MÁSTER EN INGENIERÍA DE AUTOMATIZACIÓN E INFORMÁTICA INDUSTRIAL

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA
DE COMPUTADORES Y SISTEMAS

Universidad de Oviedo

DETECCIÓN, CLASIFICACIÓN E INTERPRETACIÓN DE PATOLOGÍAS DE LA CONDUCCIÓN CARDIACA MEDIANTE EL USO DE TÉCNICAS DE DEEP LEARNING

JORGE BLANCO PRIETO

TUTOR DE EMPRESA: ÓSCAR JESÚS COSIDO COBOS

TUTOR ACADÉMICO: IGNACIO DÍAZ BLANCO

COTUTORA ACADÉMICA: ANA GONZÁLEZ MUÑIZ

FEBRERO 2021

RESUMEN

En el presente documento se exhibe un estudio sobre la *detección, clasificación automática e interpretabilidad* de patologías de la conducción cardiaca, que engloba patologías que afectan propiamente a los haces de conducción eléctrica cardiaca y patologías parenquimatosas, pudiendo dar ambas cuadros de arritmia cardiaca, mediante el uso de redes neuronales convolucionales de una dimensión (CNN1D). Para el estudio se utilizó un conjunto de datos compuesto por diversos electrocardiogramas (ECGs) con múltiples patologías. El objetivo es obtener una clasificación automática y aplicar algoritmos de interpretabilidad que permitan al personal clínico especializado confiar en los modelos de aprendizaje profundo, en ocasiones denominados con el término de *cajas negras*, haciendo referencia al desconocimiento o falta de transparencia en su proceso interno. En la búsqueda de la optimización del proceso, se realizaron varios análisis donde se debate entre el uso de técnicas de extracción de características, específicamente técnicas de segmentación de latidos, en función de los resultados que arroja el modelo:

1. El primer análisis está basado en la aplicación del modelo CNN1D sin una etapa previa de extracción de características. En sus resultados se obtuvo una precisión de la red en la clasificación de patologías por parte de los datos de entrenamiento y de test del 96% y 90% respectivamente; sin embargo, los métodos de interpretabilidad exponen que la red no extrae las características de la señal acordes a los requisitos de interpretabilidad de los especialistas sanitarios. Debido a esto, se propone realizar una etapa de extracción de características previa, enfocando el modelo a aprender sobre latidos cardiacos.

2. En el segundo análisis realizado se utilizan técnicas de extracción de características previas para ayudar a la red convolucional en su entrenamiento. Se optimizó la red y se utilizó un modelo CNN1D multi-entrada. En este caso la precisión en datos de entrenamiento y test bajó a un 70% y 68% respectivamente. La matriz de confusión expone una mala detección de varias patologías entre las que se encuentran VEB y PAC, patologías definidas por latidos ectópicos que se llegan a ver enmascaradas en las señales electrocardiográficas por la técnica de extracción de características.

3. En el tercer análisis, se suprimieron las patologías que pudieran afectar a la red neuronal, PAC y VEB, y, con la misma implementación de la red que en el análisis previo, se obtuvieron unos resultados gratamente satisfactorios con una precisión de la red de 83% en entrenamiento y 82% en test. La matriz de confusión muestra ciertas características de algunas patologías que son conocidas en el procesamiento digital de señal previo.

Posteriormente se aplican las técnicas de interpretabilidad mencionadas en base a la evaluación de la interpretabilidad definida en la metodología con el fin de proyectar luz sobre la explicación dada de las cajas negras de los modelos de deep learning y ayudar en la aceptación de estas técnicas, una aceptación que permitirá introducir su uso en entornos de alta responsabilidad como es el ámbito sanitario.

AGRADECIMIENTOS

No querría concluir este estudio sin redactar un mensaje de agradecimiento a todas las personas que han estado ayudándome para el desarrollo del trabajo. Especial atención a Ignacio Díaz por mostrarme el mundo existente en el campo de la analítica de datos y ver como desde la ingeniería se puede aportar conocimiento en un mundo totalmente distinto como es el campo de la medicina; a Ana, Diego y al Podcast DL, que ha sido un apoyo imprescindible para motivarme y poder sacar el trabajo adelante en los momentos más difíciles. Hacer una mención especial a Óscar, que me ha dado la oportunidad y motivación de poder seguir dedicándome en un trabajo apasionante vinculado con el sector sanitario en Virtual Intelligence S.L y que ha sabido guiarme dentro del trabajo realizado. Y sobre todo a mi pareja médica, Sara, una persona imprescindible para adentrarme en el mundo del análisis de señales de electrocardiograma que ha tenido que aguantarme día sí, día también, para explicarme todas las patologías y el sistema de conducción cardiaco, sin lo cual, este trabajo no habría sido posible.

LISTA DE FIGURAS

1. INTRODUCCIÓN

En el siglo XIX se evidenció la existencia de actividad bioeléctrica correspondiente a los latidos cardiacos. Las primeras pruebas consistieron en colocar alambres conductores a las muñecas de pacientes febriles con el fin de obtener un registro de los latidos del corazón. Sin embargo, no fue hasta principios del siglo XX cuando Willem Einthoven, Premio Nobel de Medicina y Fisiología en 1924, descubrió el galvanómetro de cuerda, instrumento de medición y detección de corrientes eléctricas, que permitió asignar las letras PQRST-U a las diferentes deflexiones y describir las características electrocardiográficas de gran número de enfermedades cardiovasculares. La medición del conjunto de deflexiones, así como la representación de la actividad bioeléctrica obtenida de la monitorización del corazón, es conocida como la prueba del electrocardiograma (ECG). En la actualidad, un siglo más tarde, el electrocardiógrafo, teniendo en cuenta sus innumerables mejoras en comparación con las primeras versiones gracias al desarrollo de la tecnología, se sigue utilizando y se mantiene como uno de los instrumentos electrónicos más empleados en la medicina moderna demostrando así su valor médico al ser una prueba de fácil realización y no invasiva al paciente.

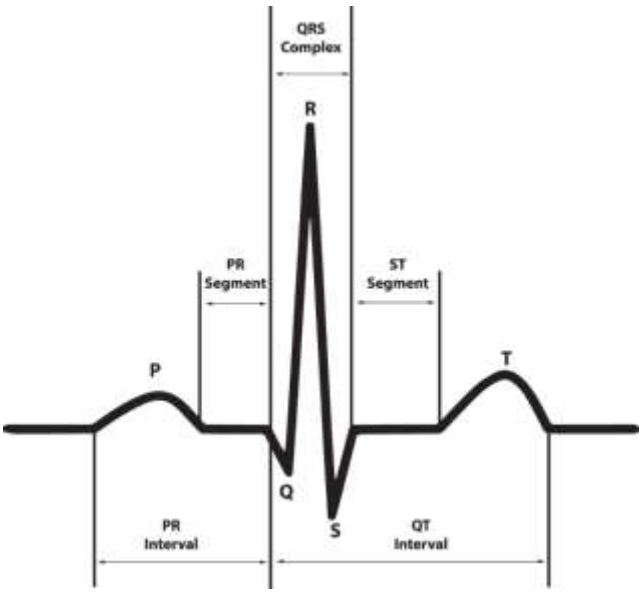

Figura 1.1. Ejemplo de un segmento PQRST que muestra los intervalos PR-QT, los segmentos PR - ST y el complejo QRS [1].

1.1. DESCRIPCIÓN DEL PROYECTO

En la actualidad, la prueba del electrocardiograma estándar de 12 derivaciones se compone por seis cables colocados en las extremidades (derivaciones proporcionaas: I, II, III, avR, avF, avL) y seis cables colocados en el pecho (derivaciones proporcionadas: V1, V2, V3, V4, V5 y V6) encargados de medir la actividad eléctrica de forma precisa desde diferentes ángulos. Este examen se ha utilizado ampliamente para diagnosticar una variedad de anormalidades cardiacas como arritmias cardiacas, alteraciones en el funcionamiento de la red eléctrica del corazón, y predice la morbilidad y mortalidad cardiovascular [1]. Las arritmias cardiacas (AC) son precursoras de enfermedades cardiovasculares y de la posible mortalidad asociada [2] . Tal es así que un estudio llevado a cabo por Statista, plataforma de gestión de

datos global, mostró las estadísticas de la principal causa de muerte en EE.UU en el año 2017, país asociado a la obesidad y consumo excesivo de comida rápida, y se observó que liderando el ranking se encontraban las enfermedades asociadas a patologías cardiacas. No es extraño encontrarse con estas estadísticas, otros países como México e incluso España siguen la misma tendencia.

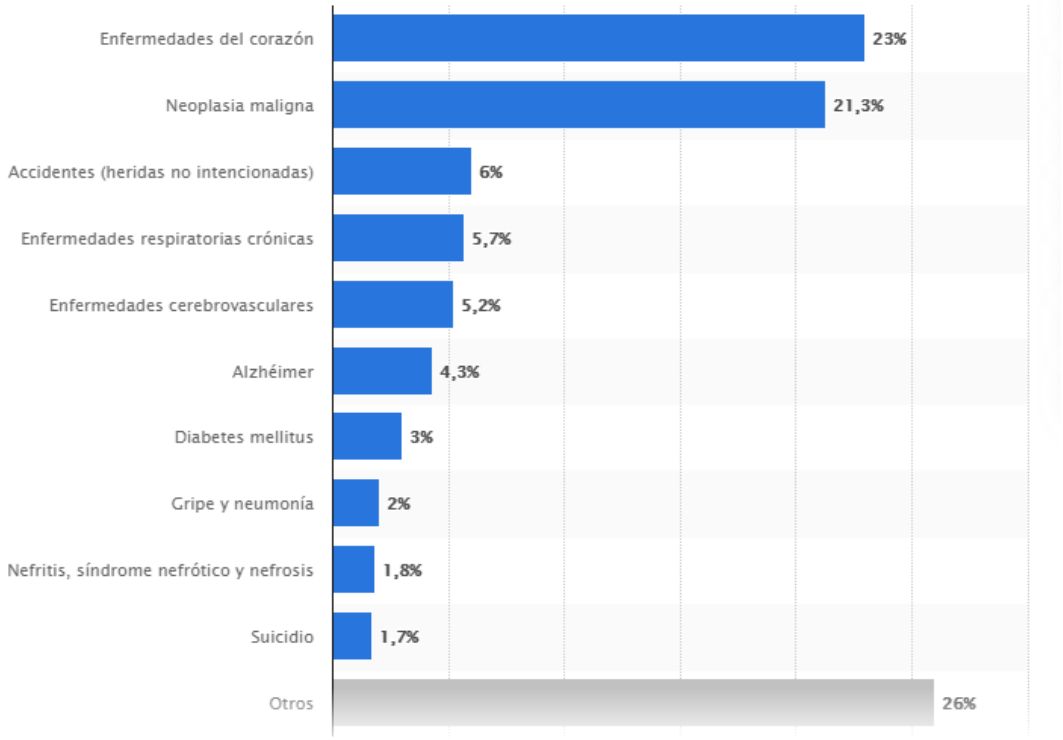

Figura 1.2. Estadísticas de la principal causa de muerte en EE.UU en 2017, obtenida de [3]

El diagnóstico temprano y correcto de las anomalías cardiacas, fundamental para la salud del paciente, emana de la observación de las señales ECG, prueba que aumenta las posibilidades de éxito de los tratamientos [4].

A pesar de que la prueba del electrocardiograma sea una de las pruebas más sencillas de realizar, la detección y clasificación manual del electrocardiograma es un trabajo arduo y requiere de personal cualificado con un alto grado de formación y que limita una gran parte de su tiempo en la interpretación del electrocardiograma. Debido a esto, la detección y clasificación automática de anormalidades cardiacas puede ayudar al personal sanitario en el diagnóstico del creciente número de señales ECGs.

1.2. PLANTEAMIENTO DEL ESTUDIO

Tradicionalmente, el análisis automático del electrocardiograma se ha basado en desplegar algoritmos de estudio de características en el procesamiento de estas señales, basadas en características estadísticas, características del dominio del tiempo y del dominio de frecuencia. Estos métodos son insuficientes debido a su limitación por la calidad de los datos y la limitación por el conocimiento de los expertos. Sin embargo, el número de datos de dominio público proporcionados correspondientes a señales ECG ha aumentado de forma considerable en los últimos años propiciando que el número de trabajos realizados utilizando métodos de aprendizaje profundo aumenten, de manera que se mejore el dominio sobre el problema y por tanto, se optimicen las técnicas utilizadas llegando a lograr prometedores resultados en muchas áreas de aplicación , siendo su ventaja la extracción automática de características de la señal [6].

No obstante, los avances obtenidos en la identificación de las patologías de las señales ECG mediante técnicas novedosas traen consigo nuevos desafíos y problemas, según los autores de [4], dentro del mismo campo: interpretabilidad, escalabilidad y eficiencia. En la definición de estos conceptos se encuentra que la escalabilidad corresponde con la propiedad o capacidad deseable de un sistema, red o proceso que indica su habilidad para reaccionar y adaptarse sin pérdida de calidad; la eficiencia corresponde con la capacidad de cumplir una determinada tarea de forma óptima y la interpretabilidad es la característica de dar o atribuir un significado al proceso, siendo este el caso que concierne al presente estudio, para poder permitir establecer una conexión entre los resultados obtenidos del proceso de aprendizaje automático con el dominio del personal sanitario especializado y proporcionar una mayor confianza en los modelos. Es debido a esto por lo que, plantear el uso de aprendizaje profundo en una investigación cuya finalidad sea la de obtener una clasificación de las señales de electrocardiograma dadas, según su determinada patología, resulta muy interesante y prometedor, pero aunar estas técnicas junto con el desarrollo de algoritmos que mejoren la escalabilidad y la eficiencia, a la par que se obtiene una interpretabilidad del modelo, aporta cierto realismo, relevancia e interés al trabajo desarrollado.

