uso de los descriptores MFCC. Esto puede ser debido a que los MFCC usa filtros más restrictivos al estar caracterizados para el uso del habla.

En el caso de la red neuronal convolucional, se ha observado que son eficientes a la hora de realizar la clasificación para cada uno de los casos de estudios, clasificación de si hay o no embarcación, clasificación de si la embarcación es grande, pequeña o no hay y para

la clasificación de las embarcaciones en cada una de las doce clases de embarcaciones que encontramos. Además, teniendo en cuenta los datos que se han obtenido se puede afirmar que el uso de los descriptores GTCC es más efectivo en los dos primeros casos, pero para la clasificación de las embarcaciones en cada una de las diferentes clases de embarcación resulta más eficiente el uso de los descriptores MFCC.

Ahora, teniendo en cuenta los dos modelos de aprendizaje automático propuestos para los tres casos de estudio, se ha podido afirmar que en el caso de clasificación si hay o no embarcación y el caso de clasificación de embarcaciones grandes, pequeñas o no hay embarcación, ambos modelos son generalmente igual de eficientes, habiendo obtenido unos valores parecidos. Sin embargo, para el caso de clasificación de las embarcaciones en los diferentes tipos de barcos, se nota una gran mejoría en la red neuronal convolucional con respecto a los datos obtenidos por el uso de árboles de decisión. Este se ha debido a que las redes neuronales convolucionales trabajan con imágenes en vez de con vectores, por lo tanto, ha sido más eficiente a la hora de realizar dicha clasificación.

Para las líneas futuras de estudio se podría, ampliando la base datos para obtener un equilibrio de las clases, realizar un estudio más completo de dichas clasificaciones para poder ponerlo a prueba en un escenario real, como por el ejemplo podría utilizarse para el uso de monitorización en tiempo real del tipo de embarcaciones que se encuentran entrando y saliendo de un puerto únicamente con el uso de un algoritmo de clasificación más preciso, y equipo capaz procesar las señales acústicas obtenidas por un micrófono submarino.

También se podría realizar este tipo de aprendizaje maquina con un mayor número de descriptores frecuenciales, ya que en este TFG solo se han usado los MFCC y los GTCC, habiendo un mayor número de descriptores con los cuales se podrían procesar estas señales acústicas.

7. Bibliografía

1. scikit-learn. 2022. 1. *Supervised learning*. [online] Available at: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
2. Rincón Llorente, C., 2022. [online] Lorien.die.upm.es. Available at: <http://lorien.die.upm.es/barra/pfcs/2007-carmenr/docs/proyecto.pdf>
3. Ortiz Ibarra, J., 2019. *Coeficientes Cepstrales de las Frecuencias de Mel*. [online] 1library.co. Available at: <https://1library.co/article/coeficientes-cepstrales-de-las-frecuencias-de-mel.6qm1558q>
4. Martínez Mascorro, G. and Aguilar Torres, G., 2013. Reconocimiento de voz basado en MFCC, SBC y Espectrogramas. *Ingenius*, (10).
5. Librosa.org. 2022. *librosa.feature.mfcc — librosa 0.10.0.dev0 documentation*. [online] Available at: <https://librosa.org/doc/main/generated/librosa.feature.mfcc.html>
6. Valero, X. and Alías, F., 2012. ANÁLISIS DE LA SEÑAL ACÚSTICA MEDIANTE COEFICIENTES CEPSTRALES BIO-INSPIRADOS Y SU APLICACIÓN AL RECONOCIMIENTO DE PAISAJES SONOROS. [online] Sea-acustica.es. Available at: <http://www.sea-acustica.es/fileadmin/publicaciones/227.pdf>
7. Spafe.readthedocs.io. 2019. *spafe.features.gfcc — spafe documentation*. [online] Available at: <https://spafe.readthedocs.io/en/latest/features/gfcc.html>
8. *Ciencia y Datos*, 2019. Aprendizaje No Supervisado en Machine Learning: Agrupación. [online] Available at: <https://medium.com/datos-y-ciencia/aprendizaje-no-supervisado-en-machine-learning-agrupaci%C3%B3n-bb8f25813edc>
9. García, L., 2020. *Aprendizaje por refuerzos; ¿Qué es? ¿Cómo se usa?*. [online] Kabel. Available at: <https://www.kabel.es/aprendizaje-refuerzos/>
10. Ariza-López, F., Rodríguez-Avi, J. and Alba-Fernández, V., 2018. CONTROL ESTRICTO DE MATRICES DE CONFUSIÓN POR MEDIO DE DISTRIBUCIONES MULTINOMIALES. *GeoFocus Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, pp.215-226.
11. scikit-learn. 2022. 1.10. *Decision Trees*. [online] Available at: <https://scikit-learn.org/stable/modules/tree.html>
12. Kortas, M., 2020. *Sound-Based Bird Classification*. [online] Medium. Available at: <https://towardsdatascience.com/sound-based-bird-classification-965d0ecacb2b>
13. Santos-Domínguez, D., Torres-Guijarro, S., Cardenal-López, A. and Pena-Gimenez, A., 2016. ShipsEar: An underwater vessel noise database. *Applied Acoustics*, 113, pp.64-69.
14. Reis, C., Padovese, L. and Oliveira, M., 2019. Automatic detection of vessel signatures in audio recordings with spectral amplitude variation signature. *Methods in Ecology and Evolution*, 10(9), pp.1501-1516.
15. Rizwan, M., 2018. *LeNet-5-A Classic CNN Architecture*. [online] Data Science Central. Available at: <https://www.datasciencecentral.com/lenet-5-a-classic-cnn-architecture/> [Accessed 31 August 2022].
16. Team, K., n.d. *Keras: the Python deep learning API*. [online] Keras.io. Available at: <https://keras.io>
17. Bahit, E., 2012. *Curso Python para principiantes*. Buenos Aires: eugeniabahit.com.
18. TensorFlow. 2022. *Guide | TensorFlow Core*. [online] Available at: <https://www.tensorflow.org/guide>

19. Vannieuwenhuyze, A., 2020. *Inteligencia artificial fácil*. Cornellà de Llobregat (Barcelona): ENI.
20. Chollet, F., n.d. *Deep learning with Python*. 1st ed.