1.3. OBJETIVOS DEL PROYECTO

Teniendo en cuenta las directrices dadas en la línea de trabajo planteada, se presenta la siguiente investigación que tiene definido de forma clara y concisa la diferenciación entre los siguientes objetivos:

1. Realizar un estado del arte del conjunto de técnicas de procesamiento de señal que se llevan a cabo para la evaluación de las patologías cardiacas en señales de electrocardiograma, con el fin de realizar una extracción de características de las señales e identificar la patología perteneciente, teniendo en cuenta el campo de aplicación: entorno sanitario.

2. El objetivo principal es desarrollar un algoritmo inteligente que sea capaz de realizar una clasificación de una serie de patologías cardiacas estudiadas de manera automática a partir de las señales de electrocardiograma. Este proceso trata de integrar métodos tradicionales centrados en el procesamiento digital de señal aplicados a señales no estacionarias junto con técnicas modernas de aprendizaje profundo, deep learning. Los métodos tradicionales se utilizarán para realizar un preprocesamiento de los datos y prepararlos de forma óptima para el entrenamiento de la red. La arquitectura de la red de deep learning se basará en la utilización de una red convolucional 1D y su estudio radica en obtener una arquitectura de red que optimice

el resultado del problema. Obtener la mayor tasa de acierto es necesario teniendo en cuenta la alta responsabilidad que exige este campo.

3. Aplicar técnicas de interpretabilidad del modelo con el fin de dar una mayor credibilidad a la existencia de sistemas inteligentes en el mundo de la ingeniería de la inteligencia artificial aplicada a campos como la medicina. Así mismo se realizará un escueto estudio de evaluación sobre la interpretabilidad que muestra el modelo con el objetivo de arrojar luz sobre la interpretabilidad de la persona en referencia con el sistema inteligente.

El trabajo desarrollado trata de proporcionar una alternativa a los métodos convencionales en el diagnóstico de patologías cardiacas y añade técnicas de interpretabilidad con el fin de aumentar la aceptación de estas técnicas en un entorno sanitario real y, desde el punto de vista sanitario.

2. MÉTODOS Y TÉCNICAS

2.1. DATOS

Las señales electrofisiológicas, en este caso asociadas al sistema cardiovascular, nacen en el fenómeno denominado potencial de acción originado en las membranas celulares de los tejidos por células auto rítmicas. Estas células que componen estas membranas poseen en su interior iones de sodio y potasio, cuya concentración es regulada por canales que controlan su flujo de entrada y de salida. Los cambios existentes en dicha concentración causan despolarización y polarización, las cuales son el principio de los movimientos musculares. La despolarización se difunde a través de unas uniones de hendidura intermedias hacia las células contráctiles cardiacas, tal y como se puede observar en la siguiente figura.

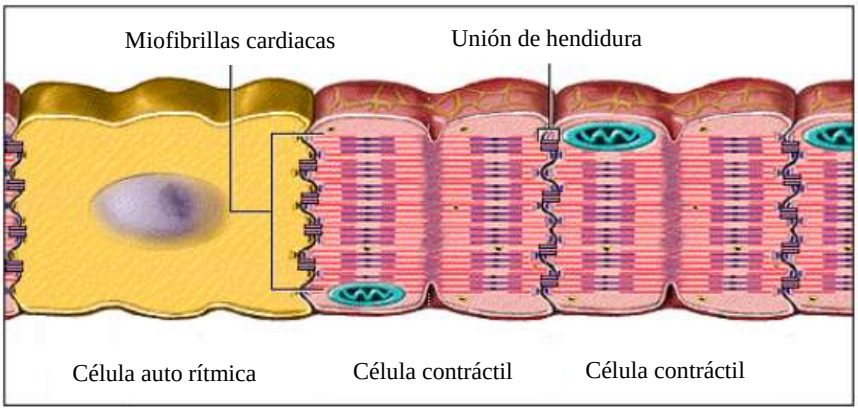

Figura 2.1. Esquema de polarización y despolarización del tejido del cuerpo humano [17].

La diferencia de potencial generada permite conocer, mediante la colocación de electrodos, el comportamiento eléctrico del sistema fisiológico cardiovascular [17]. Sin embargo, la contracción de los músculos, al igual que la contracción del corazón, contribuyen al aumento del ruido en las señales cardiacas. Cuando los músculos en la vecindad de los electrodos se contraen, se generan ondas de despolarización y repolarización que son recogidas por el electrocardiograma. La gravedad de la diafonía depende de la cantidad de contracción muscular del sujeto y de la calidad de las sondas. Además de las distorsiones causadas por los denominados artefactos fisiológicos, como los derivados del sistema de conducción eléctrico del corazón o la respiración del paciente en la realización de las pruebas, hay otro tipo de distorsiones originadas por los denominados artefactos artificiales, perteneciendo ambos tipos de artefactos a la banda de altas frecuencias. Los artefactos artificiales son aquellos que se originan debido a los dispositivos eléctricos y electrónicos presentes en la etapa de adquisición de la señal, siendo uno de los más frecuentes, el sistema eléctrico de alimentación.

Debido a esto, las técnicas de preprocesamiento de señales adquieren gran importancia como paso previo a la extracción de características realizada.

2.2. PREPROCESAMIENTO DE LOS DATOS

Entre las técnicas más importantes comúnmente realizadas en esta fase de preprocesamiento, se pueden destacar la reducción de la línea de interferencia base y la reducción de ruido propia de la señal.

Las señales ECG contienen diversas fuentes de ruidos, interferencias o desviaciones. Una fuente de ruido típica de estas señales es la denominada desviación de la línea de interferencia base, Figura 2.2, que, en ocasiones, se caracteriza por mostrar en las señales ECG una onda cuya frecuencia varía entre los 0.5 Hz y 0.6 Hz [13]. Esta interferencia en la lectura de la actividad eléctrica del corazón viene dada principalmente por la respiración del paciente, el movimiento del paciente durante la realización de la prueba y la interacción de los electrodos y la piel que provocan el desvío de la señal ECG. Además, en algunas ocasiones puede ser causada por variaciones en la temperatura y el sesgo de la instrumentación.

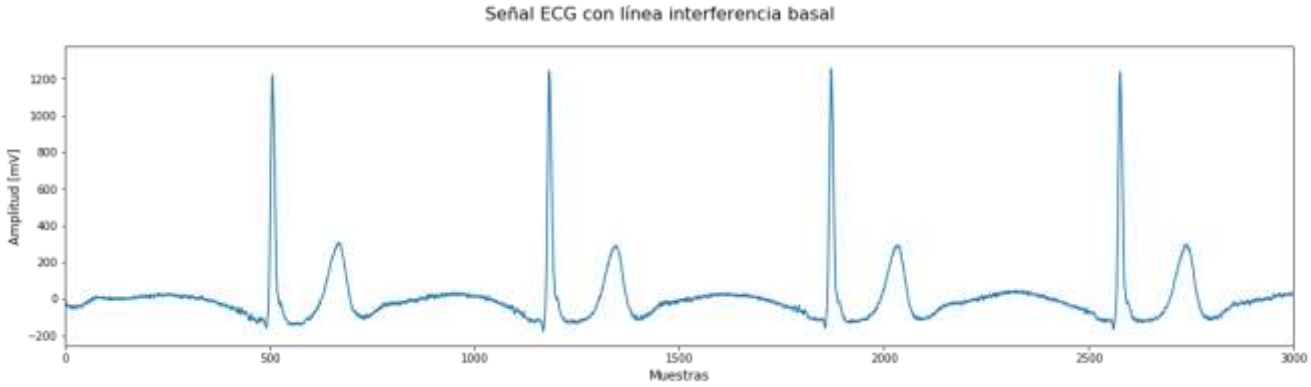

Figura 2.2. Señal de electrocardiograma con línea interferencia basal. Se puede observar un aspecto de onda entre los picos máximos R.

Por otro lado, se ha de impedir trabajar con señales electrofisiológicas en presencia de todo tipo de ruido, evitando que la efectividad de los métodos de extracción de características se pueda ver perjudicada por los latidos contaminados con ruido que alteren la morfología de la señal.

2.2.1. REDUCCIÓN LÍNEA DE INTERFERENCIA BASE

El procesamiento de señal dedicado a la eliminación de la línea de interferencia base de las señales ECG no es una tarea trivial. Esta desviación de la línea base puede enmascarar ciertas características importantes de la señal ECG que deriven en el dictamen de varias enfermedades cuyo centro de diagnosis radica en el segmento ST, como por ejemplo la elevación del segmento ST (STE), la disminución del segmento ST (STD), isquemia o infarto, entre otros; por lo tanto, es conveniente eliminar este ruido para un análisis adecuado de la señal. Varios estudios, como [6][15], se centran en la comparación de diferentes enfoques, desde el punto de vista del procesamiento digital de señal, para la supresión eficaz de la línea base.

En el proceso de aplicación de técnicas de filtrado de la línea de interferencia base en el entorno sanitario resulta interesante tener en consideración dos puntos, que servirán como base para la elección de la técnica a aplicar en el estudio: reducción de la línea interferencia base sin perjudicar la señal y realizar esta reducción con un coste computacional mínimo que garantice una operación en tiempo real estricto ya que, hay que tener en cuenta que, estas operaciones se ejecutan dentro de un ambiente sanitario.

Los enfoques más utilizados en la literatura para el análisis de la línea de interferencia base tienen en común la cancelación de las componentes de baja frecuencia de las señales ECG. El segmento ST está compuesto por componentes de baja frecuencia, entre 0.5 y 3 Hz, siendo esta la razón por la que este segmento se ve perjudicado. [6] realiza un estudio comparativo entre las técnicas de filtrado utilizadas regularmente en la literatura, *Filtro de Butterworth, Mediana móvil, spline approximation and subtraction, Cancelación línea basal basado en wavelets* y *Filtrado de paso alto basado en wavelet*.

El estudio se basa en el diseño y utilización de los filtros mencionados cuyo objetivo se centra en minimizar el proceso de supresión de las características de las señales ECG tras la aplicación de técnicas de filtrado. Sin embargo, el proceso planteado acarrea consigo un problema: es objetivamente complicado evaluar el rendimiento de las técnicas de filtrado sobre el segmento ST, segmento que une el pico S con el pico de la onda T (véase Fig. 1.1), al no existir señal libre de artefactos. Para resolver este problema, en [6] se generan señales ECG sintéticas, de manera que ningún artefacto altera la señal y todos los picos u ondas mantienen una anotación precisa en su eje de tiempo; no obstante, para evitar la generación de señales perfectas que se alejen de una situación realista, se recurre a una simulación de electrofisiología cardiaca que permite recrear patologías a un nivel multiescalar. Como conclusión en el artículo mencionado, se obtiene una tabla comparativa de las técnicas de filtrado aplicadas a las señales generadas que resume el rendimiento de los filtros en base a varios métodos: *correlation*, la correlación de la morfología de las señales ECG filtradas con las señales originales, *l_operator*, valor que indica el método que mejor restaura la señal ECG filtrada con respecto a la original, *KP deviation*, método por el cual el segmento ST ha resultado menos perjudicado y *computation time*, que se refiere al coste computacional que requiere el uso de la técnica de filtrado.

Tabla 1.- Resumen de los resultados obtenidos para la evaluación del rendimiento entre los filtros mencionados. Los valores se dan en $MED \pm IQR$ (Median & Interquartile range). Los valores de p *muestran el significado estadístico del filtro con mejor rendimiento de cada categoría (numeros en negrita) [6].*

Filter/indexes	Correlation	l_operator	KP deviation [mV]	Computation time [s]
No filter	0.7794 ± 0.3764	0.7797 ± 0.4452	0.0000 ± 0.2802	0 ± 0
Butterworth	0.9851 ± 0.0392	0.9859 ± 0.0372	0.0010 ± 0.0595	$\mathbf{0.0059 \pm 0.0434}$
Median	0.9904 ± 0.0299	0.9872 ± 0.0413	-0.0033 ± 0.967	1.8464 ± 0.0434
Spline	0.9813 ± 0.0486	0.9716 ± 0.0629	0.0037 ± 0.0628	0.0074 ± 0.0434
Wavelet cancellation	$\mathbf{0.9928 \pm 0.0194}$	$\mathbf{0.9933 \pm 0.0184}$	$\mathbf{0.0000 \pm 0.0419}$	0.3892 ± 0.0434
Wavelet high-pass	0.9672 ± 0.0827	0.9689 ± 0.0816	0.0000 ± 0.0885	0.4206 ± 0.0189
p values	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$	$< 10^{-6}$

En esta simulación se puede observar en la Tabla 1 que la técnica *wavelet-based baseline cancellation* ha logrado mayor efectividad que el resto de las técnicas en ¾ métodos de evaluación. Según indica [6], estos resultados revelan que esta técnica de filtrado es, no solo más precisa que el resto de las técnicas analizadas al tener su valor MED superior, sino más robusta (su IQR es mayor). La razón de su efectividad radica en las propiedades de la descomposición de las *wavelet* en varios niveles. Sin embargo, la mayor efectividad de las técnicas, además de resultar en técnicas precisas y robustas en todos los métodos de evaluación morfológicos, radica en adecuarse al ámbito de análisis, que no deja de ser otro que la aplicación en un ámbito médico. Como se puede observar, el filtro *Butterworth high-pass filter* obtiene unos resultados muy superiores al resto de técnicas en el método de evaluación *computation time [s]*, cumpliendo una ejecución en tiempo real ($< 0.001\ [s]$) y adecuándose a las restricciones, en tiempo, y responsabilidades, en precisión, de un entorno sanitario.

A pesar de esto, se ha observado que ninguno de los métodos planteados en el estado del arte es capaz de restablecer el segmento ST a la perfección pudiendo llegar a ocasionar errores en el diagnóstico de una patología cuyo centro de atención resida en un cambio en este segmento.

Tras realizar un estudio del estado del arte de las técnicas y realizar una comparativa de su rendimiento, [6] señala como interesante el estudio y aplicación del filtro *Butterworth high-pass filter* debido a su bajo coste computacional (0.0059 ± 0.0434 [s]) y sus valores altamente competitivos en los demás métodos estudiados con respecto a las técnicas de filtrado del estado del arte.

2.2.1.1. BUTTERWORTH HIGH-PASS FILTER

Los filtros de *Chebyshev* son filtros que se utilizan para separar dos bandas de frecuencias. Su diseño se basa en una estrategia matemática, *z-transform*, caracterizada por su rapidez a la hora de minimizar al máximo el rizado "ripple". La siguiente figura muestra tres casos de respuesta en frecuencia de un filtro paso bajo de *Chebyshev*. Se puede observar cómo, a medida que crecen las ondulaciones, el corte se vuelve más nítido [16].

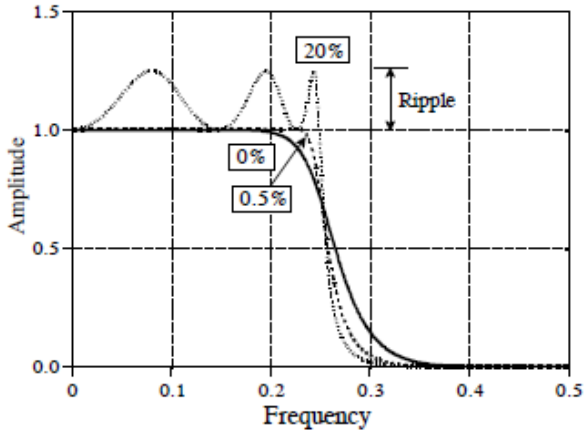

Figura 2.3. Respuesta del filtro Chebyshev. Cuando las ondulaciones cercanas al corte tienden a 0%, el filtro se denomina filtro de Butterworth. Se puede observar ondulaciones del 0%, del 0.5% y del 20% [16].

Cuando las ondulaciones previas al corte tienden a 0%, el filtro se denomina filtro de *Butterworth*. Por lo tanto, el filtro de *Butterworth* es un filtro electrónico diseñado para producir una frecuencia de amplitud monótona a una respuesta de frecuencia cero casi hasta la frecuencia de corte, donde empieza a disminuir a razón de $n * 20\ dB$, siendo n el orden del filtro, el cual, a menor número de orden, mejor rendimiento ofrece en dominio del tiempo y a un número de orden más alto, mejor rendimiento en dominio de la frecuencia.

Para el diseño de los filtros se han de tener en cuenta varios parámetros: tipo de respuesta (*high-pass* o *low-pass*), frecuencia de corte, porcentaje de ondulaciones en el corte y número de polos. Siguiendo las indicaciones mostradas en [16], Mahesh y Mahadev [17], diseñaron e implementaron un filtro *Butterworth* paso alto con la finalidad de suprimir la línea de interferencia de las señales ECG sin afectar a las características de la señal. Su diseño inicial se basa en un filtro de orden 3 con frecuencia de corte de 1000 Hz. La implementación del filtro se realiza sobre la derivación II de las señales ECG. En la siguiente figura se puede observar, a la izquierda, la ventana con la señal ECG original, mientras que, a la derecha, se observa la misma señal ECG tras haberla filtrado con el filtro *Butterworth* diseñado.

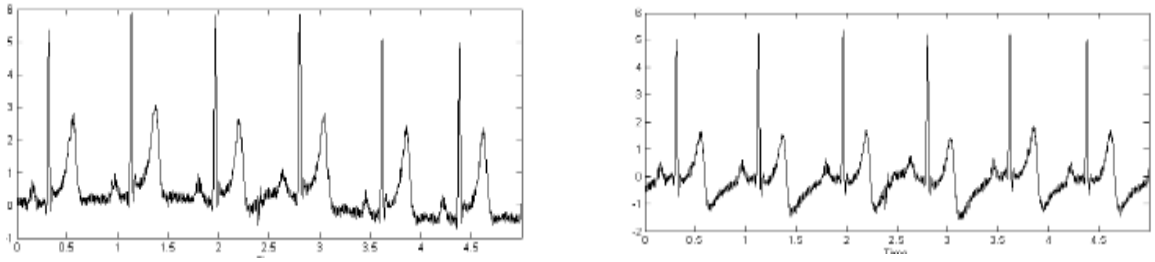

Figura 2.4. La gráfica de la izquierda representa la derivación II de una señal ECG sin procesar, mientras que la figura de la derecha muestra la misma ventana de la señal ECG tras haber aplicado el filtro de Butterworth diseñado en [17].

En el ejemplo expuesto en la figura anterior, se puede observar que la línea de interferencia base se ha suprimido con éxito y que, además, se han desenmascarado ciertas características que en la señal inicial no se podían observar (ondas negativas tras las ondas T en los instantes de tiempo 0.6, 1.5, 2.4, 3.1, 4 y 4.7 s).

2.2.2. REDUCCIÓN DE RUIDO – FILTRADO DE LA SEÑAL

El uso de métodos adaptativos y filtros digitales son la respuesta clásica para el procesamiento del ruido de las señales estacionarias. No obstante, las señales ECG son señales no estacionarias, es decir, sus características se ven modificadas a lo largo del tiempo. Debido a esto, este tipo de señales contienen una dificultad añadida para eliminar/suavizar el ruido sin perder información. En el procesamiento de las señales de electrocardiograma, Daqrouq, [8], utilizó el método *Discrete Wavelet Transform* (DWT) para la reducción de la desviación de la línea base del ECG, análisis recogido en la Tabla 1, mostrada en el apartado anterior. La transformación ondulatoria discreta tiene las propiedades que permiten una buena representación de la señal no estacionaria como la señal ECG y divide la señal en diferentes bandas de frecuencia. Posteriormente Zhang, [9], utilizó este método para la reducción del ruido de la señal. La subdivisión de la señal obtenida tras aplicar técnicas de DWT permite la detección seguida de la reducción del ruido en subseñales de baja frecuencia

Las técnicas de procesamiento digital de señal basadas en frecuencia (FBTs) han sido ampliamente utilizadas para el análisis de señales estacionarias. Por otro lado, las señales no estacionarias utilizan técnicas tiempo-frecuencia (TF) como por ejemplo *Short-Time Fourier Transform* (STFT), *Wavelet Transform* (WT), *Ambiguity Function* (AF) y *Wigner-Ville Distribution* (WVD).

STFT utiliza una transformada de Fourier estándar para varios tipos de ventanas en el procesamiento de la señal. Las técnicas basadas en *wavelets* aplican una *mother wavelet* con escalas discretas o continuas, *Discrete Wavelet Transform* y *Continuous Wavelet Transform* respectivamente, a una forma de onda determinada con el fin de resolver los problemas de resolución de tiempo-frecuencia inherentes al STFT. *Wavelet Transform* es una técnica ampliamente utilizada en aplicaciones computacionales debido a su bajo coste computacional. AF y WVD son representaciones cuadráticas en tiempo-frecuencia que usan técnicas muy avanzadas para combatir las dificultadas de resolución proporcionadas por STFT; no obstante, sufren por interferencias y producen resultados con mayor ruido granular[1] que la aplicación de otras técnicas como *Wavelet Transform*.

[1] Ruido granular es el resultado de la utilización de un escalón con una amplitud elevada frente a tramos de la señal con escasa variación de pendiente.

Dentro de las técnicas basadas en *wavelets*, se observa que, especialmente DWT, son técnicas ampliamente utilizadas para codificar y decodificar señales debido a su rapidez computacional, mientras que otro tipo de técnicas *wavelets*, como *Wavelet Packet Analysis* (WPA) o CWT, son técnicas cuyo éxito radica en el reconocimiento y en la extracción de características de las señales [19][20].

En aplicaciones, las técnicas FBTs son típicamente utilizadas en el análisis de ruido y vibraciones. Este tipo de técnicas proporcionan un promedio de la energía en el tiempo de un segmento de señal en dominio de la frecuencia, pero no permanece en el dominio del tiempo. Por otra parte, las técnicas *Wavelet Transform* permiten cambiar la composición del espectro de una señal no estacionaria para ser medidas y analizadas en el dominio de tiempo y frecuencia, definiendo una excelente herramienta para este tipo de señales. Particularmente, este tipo de técnicas son ampliamente utilizadas en aplicaciones biomédicas, ya sea para realizar una comprensión de las señales ECG y evitar así el coste computacional que requiere el análisis del gran volumen de datos que se obtiene, para reducción de ruido de las señales sin afectar la composición de las ondas e incluso para la extracción de características PQRST de las señales [20], entre otros. Para entrar más en detalle de las técnicas mencionadas de las *wavelet*, se introducen brevemente las técnicas CWT, DWT y WPA.

2.2.2.1. CONTINUOUS WAVELET TRANSFORM

Continuous Wavelet Trasnform (CWT) es una técnica de análisis en tiempo – frecuencia que permite arbitrariamente localizar y extraer en el tiempo las características de frecuencia de una señal. Esta técnica utiliza una ventana variante a lo largo de la señal relacionada con el factor de escala, constante que varía en un rango de valores predefinidos conexos con el punto de observación de la señal en función de su cálculo sobre la discretización en tiempo – frecuencia [19][20]. Esta discretización se basa en el cálculo de una integral que recorre el mapa discretizado por las escalas, a, y las posiciones en el eje del tiempo de dichas frecuencias, b. Para el análisis de CWT hay que tener en cuenta el tipo de *wavelet* a utilizar, que variará en función del tipo de señal que se analiza y de su aplicación. Por lo tanto, dada una función o señal, $x(t)$, y un tipo de la familia de las *wavelet*, ψ, la ecuación de *Continuos Wavelet Transform* se puede expresar como:

$$CWT_x(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \cdot \psi\left(\frac{t-b}{a}\right) dt \qquad (2.1)$$

Finalmente, tras aplicar CWT se extrae la información local en el dominio de frecuencia y tiempo que refleja las características de la señal analizada. La señal inicial puede ser reconstruida por la siguiente ecuación, denominada "*identidad de la resolución*", *a* través de la transformada resultante $CWT_x(a,b)$ obtenida anteriormente.

$$x(t) = \frac{1}{C_\psi} \cdot \int_0^{\infty} \int_{-\infty}^{\infty} W_x(a,b) \cdot \psi_{a,b}(t) \cdot \frac{1}{a^2}\, da\, db \qquad (2.2)$$

Las ventanas utilizadas por las familias de las *wavelet*, $\psi(t)$, están superpuestas unas con otras, derivando en una redundancia de la información en la técnica CWT. Esta es una desventaja a la hora de

comprimir la señal sin pérdida de información. Para reducir esta redundancia, *wavelet transform* puede ser calculada discretamente en el mapa de tiempo – frecuencia.

2.2.2.2. DISCRETE WAVELET TRANSFORM

Discrete Wavelet Transform (DWT) es una técnica de procesamiento que representa las características de la señal en un modo de multiresolución en tiempo y frecuencia. En su implementación, DWT utiliza un set de escalas discretas *wavelet* y descompone la señal inicial en un conjunto de ondas mutuamente ortogonales, evitando así la redundancia de información en el enventanado y siendo, por lo tanto, la principal diferencia con CWT [21]. La ecuación de la transformada wavelet es la siguiente:

$$WT_x(j,k) = \frac{1}{\sqrt{a_0^j}} \cdot \int x(t) \cdot \overline{\Psi\left(\frac{\mathrm{t} - \mathrm{k} \cdot \mathrm{a}_0^{\mathrm{j}} \cdot b_0}{a_0^j}\right)} dt = \langle x(t), \Psi_{jk}(t) \rangle \qquad (2.3)$$

Esta descomposición es lo que se denomina *Discrete Wavelet Transform*. Una vez realizada la descomposición de la señal, la recuperación de la señal inicial debe realizarse bajo la satisfacción de la siguiente condición, denominada *condición de estabilidad*, donde los parámetros A y B son constantes:

$$A \leq \sum_{j=-\infty}^{\infty} \left|\Psi\left(2^j \omega\right)\right|^2 \leq B \qquad (2.4)$$

Seguidamente, se establece una escala invariante de tiempo, es decir, únicamente se calcula la transformada en el plano de frecuencia, obteniendo los coeficientes de la siguiente ecuación, $d_j(k)$, tras aplicar la transformada wavelet a los distintos niveles, j:

$$d_j(k) = WT_x(j,k) \qquad (2.5)$$

Finalmente, para la recuperación de la señal inicial, $x(t) \in L^2(R)$, se utiliza el método *wavelet series* que computa el sumatorio de los coeficientes de las señales analizadas en los distintos niveles como resultante de la señal inicial:

$$x(t) = \sum_{j=0} \sum_{k=-\infty} d_{j(k)} \cdot \hat{\psi}_{jk}(t) \qquad (2.6)$$

Siendo en este caso $\hat{\psi}(t)$ la wavelet dual de $\psi(t)$ utilizada para la recuperación de la señal original. Esta wavelet dual tiene la misma escala y desviación que la original:

$$\Psi_{jb}(t) = \frac{1}{\sqrt{2^j}} \cdot \Psi \cdot \left(\frac{t-b}{2^j}\right) \qquad (2.7)$$

2.2.2.3. MULTIRESOLUTION ANALYSIS

Multiresolution Analysis (MRA) en el espacio cerrado $L^2(\mathbb{R})$ de una secuencia $\{V_j\}_j$, consiste en una secuencia de subespacios anidados que satisface ciertas relaciones de semejanza en el tiempo (espacio) y frecuencia (escala). Es un procedimiento de análisis de la señal que tiene en cuenta su representación a múltiples resoluciones [21]. La principal idea de esta técnica, descrita en las figuras 2.5 y 2.6, reside en la descomposición de una secuencia dada mediante *Discrete Wavelets*, $\{V_j\}_j$, manteniendo la relación entre espacios adyacentes, V_j y V_{j+1}, por factor de escala de 2. Por lo tanto, basándose en una base *wavelet*, W_j, el espacio de la serie satisface la relación $V_j \oplus W_j \subset V_{j-1}$ del subespacio $L^2(\mathbb{R})$, su descomposición sea:

$$V_0 = W_1 \oplus W_2 \oplus W_3 \ldots W_j \oplus V_j \qquad (2.8)$$

y, por lo tanto, el espacio sea,

$$L^2(\mathbb{R}) = \bigoplus_{m=-\infty}^{\infty} W_m \qquad (2.9)$$

Debido a esta serie de descomposición, los componentes de cada espacio W_j contienen distintos detalles, bandas de frecuencia, de la función, resultando en una descomposición de filtros ortogonales de la señal original. Por lo tanto, en el análisis de esta técnica, los detalles de la señal se obtienen mediante productos escalares entre las señales escaladas y las *wavelets*. Para implementar una descomposición *wavelet* computacionalmente rápida, Mallat, en 1989, diseño un sistema por el cual la señal es analizada en función de la jerarquía de los niveles de resolución, difiriendo por un factor de escala de 2.

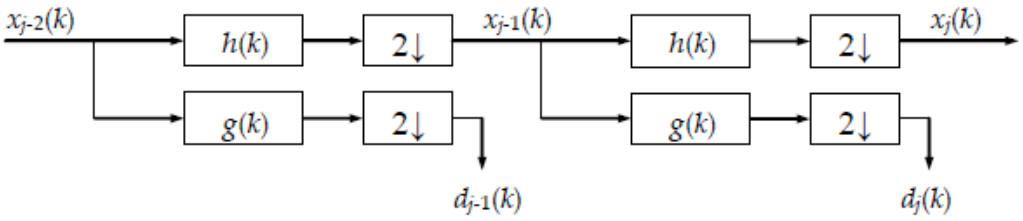

Figura 2.5 Descomposición de una señal en niveles según el algoritmo de Mallat [21].

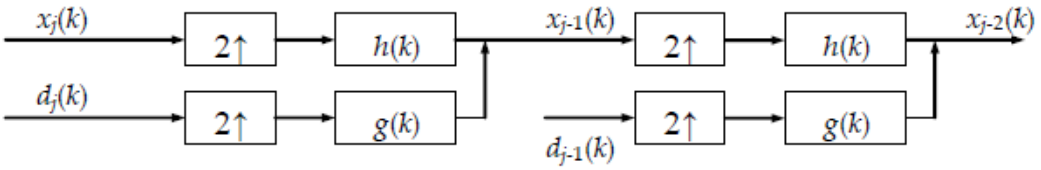

Figura 2.6. Reconstrucción de la señal original a partir de las señales dadas por la descomposición de los niveles según el algoritmo de Mallat [21].

2.3. EXTRACCIÓN DE CARACTERÍSTICAS

La extracción de características en una señal electrocardiográfica se basa en la detección de los picos de la onda R. Estos picos son detectados usando un algoritmo basado en *Wavelet Transform* (WT), debido a que estos picos del complejo QRS producen un par de módulos máximos, pero con signo opuesto, donde el pico R es uno de ellos. Para aumentar el rendimiento frente a perturbaciones de la señal y aumentar la efectividad de detección del QRS del algoritmo, se utiliza una técnica de umbral adaptativo, evitando de este modo falsos positivos debidos a los artefactos o falsos negativos debidos a ondas R de escasa amplitud [11].

Dentro de esta etapa de extracción de características, además de obtener un enventanado de los datos de entrada de la red para facilitar el aprendizaje de la red neuronal utilizada, se realizan técnicas de escalado de la señal, ya que los datos provienen de varias fuentes de datos sin ninguna relación entre ellas en los que se usan sistemas de medición de las señales ECG con diferentes características y se proponen técnicas de aumento de datos sobre las señales ECG, parámetro necesario en el caso de no disponer de una gran fuente de datos, en vista de los utilizados en la literatura científica.

2.3.1. TÉCNICAS DE ESCALADO DE CARACTERÍSTICAS

El preprocesamiento de los datos no es únicamente un paso correspondiente con la transformación de datos en bruto, datos originales obtenidos de una única fuente o diversas fuentes, en datos procesados, datos adaptados, sino que esta secuencia de procedimientos se utiliza para mejorar el rendimiento de los algoritmos de machine learning [26].

Las técnicas de escalado de características se definen como un conjunto de técnicas utilizadas para comprimir o extender los valores de la variable analizada propiciando la definición de un rango. Existe un gran abanico de técnicas de escalado conocido como métodos de normalización y/o estandarización utilizados en el estado del arte. Este tipo de técnicas es muy sensible por algunos tipos de algoritmos de aprendizaje automático como las regresiones lineales, logísticas, redes neuronales o algoritmos basados en distancias como K-Means, KNN o SVM, entre otros.

Los algoritmos de aprendizaje automático que utilizan el descenso de gradiente como técnica de optimización requieren que los datos se escalen. La razón se encuentra en la ecuación del descenso de gradiente:

$$\theta_j := \theta_j - \alpha \cdot \frac{1}{m} \cdot \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) \cdot x_j^{(i)} \qquad (2.10)$$

Teniendo en cuenta la presencia de la matriz de los datos de entrada, $x^{(i)}$, en la ecuación anterior, el rango de valores utilizados afectará en su análisis. La diferencia en los rangos provocará diferentes tamaños de paso, rangos dispares de θ, derivando en una convergencia más lenta del algoritmo.

Por otra parte, los algoritmos basados en distancia requieren de la utilización de este tipo de técnicas de escalado para evitar la posibilidad de tener datos con distintos rangos de valores dentro del análisis de la misma característica.

No obstante, las técnicas de escalado pueden resultar insensibles a la escala de características utilizada cuando las funciones empleadas en la resolución de los algoritmos utilizan una única característica. Esta homogeneidad explica su nula influencia por otros factores. Los algoritmos basados en árboles pertenecen a este grupo, ya que divide un nodo en función de una única característica, por lo tanto, no hay efecto de las funciones restantes en la división. Las técnicas de escalado más empleadas son la normalización y la estandarización:

- La **normalización**, técnica que pertenece al abanico de métodos de preprocesamiento de datos, tiene como objetivo transformar los valores de los datos iniciales en un conjunto de datos con una escala común, entre 0 y 1, sin distorsionar las diferencias en los rangos de valores.

$$X' = \frac{X - X_{min}}{X_{\max} - X_{min}} \tag{2.11}$$

- La **estandarización** es otra técnica de escalado donde los valores se centran alrededor de la media con una desviación estándar unitaria.

$$X' = \frac{X - \mu}{\sigma} \tag{2.12}$$

2.3.2. AUMENTO DE DATOS

Para el aprendizaje de series temporales, las redes convolucionales requieren una gran cantidad de datos, siendo posiblemente necesario un aumento de datos. Las técnicas de aumento de datos en series temporales residen en una modificación de las características de las señales. Hongen Liao, [12], propone técnicas de reducción de ruidos mediante DWT, utilizando varios tipos de wavelets para filtrar la señal ECG en sus subniveles: Daubechies 4 wavelet (db4), Daubechies 6 wavelet (db6) y Symlets 8. Con esta técnica, [12] obtiene un conjunto de datos cuatro veces mayor, compuesto por los datos originales junto con los datos filtrados con su respectiva wavelet.

Otras técnicas se centran en la modificación de la serie temporal de las señales, [13], como *Window Warping* (WW), o la deformación de las ventanas segmentadas. Sin embargo, una característica muy importante y común al análisis de todo tipo de problemas es conocer el ámbito en el que se está trabajando; realizar una modificación de características frecuenciales de las señales de electrocardiograma sin ninguna regla en la que basarse, puede llevar consigo modificaciones de las patologías que la señal ECG contiene y enmascarar características importantes para el diagnóstico de estas patologías cardiacas. Por lo tanto, para realizar técnicas de aumento de datos hay que analizar detalladamente qué patologías se están utilizando y saber cómo van a ser los datos de alimentación de la red, con el fin de evitar modificar partes de la señal que sean significativas para su diagnóstico. En el capítulo siguiente se redactan las técnicas utilizadas para el entrenamiento de la red neuronal convolucional.

2.3.3. SEGMENTACIÓN DE LATIDOS

En el análisis de la señal electrocardiográfica es muy importante la detección de ciertos elementos para obtener la duración y amplitud de las ondas con el fin de aislar los latidos del corazón ya que la actividad eléctrica en el corazón se puede medir como una secuencia de amplitudes lejos de una señal electrocardiográfica basal. La separación de los latidos facilita una visión específica de la morfología PQRST, proporcionando la base para mediciones útiles en la evaluación de la salud general del corazón humano y en la presencia de patologías cardiacas.

Para realizar esta separación, técnica denominada segmentación de latidos o enventanado, se debe localizar los picos R de cada latido cardiaco, véase figura 1.1, y averiguar los intervalos que unen estos picos, intervalos RR. Ivaylo I. Christov, [34], proponen un método para la detección de los QRS en tiempo real basado en la comparación entre valores absolutos del sumatorio de una o varias derivaciones de los electrocardiogramas y un umbralizado adaptativo. El método propuesto se ha dividido en dos algoritmos complementarios: uno que detecta el latido actual y otro método que tiene un componente de análisis del intervalo RR. Los algoritmos se autoajustan a los umbrales establecidos y a las constantes de ponderación, independientemente de la resolución y la frecuencia de muestreo utilizada, parámetros que son utilizados como datos de entrada al algoritmo. El desarrollo de estos algoritmos permite su funcionamiento con cualquier derivación del electrocardiograma, se auto sincronizan a la morfología del QRS o del latido y se adaptan a los intervalos entre ellos.

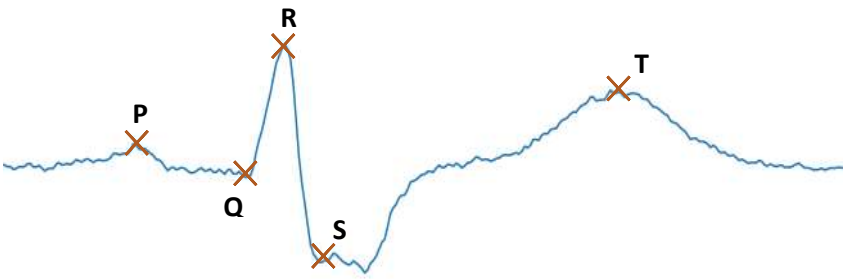

Figura 2.7. Ejemplo de la segmentación de un latido. El segmento está constituido por PQRST propio de la señal.

Tras esta etapa de extracción de características, la siguiente tarea reside en estudiar la técnica de aprendizaje profundo que tratará de utilizar estos datos para realizar una clasificación del estudio planteado.

2.4. CNN-1D

Las Redes Neuronales Convolucionales o *Convolutional Neural Networks*, CNN, son un tipo de redes de aprendizaje profundo, *deep learning*, con aprendizaje supervisado[2], que procesan los datos de entrada mediante capas intermedias que imitan al córtex visual del ojo, con el objetivo de identificar características específicas de estos datos que proporcionan la información esencial para su clasificación. Su analogía a las redes neuronales densas radica en la optimización automática de las propias neuronas de las capas ocultas a través del aprendizaje [35]. Estas neuronas seguirán recibiendo una entrada y realizando un conjunto de operaciones (producto escalar seguido, si así se requiere, de una función de activación no lineal) para obtener una salida. Desde los vectores o serie de datos temporales introducidos en el canal de entrada, a la obtención de la clasificación de la salida de la red, la función de coste seguirá basándose en los pesos calculados. Las últimas capas contendrán funciones de pérdida asociadas a las clases.

Por otro lado, las principales diferencias de las redes convolucionales con las redes neuronales densas es la reducción del número de parámetros de la red y la conexión de las neuronas entre las capas profundas. En una red convolucional, cada neurona de una capa no se encuentra conectada con todas las neuronas de la capa previa, propiciando que, a medida que se avanza en las capas convolucionales, las neuronas sean más específicas con relación a la información que abstraen, reduciendo drásticamente el número de operaciones a realizar y obteniendo más información de las características de los datos de entrada. De esta manera es como las redes neuronales convolucionales consiguen modelar una gran cantidad de datos con una reducción de parámetros: dividiendo el problema en partes aisladas para conseguir predicciones más sencillas y precisas [36]. Esta propuesta por parte de la red convolucional ha supuesto que, tanto científicos como desarrolladores, puedan aprovecharse de modelos con grandes estructuras para resolver tareas complejas, donde las redes neuronales artificiales no llegaban a satisfacer.

Las redes neuronales convolucionales basan sus fundamentos en el sistema Neocognitron, introducido por Kunihiko Fukushima en 1980 [37], combinado con el método de aprendizaje básico de propagación hacia atrás o *back-propagation* que ayuda al sistema a entrenar correctamente. Este tipo de redes de aprendizaje profundo han evolucionado hasta ser uno de los modelos más exitosos dentro del ámbito de reconocimiento de objetos y extracción de características de imágenes. Estas arquitecturas se emplean típicamente en problemas de tratamiento de imagen; sin embargo, multitud de datos se almacenan en forma de series temporales: medidas climáticas, pruebas médicas, vibraciones [45], análisis de motores, audios, etc. En series temporales, las redes neuronales convolucionales, en este caso, de una dimensión, CNN-1D, han demostrado que tienen varias ventajas sobre otros métodos tradicionales en este ámbito, sobre todo frente a aquellos modelos que incorporan ingeniería de características previamente:

- Las CNN-1D son modelos altamente resistentes al ruido.
- Las CNN-1D son capaces de extraer características, independientes del tiempo, y crear representaciones de series temporales automáticamente

Una de las particularidades más importante en los problemas que son resueltos con redes neuronales convolucionales es que las características, *features*, de los datos de entrada no deberían de ser espacialmente dependientes de la posición, es decir, la red convolucional debe ser capaz de detectar las características de los datos independientemente de su posición, ya sea en la imagen o en el eje de la serie.

[2] El aprendizaje supervisado corresponde con aquellos modelos que se aprenden relaciones que asocian entradas con salidas y que, en función del error obtenido, ajustan dicha relación para obtener la salía deseada.

2.4.1.ESTRUCTURA GENÉRICA DE CNN

Las redes neuronales convolucionales tienen como objeto reducir el número de pesos y el tamaño de los datos de entrada de forma que el procesamiento sea asequible para la red sin perder ninguna característica que sea crítica para realizar una buena predicción. Debido a esto, el correcto diseño de la arquitectura de la red se convierte en una tarea importante, no solo a nivel de características de los datos, sino que la red debe ser escalable a grandes fuentes de datos y evitar el sobreajuste, comúnmente denominado *overfitting*[3].

Las redes convolucionales utilizan una arquitectura de red dividida en dos partes: *Extracción de características* y *Clasificación*. La primera parte, *Extracción de características*, se compone de 4 tipos de capas: capas convolucionales, capas de *pooling*, capas de normalización y capas de salida. La segunda parte, *Clasificación*, corresponde con una capa de salida denominada *Fully-Connected*.

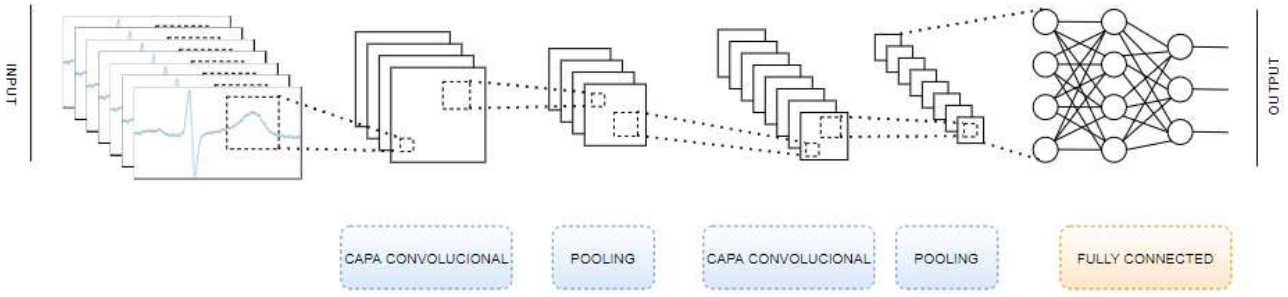

Figura 2.8. Estructura genérica de una red convolucional multi-input.

- **Capas convolucionales**: Estas capas suelen corresponderse con el grueso de la arquitectura de la red convolucional y se caracterizan por aplicar operaciones de convolución [36] en lugar de productos entre matrices. Estas capas obtienen los datos de entradas en formato matricial y realizan operaciones de convolución con filtros, *kernels*, que son matrices cuyas dimensiones son de un tamaño prefijado, para obtener una salida.
  El objetivo de la capa de convolución es extraer características de los datos de entrada. A lo largo de la arquitectura de la red, dependiendo de la complejidad del estudio, es necesario aplicar varias capas convolucionales ya que, las primeras capas se caracterizan por extraer información específica o de *bajo nivel* (ejes, vértices, colores, orientaciones de gradientes, etc.) de los datos de entrada, pero a medida que se apliquen capas convolucionales, la información obtenida serán zonas genéricas de los datos que ayuden a su predicción en la clase correspondiente. [38]

[3] Se considera como el efecto de sobreentrenar un algoritmo con unos ciertos datos para los que se conoce el resultado deseado. También conocido como sobreajuste, hace referencia al fallo de un modelo al generalizar.

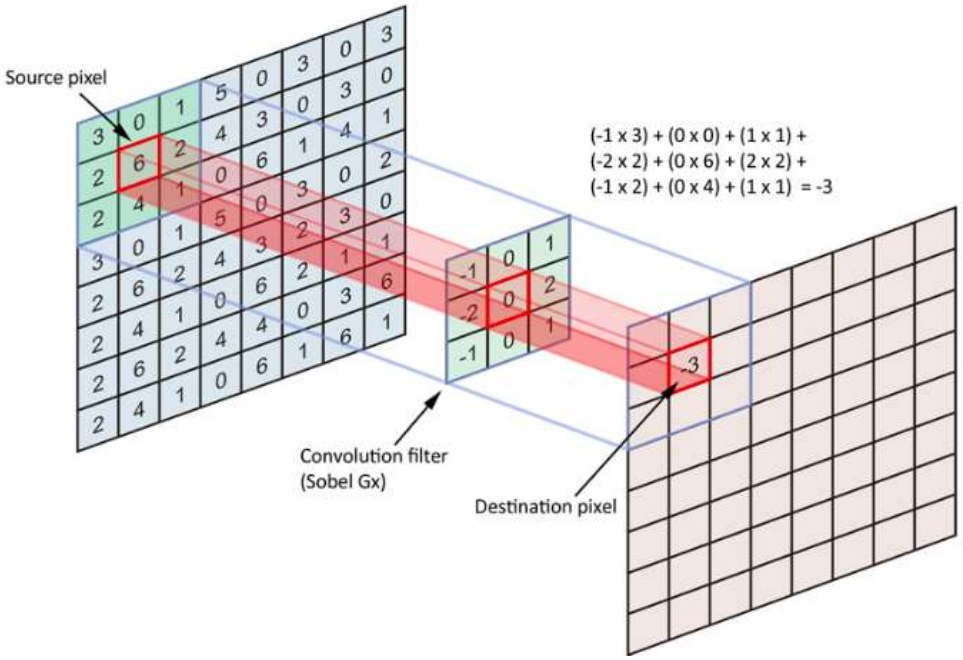

Figura 2.9. Representación de capas convolucionales con su respectiva capa de filtrado (kernel) [39].

- **Capas de *pooling***: Las capas de *pooling* o submuestreo, se sitúan a la salida de las capas convolucionales y su principal propósito es reducir el número de parámetros de los datos de entrada. De esta forma, estas capas ayudan al modelo a controlar el sobreajuste de los datos, extraer las características dominantes para la identificación de los datos de entrada y reducir notoriamente el coste computacional del proceso, así como aumentar la eficiencia de la red.

  La operación de *pooling* requiere determinar el tamaño de la región, también denominado *kernel*, sobre la que se quiere realizar el submuestreo de los datos de entrada y el tipo de *pooling* que se requiere. Hay dos tipos de submuestreo: *Max. Pooling*, que devuelve el valor máximo de la región cubierta por el kernel, y *Average Pooling*, que devuelve el valor de la media de todos los valores de la región cubierta por el *kernel* [38].

- **Capas de normalización**: Existen multitud de tipos de capas de normalización propuestas para la arquitectura de una red convolucional; sin embargo, estas capas han resultado ineficientes en la práctica por su mínimo resultado significativo [41].

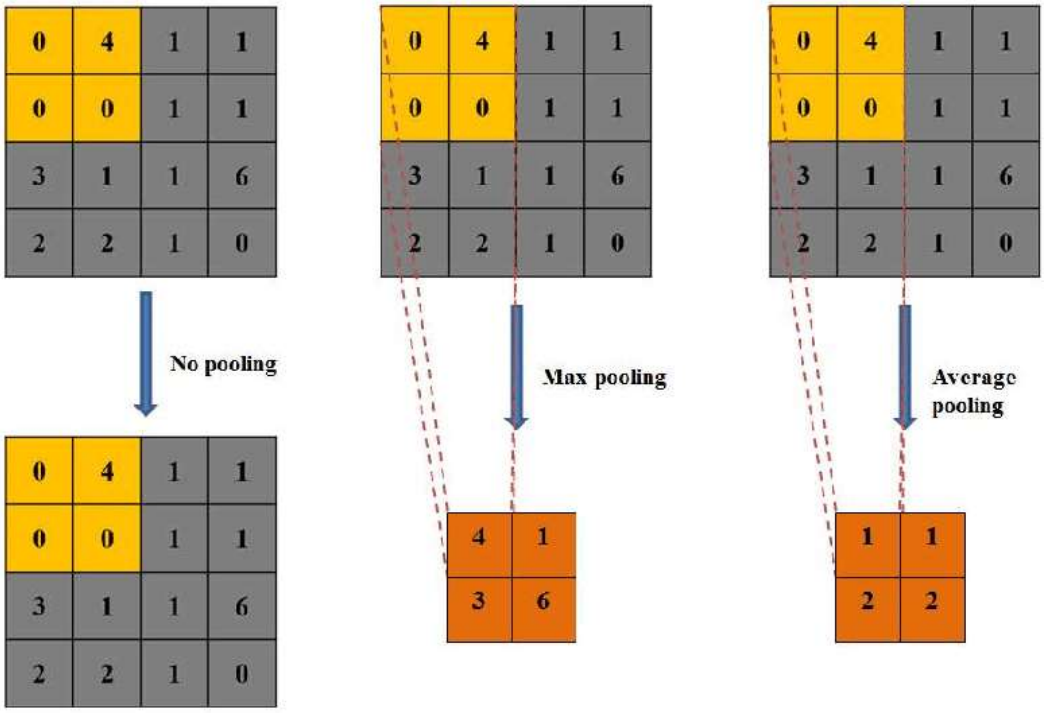

Figura 2.10. Representación de capas de pooling (No pooling, Max Pooling y Average Pooling) [40].

Por último, tal y como se puede ver en la figura de la estructura de la red, Fig. 2.8, se encuentra la parte de clasificación de la red compuesta por la capa de salida denominada *Fully-Connected*. La arquitectura de la red convolucional se divide en *Extracción de características*, debido al objetivo de las capas convolucionales y capas de *pooling* de extraer información dominante de los datos de entrada, y en capas densas y capa de salida denominada *Fully-Connected*. Previamente a la capa de salida *Fully-Connected*, la red convolucional no realiza ningún tipo de clasificación debido a que las neuronas de su propia capa de salida no serían capaces de generalizar, ya que, como se ha explicado, carecen de interconexiones y se encuentran especializadas, únicamente, en una parte de los datos de entrada. Es por ello que, a pesar de que la capa de salida *Fully-Connected* no sea necesaria para realizar una red convolucional, es un pilar muy importante en su desarrollo.

- **Capa de salida o *Fully-Connected***: La capa de salida o *Fully*-Connected es una capa donde las neuronas se encuentran interconectadas con todas las funciones de activaciones de las capas previas, tal y como sucede en las RNA. Sus activaciones pueden ser obtenidas por medio de multiplicaciones matriciales seguidas de un offset.

  Con la utilización de esta capa se obtiene una manera de realizar combinaciones lineales entre las características de alto nivel de la última capa convolucional.

2.5. INTERPRETABILIDAD

El machine learning o aprendizaje automático tiene un gran potencial para mejorar cualquier tipo de proceso convencional, productos e investigación. Sin embargo, el uso de técnicas inteligentes en el desarrollo de un proyecto, frecuentemente, es referido al concepto de *caja negra*. Este concepto plantea emular la utilización de dispositivos, sistemas, objetos o algoritmos matemáticos que pueden ser definidos en términos de entradas al sistema y salidas, o ciertas características del proceso, pero sin ningún conocimiento del funcionamiento interno.

Figura 2.11. Ejemplo de una caja negra como una representación hacia el uso de técnicas de Machine learning.

La existencia de este tipo de caja negra supone que el algoritmo empleado no explique su resultado en la predicción, derivando en una barrera de adopción del aprendizaje automático, además de proponer una barrera de aceptación de la tecnología empleada. ¿Cómo saber si los resultados obtenidos por el modelo de aprendizaje automático empleado son correctos o no? Si se desea realizar un algoritmo de clasificación y tras la aplicación del modelo de machine learning se obtiene una precisión del 95% en datos de test, ¿cómo saber en qué parte de los datos input se está fijando el modelo para realizar dicha clasificación? Conocer las decisiones que adopta el modelo para realizar su aprendizaje es lo que se conoce como interpretabilidad.

Según Miller, [25], el término genérico de interpretabilidad puede definirse como el grado en que un humano es capaz de comprender la causa de una decisión. Si esta definición es adoptada como válida y se deriva al campo de la inteligencia artificial, crea sentido distinguir que el proceso de interpretabilidad de un modelo reside sobre la interpretación humana, es decir, cómo de bien es capaz un humano de entender las decisiones tomadas por el modelo para obtener la salida deseada. Por otra parte, el departamento de Ingeniería Informática de la Universidad Normal de la Capital de Beijing, China [24], proporciona una definición más técnica correspondiente con el objeto del presente proyecto. Estos denominan al aprendizaje basado en la interpretabilidad como aprendizaje explicativo o *Explained Learning.* Este tipo de aprendizaje trata de analizar y responder la instancia en la que se encuentra el modelo desarrollado a través de los conocimientos previos en el campo y a través del conocimiento conceptual del propio aprendizaje.

Por ende, se observa como el concepto de interpretabilidad adoptado por el área de la inteligencia artificial es dependiente de su significado primitivo, la interpretación que los seres humanos damos al modelo, ya sea por la facilidad de entender la decisión tomada o bien por un conocimiento conceptual previo en el campo de aplicación del modelo.

Dada esta introducción, se puede considerar que la interpretabilidad de las decisiones tomadas por un algoritmo de Machine learning pasa a un plano muy importante, ya que proporciona comprensión sobre el modelo, ayuda en la toma de decisiones, información sobre la ingeniería de características, seguridad y robustez en su desarrollo, justificación de acciones y/o una mejora sustancial en el propio desarrollo. No obstante, la interpretabilidad de un modelo de Machine learning es considerada ineficaz cuando no tiene un impacto significativo en el problema planteado o en su respectivo área de aplicación, cuando el área de aplicación no se encuentra bien estudiado, es decir, no hay un conocimiento conceptual previo a la aplicación o cuando el problema está tan bien estudiado que los algoritmos de interpretabilidad no proporcionan ninguna ventaja adicional.

La importancia de la interpretabilidad, el porqué, es realmente significativa para los seres humanos. Si un modelo de Machine learning obtiene un rendimiento altamente eficaz, ¿cuál es la razón por la que los humanos terceros, es decir, aquellos ajenos al desarrollo, no creen en el modelo de Machine learning e ignoran el porqué de su decisión final? Según [26], "El problema es que una simple métrica como la precisión en la clasificación del modelo, es una incompleta descripción de las tareas del mundo real". Debido al factor de la curiosidad humana, la necesidad de la explicación del modelo se convierte en un problema, ya que no es suficiente obtener la predicción, el qué, sino que el modelo debe explicar el porqué de dicha predicción para solucionar el problema inicial. No obstante, no hay un consenso real sobre qué es la interpretabilidad en Machine learning ni cómo realizar una medición sobre ella [28].

Robnik-Sikonja y Bohanec, [29], por su parte, proponen en 2018 unas propiedades de explicación de los métodos y modelos que pueden ser utilizadas para definir y/o formalizar la evaluación de la interpretabilidad:

- ***Precisión del modelo***: La obtención de una alta precisión/rendimiento en el desarrollo del algoritmo es importante frente a la explicación del modelo de Machine learning.
- ***Fidelidad***: Característica que se relaciona con la aproximación de la explicación de la predicción del modelo. La alta fidelidad es una de las propiedades más importantes de una explicación, ya que una explicación con baja fidelidad es inútil para explicar el modelo de Machine learning. La precisión y la fidelidad se encuentran estrechamente relacionadas.
- ***Consistencia***: Esta propiedad se encuentra sujeta a las diferencias que puedan existir entre las predicciones propuestas por dos modelos distintos sobre una misma aplicación. Si los modelos ofrecen predicciones similares y se computa la interpretabilidad utilizando un método de libre elección para cada uno de ellos, la consistencia se verá afectada por la similitud o lejanía de las explicaciones obtenidas. Si las explicaciones son muy similares, indica que son altamente consistentes.
- ***Estabilidad***: La estabilidad es similar a la propiedad de la consistencia; sin embargo, la propiedad de la estabilidad compara las explicaciones entre instancias/datos similares en un modelo fijo. Una alta estabilidad de las explicaciones significa que las características obtenidas por el modelo pueden llegar a sufrir ligeras variaciones, pero no afectan a la interpretabilidad del modelo. En el caso contrario, si el modelo carece de estabilidad, se deriva en un grado alto de varianza lo que significa que el modelo no puede reconocer correctamente una característica.
- ***Comprensibilidad***: Esta propiedad de la interpretabilidad se encarga de definir y medir qué tan bien entienden los humanos ajenos al desarrollo del algoritmo las explicaciones. La medición de esta propiedad se basa en el tamaño de la explicación (número de características o reglas de decisión existentes), predicción del comportamiento del modelo y/o comprensibilidad de las características resultantes de la transformación del modelo de las características originales.
- ***Certeza***: Esta propiedad indica el grado de confianza en la predicción del modelo.
- ***Grado de importancia***: Propiedad que indica la importancia de las características de los datos originales frente a la predicción del modelo y explicación.

- ***Novedad***: La novedad actúa frente a los datos obtenidos a raíz de una fuente de datos alejada de la distribución de datos de capacitación. Este concepto está relacionado con el concepto de certeza. Cuanto mayor sea la novedad, más probable es que el modelo tenga una baja certidumbre debido a la falta de datos.
- ***Representatividad***: Las explicaciones pueden abarcar todo el modelo. Esta propiedad hace referencia al número de instancias que pueden abarcar todo el modelo.

Esta evaluación de las propiedades puede llevarse a cabo para definir cómo de explicable e interpretable es un modelo de Machine learning. Los seres humanos no preguntan el porqué de una probabilidad en una predicción, se preguntan el porqué de la obtención de dicha predicción.

Debido a esto, han surgido técnicas recientemente sobre la interpretabilidad de las redes neuronales con el fin de clarificar el aprendizaje del modelo. En el presente proyecto se examinan dos tipos de técnicas de interpretabilidad muy utilizadas en el estado del arte: *Class Activation Maps* (CAM) y *Shapley Additive Explanations* (SHAP).

2.5.1.CLASS ACTIVATION MAPPING (CAM)

Las Redes Neuronales Convolucionales han demostrado tener un comportamiento prometedor frente a la detección de objetos gracias al aprendizaje por regiones que emplean las unidades de sus capas convolucionales. Sin embargo, una red neuronal convolucional utilizada con fines de clasificación y predicción requiere, en la mayoría de los casos, una capa de salida que congregue el aprendizaje específico de las regiones, dando lugar a una nueva estructura, explicada en apartados anteriores, que requiere de la agregación de una capa *Fully-Connected*. Debido a esta agregación, la red convolucional pierde su habilidad de extracción de características.

Recientemente, en el 2016 indica [30], varias redes neuronales populares como GoogLeNet o Network in Network (NIN) propusieron suprimir la capa *Fully-Connected* para minimizar el número de parámetros de la red manteniendo su alto rendimiento. Con este objeto, se empieza a utilizar una capa denominada *Global Average Pooling* que actúa como un regularizador estructural, previniendo de sobreajuste del problema durante el entrenamiento, a pesar de una pequeña minimización del rendimiento de la red.

No obstante, [30] encontró una peculiaridad añadida en la utilización de la capa *Global Average Pooling* más allá de actuar como un regularizador: el mantenimiento de la red de los pesos en la tarea de detección de objetos de la estructura convolucional. Este ajuste permite identificar y localizar las regiones de la imagen de las clases que analiza el modelo de aprendizaje profundo.

Por lo tanto, para obtener una visualización de estas regiones y obtener una interpretabilidad de los modelos de redes neuronales convolucionales en la detección de objetos, [30] propuso el empleo de *class activation maps* (CAM) utilizando *Global Average Pooling*. Un mapa de activación de clases o *class activation maps* de una categoría, clase o tarea particular indica las regiones discriminativas de la imagen, en el caso de tener como dato de entrada imágenes, utilizadas por la red neuronal convolucional para su clasificación en su respectiva categoría.

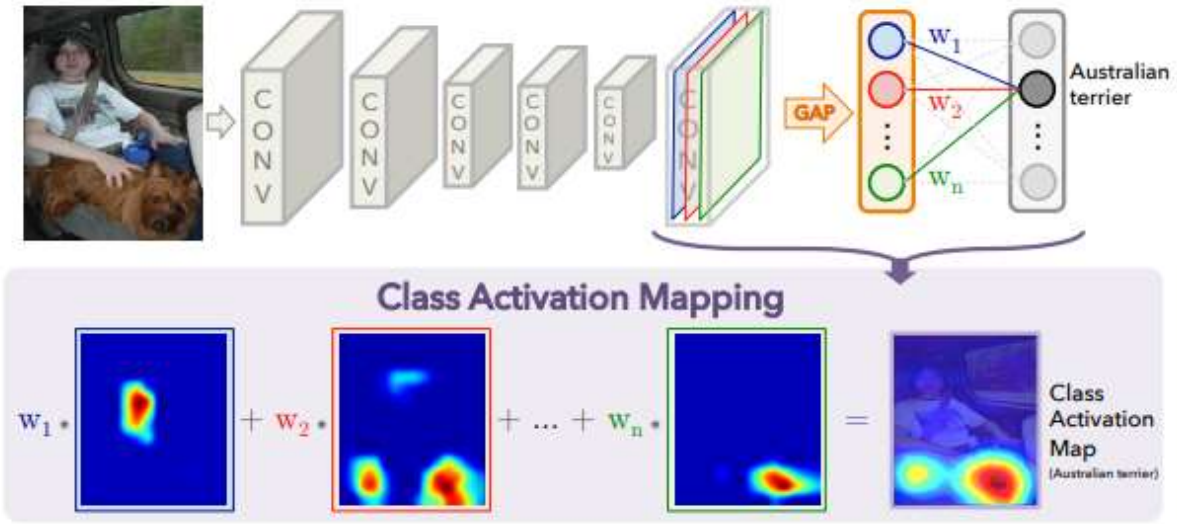

Figura 2.12. Representación de los resultados obtenidos tras aplicar Class Activation Maps [29].

Tomando como base [29], para un dato de entrada dado, por ejemplo, una imagen, sea $f_k(x, y)$ la representación de la unidad de activación k de la última capa de la estructura convolucional en cuya localización se centre en(x, y). Para esta unidad,k, el resultado de aplicación de la capa *Global Average Pooling* es:

$$F^k = \sum_{x,y} f_k(x, y) \tag{2.13}$$

Por lo tanto, dada una clase c, la entrada a su respectiva función de activación *Softmax*, S_c, se obtiene la siguiente ecuación donde w_k^c es el peso correspondiente para la clase c en la unidad $'k'$:

$$S_c = \sum_k w_k^c \cdot F_k \tag{2.14}$$

En otras palabras, w_k^c representa la importancia de F_k para la clase c. Finalmente, la salida de la *Softmax* corresponde con la probabilidad, P_c, para su clase determinada:

$$P_c = \frac{\exp(S_c)}{\sum_c \exp(S_c)} \tag{2.15}$$

Teniendo en cuenta el procedimiento anterior, se ignora el término del sesgo, utilizándolo explícitamente como nulo. Utilizando, entonces, la puntuación de la clases se obtiene:

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x,y) = \sum_{x,y} \sum_k w_k^c \cdot f_k(x,y) \quad (2.16)$$

Finalmente, se define el mapa de la activación de la clase c como:

$$M_c(x,y) = \sum_k w_k^c \cdot f_k(x,y) \quad (2.17)$$

La obtención de S_c y M_c indican directamente la importancia de la activación del mapa generado en una región determinada (x, y) permitiendo la visualización de la clasificación de la imagen para cada determinada clase.

Una resolución similar a la expuesta por CAM viene dada por [31] mediante la utilización de la capa *Global Max Pooling*, en vez de *Global Average Pooling*. La principal diferencia radica en que la *Global Max Pooling* limita su localización a un punto de la región del objeto mientras que la *Global Average Pooling* se centra en una región completa.

2.5.2.SHAP

El método SHAP nace de la observación de Scott M. Lundberg y Su-In Lee, [33], debido a la gran importancia que se requiere para llegar a interpretar correctamente la predicción o clasificación de un modelo. En este proceso de observación, adquiere un peso importante la evolución de los modelos empleados, así como las modernas bases de datos, instrumentación utilizada y el incremento de los beneficios de la aplicación de la variedad de modelos complejos a los que están expuestos.

SHAP, *Shapley Additive Explanations*, es una librería que trata de unificar métodos para la medición de la interpretabilidad de estos modelos, basada su elaboración en un tipo de aprendizaje nombrado previamente, *Explained Learning*. Esta librería obtiene unos valores *shapley* que ayudan a comprender y responder, desde el punto de vista del desarrollador, el por qué un modelo de Machine learning proporciona una salida determinada. La función de estos valores para ayudar a comprender el modelo empleado es tratar de cuantificar la contribución de las características obtenidas por el modelo tras su etapa de extracción de características y así conocer cuáles de ellas son más significativas para cada predicción o clasificación realizada.

Para obtener los valores *shapley* de un modelo, Lundberg y Su-In Lee [33], se basan en la idea de considerar el resultado de cada posible combinación posible llevada a cabo en su entrenamiento de las características que conforman los datos iniciales, denominadas f, siendo por tanto el rango posible de f de 0 al número total de características existente, denominado F. Para visualizar la relación entre los subconjuntos de características utilizadas, se puede hacer uso del término matemático Conjunto Potencia, que se puede representar por medio de un árbol:

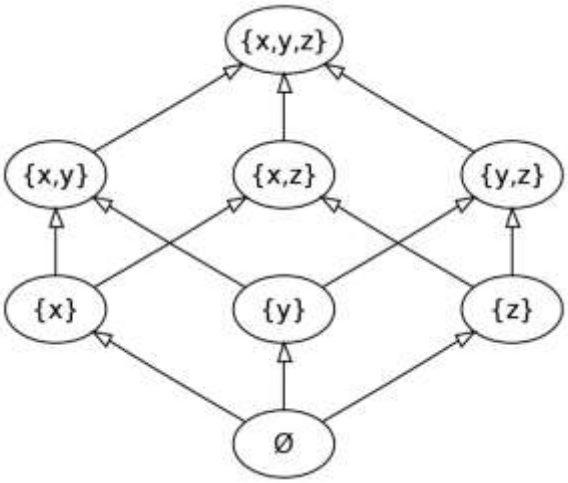

Figura 2.13. Representación en forma de árbol del conjunto potencia de las características x, y, z.

Cada nodo del árbol representa un subconjunto de características asociadas. El aprendizaje desarrollado en SHAP requiere entrenar un modelo distinto para cada subconjunto obtenido, es decir, $2\hat{}F$ modelos. Estos modelos que requieren esta metodología son equivalentes entre sí en lo que respecta a sus hiperparámetros y sus datos de entrenamiento; lo que cambia es el conjunto de características incluidas en el modelo.

El promedio de los valores *shapley* estimados de los distintos modelos aplicados producen la diferenciación en la importancia de características, elemento final del aprendizaje explicativo.

3. RESULTADOS

En el siguiente capítulo se exponen los resultados obtenidos del análisis de las señales electrocardiograma en función de las patologías cardiacas estudiadas. Primeramente, se expone el procedimiento seguido de la metodología en función del conjunto de datos utilizado, para seguir con un análisis detallado de los mismos. Finalmente, se realiza una evaluación de la interpretabilidad de los algoritmos implementados.

3.1. TRABAJO REALIZADO

3.1.1. DESCRIPCIÓN DEL CONJUNTO DE DATOS

Para llevar a cabo la aplicación de la metodología expuesta en el capítulo anterior, entre los que se encuentra el uso de técnicas de procesamiento de señal, extracción de características y sistemas inteligentes mediante el aprendizaje profundo, que permitan realizar una detección y clasificación automática de anomalías cardiacas, se debe disponer de una cantidad inmensa de datos iniciales, señales de electrocardiograma. Las señales de electrocardiograma utilizadas en el presente proyecto para el desarrollo de los algoritmos se han obtenido de un reto lanzado por *Physionet*, organización compuesta por socios de distintas industrias de EE.UU establecidos en el año 1999 que buscan proporcionar un acceso web gratuito a grandes cantidades de datos, comprendiendo señales fisiológicas u otros tipos de datos dentro del campo de la medicina, además de software de código abierto. Al mismo tiempo, *Physionet* proporciona retos de libre acceso a la comunidad de programadores en relación con los datos publicados periódicamente. Entre los distintos retos propuestos, se encuentra un área entera dedicada exclusivamente a retos de actividades y patologías cardiacas en colaboración con *Computing Cardiology*, conferencia científica internacional que se celebra cada año desde 1974. Anualmente, *Physionet* y *Computing Cardiology* proponen retos dentro de este ámbito y facilitan los datos de señales de electrocardiogramas. El objetivo de *Physionet* con la propuesta del reto del año 2020 es identificar de manera automática y clasificar patologías cardiacas sin la necesidad de una interpretación manual por parte de un profesional sanitario.

Hasta ahora, los organizadores del reto de cardiología aportaban señales de electrocardiograma incompletas, proporcionando únicamente la segunda derivación para el análisis de las patologías. Sin embargo, en Febrero de 2020 lanzan el siguiente reto: Physionet: Classification of 12-lead ECGs, el cual inicialmente proporcionaba 6877 señales ECG, con todas las derivaciones existentes, comprendiendo 9 tipos de patologías. Posteriormente, los organizadores del reto aumentaron el dataset ofertado hasta llegar a una cifra de 43.068 señales de electrocardiograma, incluyendo hasta 111 patologías. Los datos recabados por este reto son proporcionados por múltiples fuentes de manera pública, con el propósito de su tratamiento y experimentación por parte de sus participantes:

- CPSC (2018): La primera fuente corresponde con los datos propuestos en el desafío de señales fisiológicas de China de 2018 durante la Séptima Conferencia Internacional de Ingeniería Biomédica y Biotecnología en Nanjin, China. Los datos proporcionados se componen de un total de 6877 señales ECG donde 3699 corresponden con señales ECGs de hombres y 3.178 con señales ECGs de mujeres. Las señales registradas tienen una duración de entre 6 y 60 segundos de duración y fueron muestreadas con una frecuencia de muestro de 500 Hz.

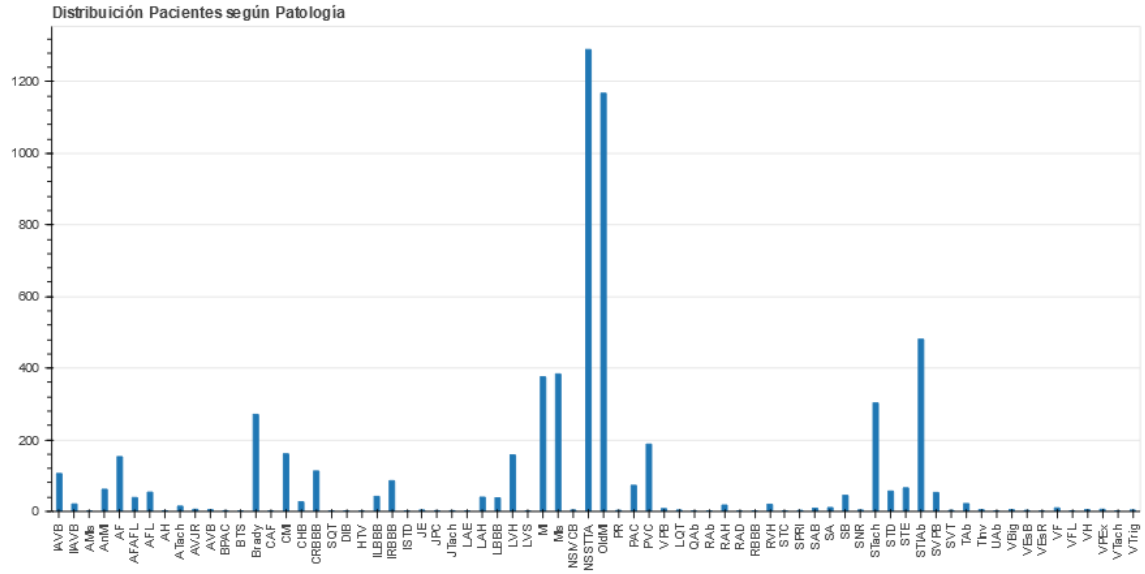

Figura 3.1.Conjunto de datos de la Séptima Conferencia Internacional de Ingeniería Biomédica y Biotecnología en Nanjin, China. (CPSC 2018)

- **INCART**: La segunda fuente corresponde con un dataset proporcionado de forma pública por el Instituto Técnico de Cardiología (INCART) en St. Petersburg, Rusia. Este dataset se compone de 75 señales electrocardiograma con una duración de 30 minutos y muestreados a 257 Hz.

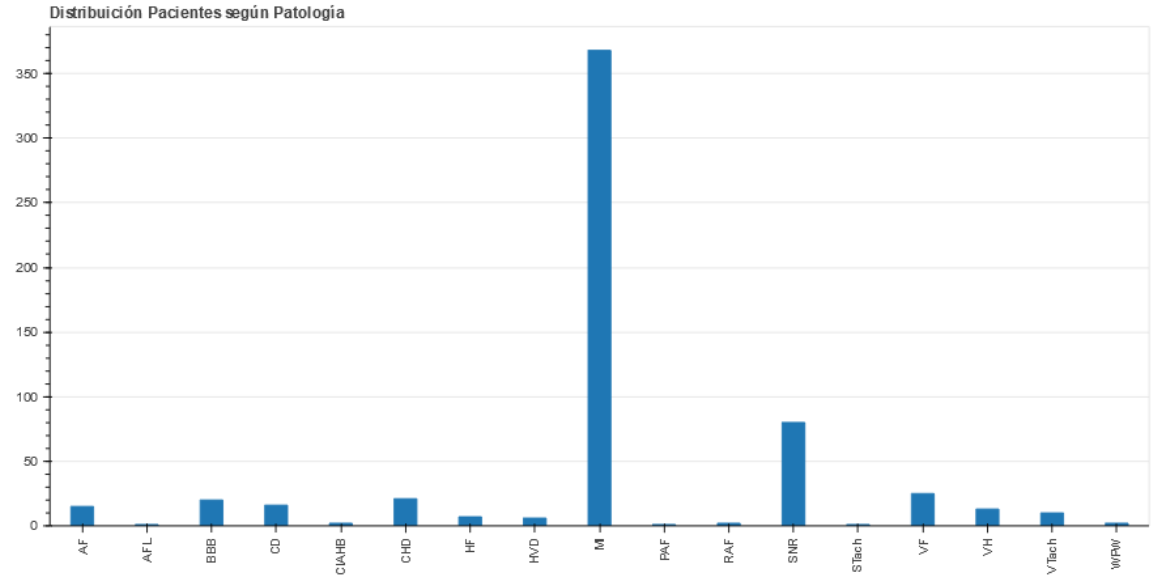

Figura 3.2. Conjunto de datos del Instituto Técnico de Cardiología (INCART).

- **PTB & PTB-XL**: La tercera fuente de datos corresponde con el dataset ofertado por Physikalisch-Technische Bundesanstalt (PTB) de Brunswick, Alemania. Este dataset incluye dos fuentes de datos: El dataset de diagnóstico de señales de electrocardiograma PTB y el PTB-XL, que hace referencia a un dataset mayor. El primero contiene 549 registros de señales ECG muestreadas con una frecuencia de 1000 Hz. El segundo contiene 21.837 señales ECG de 10 segundos de duración muestreadas a 500 Hz.

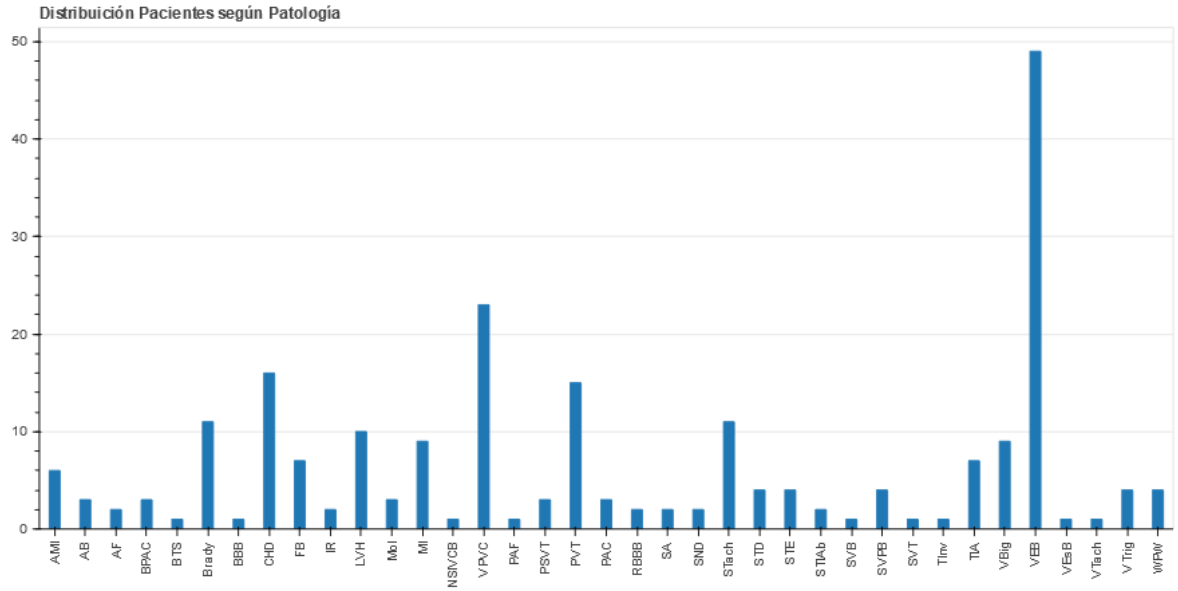

Figura 3.3. Conjunto de datos del Physikalisch-Technische Bundesanstalt (PTB) de Alemania.

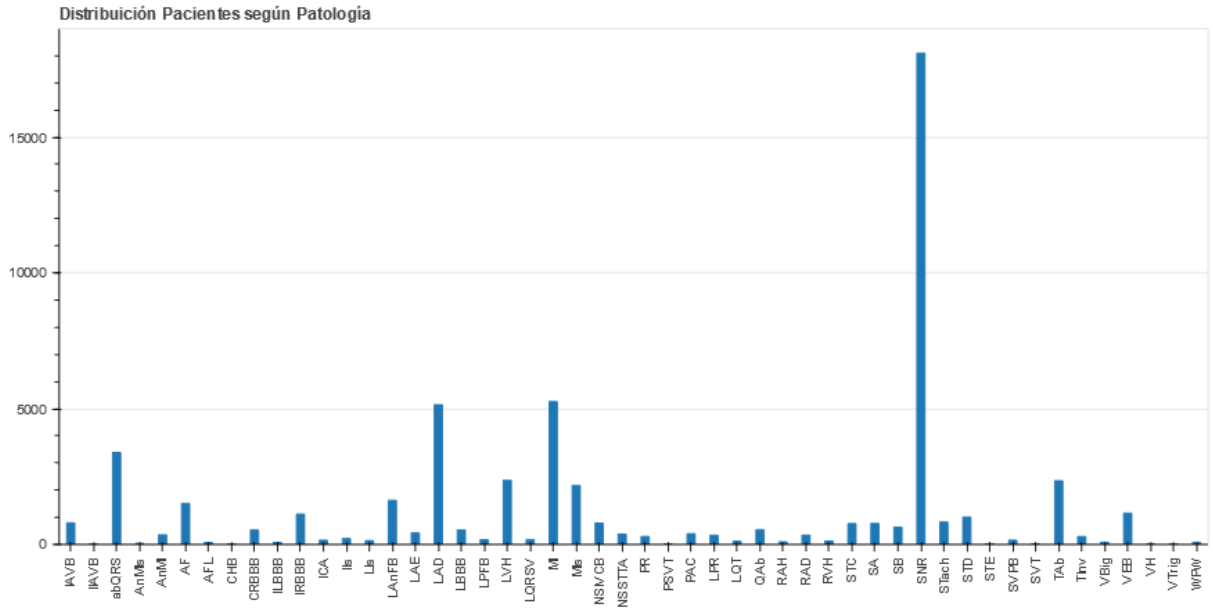

Figura 3.4. Conjunto de datos del Physikalisch-Technische Bundesanstalt (PTB-XL) de Alemania.

- **G12EC**: La cuarta fuente corresponde a los datos proporcionados por el desafío de señales ECG de 12 derivaciones de la Universidad Emory, Atlanta, EE.UU. Este dataset, compuesto por señales ECG de población del sudeste de Estados unidos, contiene 10.344 señales de electrocardiograma de 10 segundos de duración a una frecuencia de 500 Hz.

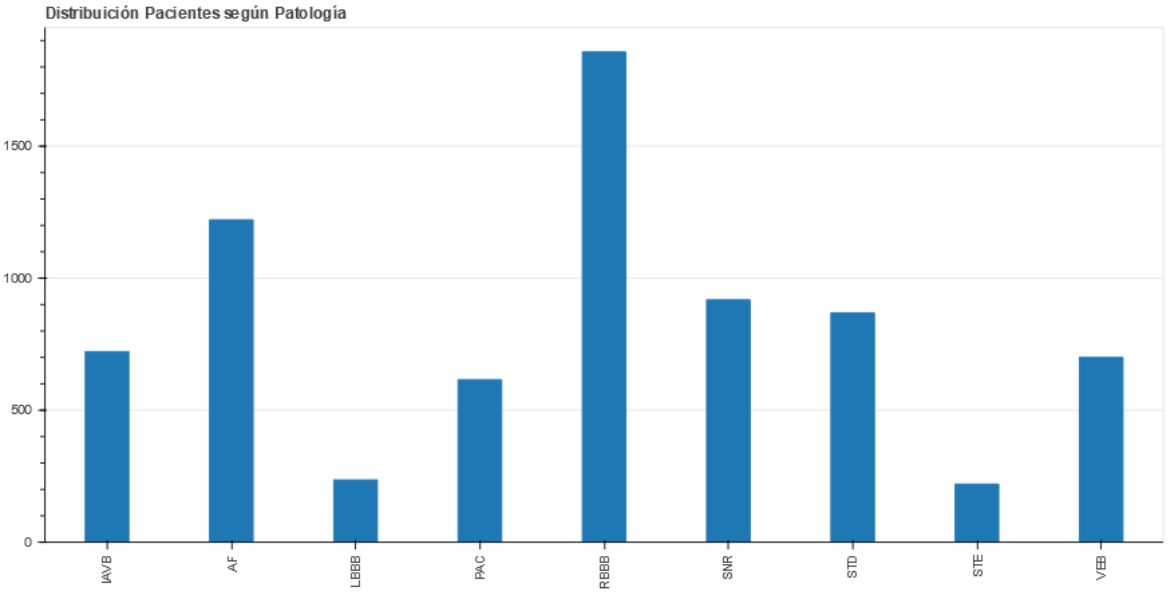

Figura 3.5. Conjunto de datos de la Universidad Emory de Atlanta (G12EC).

En la siguiente tabla se recaba la información general proporcionada por todos los conjuntos de datos:

Tabla 2. Análisis/Resumen de todos los conjuntos de datos obtenidos, separando sus características principales (duración del registro de la señal ECG, población de mujeres y hombres, número total de señales y la frecuencia de muestreo empleada para cada conjunto de datos).

DATASETS	*FRECUENCIA DE MUESTREO [Hz]*	*DURACIÓN [s]*	*Nº DE SEÑALES ECG EN MUJERES*	*Nº DE SEÑALES ECG EN HOMBRES*	*Nº TOTAL DE SEÑALES ECG*
CPSC (2018)	**500**	**6-60**	**3178**	**3699**	**6877**
INCART	**257**	**30 min**	**-**	**-**	**75**
PTB	**1000**	**10**	**377**	**139**	**549**
PTB-XL	**500**	**10**	**11.379**	**10.458**	**21.837**
G12EC	**500**	**10**	**5551**	**4793**	**10.344**

Cada señal de electrocardiograma del dataset lleva consigo un archivo de encabezado, formato .*hea, que contiene una descripción de la señal ECG y atributos de su paciente, incluyendo un diagnóstico diferencial de la patología cardiaca dada por personal sanitario del centro de su respectiva fuente de datos. El diagnóstico de cada paciente, representado en el archivo de cabecera como '*Dx'*, puede darse como una patología cardiaca o la suma de varias patologías cardiacas, lo que convierte el problema planteado en una clasificación multi-etiqueta.

Como se puede observar en la Tabla 2, el conjunto de señales que conforman el dataset que dará lugar al estudio planteado son señales variadas, con distintas frecuencias de muestreo, propias de cada procedencia de la fuente de datos, con distintas características de longitud, cantidad y corresponden con señales recogidas directamente de la instrumentación del electrocardiógrafo, sin haber tenido ningún tipo de procesamiento previo.

```
A0001 12 500 7500 05-Feb-2020 11:39:16
A0001.mat 16+24 1000/mV 16 0 28 -1716 0 I
A0001.mat 16+24 1000/mV 16 0 7 2029 0 II
A0001.mat 16+24 1000/mV 16 0 -21 3745 0 III
A0001.mat 16+24 1000/mV 16 0 -17 3680 0 aVR
A0001.mat 16+24 1000/mV 16 0 24 -2664 0 aVL
A0001.mat 16+24 1000/mV 16 0 -7 -1499 0 aVF
A0001.mat 16+24 1000/mV 16 0 -290 390 0 V1
A0001.mat 16+24 1000/mV 16 0 -204 157 0 V2
A0001.mat 16+24 1000/mV 16 0 -96 -2555 0 V3
A0001.mat 16+24 1000/mV 16 0 -112 49 0 V4
A0001.mat 16+24 1000/mV 16 0 -596 -321 0 V5
A0001.mat 16+24 1000/mV 16 0 -16 -3112 0 V6
#Age: 74
#Sex: Male
#Dx: 426783006
#Rx: Unknown
#Hx: Unknown
#Sx: Unknown
```

Figura 3.6. Archivo de encabezado proporcionado en el dataset para cada paciente registrado.

3.1.2. PREPROCESAMIENTO DE LOS DATOS

En esta sección se aborda la tarea del procesamiento de las señales de electrocardiograma de los pacientes. En concreto, se examinarán las técnicas explicadas en el capítulo previo de la metodología para afrontar el procesado de los datos y, en base a estas técnicas, obtener las señales de entrada para la aplicación de algoritmos inteligentes.

A la hora de aplicar algoritmos inteligentes como la red convolucional 1D, es importante realizar un procesamiento de los datos; más aún cuando los datos provienen de bases de datos distintas sin compartir ningún tipo de metodología en común. Es por ello por lo que la secuencia utilizada para realizar el procesamiento de las señales de las fuentes de datos descritas es la siguiente:

1. Limpieza de los datos
2. Procesamiento de la línea de interferencia base
3. Técnicas de reducción de ruido

3.1.2.1. LIMPIEZA DE LOS DATOS

Como se mencionó al comienzo de la presente tesis, en el planteamiento del problema, el análisis automático de las señales ECG se ha basado en desplegar algoritmos de estudio de características basándose en el dominio del tiempo y el dominio de la frecuencia. Estos métodos requieren de personal experto en el campo para realizar un trabajo de análisis y etiquetado de los datos.

Por otra parte, en los métodos de aprendizaje profundo, especialmente en el uso de redes convolucionales, no se requiere la labor de esta persona o grupo de personas, ya que estos métodos son capaces de realizar una extracción de las características, gracias a las capas convolucionales explicadas, sin previo dominio del campo de aplicación. Sin embargo, la utilización de datos sin procesar puede

acarrear grandes problemas a la hora de extraer características en común y más aún cuando se trabaja con distintas escalas entre los datos, por lo que se debe tratar de evitar esta situación en los datos de entrada.

El proceso de lectura de las señales de electrocardiograma puede verse perjudicado por errores en el registro dados por la instrumentación, movimientos del paciente o mala colocación de la sensórica utilizada. por parte del personal sanitario. Errores humanos o tecnológicos pueden causar que las señales ECG registradas tengan valores atípicos en todas las derivaciones, resultando inservible el electrocardiograma en su totalidad o también puede darse el caso de que únicamente una o varias derivaciones sean afectadas por los errores de lectura, lo que, en función del tipo de derivación puede llegar a ser perjudicial. La utilización de señales erróneas puede derivar en un procesamiento y en una extracción de características erróneas, siempre y cuando los errores atípicos de lectura predominen en los datos analizados.

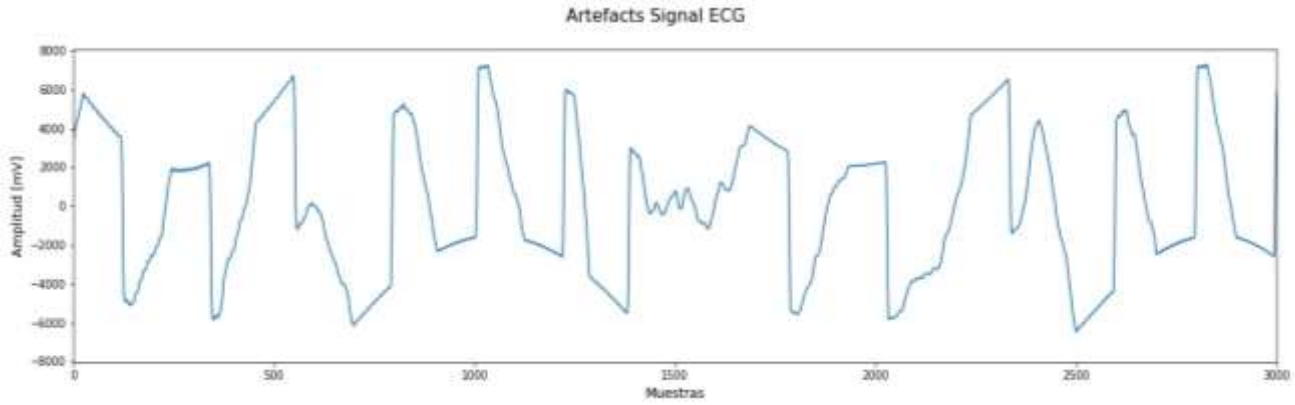

Figura 3.8. Señal de electrocardiograma de la derivación V6 compuesta por valores atípicos de la señal (artefactos) debido a una lectura errónea.

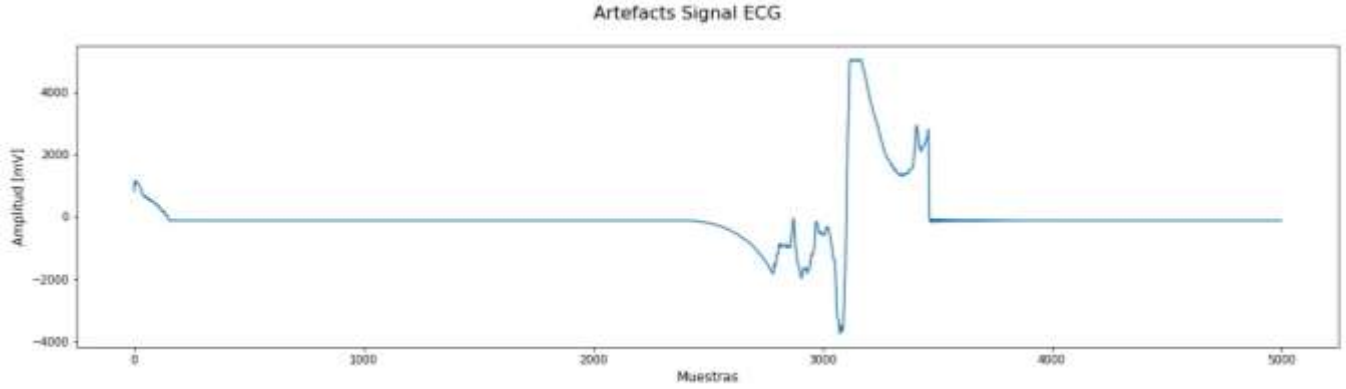

Figura 3.7. Señal de electrocardiograma de derivación V6 compuesta por una señal nula y un artefacto alrededor de la muestra 3000 debido a una lectura errónea de la señal.

Es importante localizar el mayor número de señales ECG que proporcionen falsos registros y suprimirlos del dataset analizado previamente al procesamiento de las señales para reducir el coste computacional generado. Para su localización se diseño un algoritmo que buscaba picos R en las señales. En el caso de no encontrar picos, se procedía a su eliminación, ya sea por los efectos señalados en la Fig. 3.8 o en la Fig. 3.9.

3.1.2.2. PROCESAMIENTO DE LA LÍNEA DE INTERFERENCIA BASE

Tras haber identificado señales cuya lectura resultó errónea y evita que se puedan utilizar en los pasos posteriores de extracción de características, se sigue con un parámetro muy importante en el procesamiento de las señales de electrocardiograma, siendo este el procesamiento de la línea de interferencia base.

La elección de la técnica de filtrado reside en la mejor puntuación obtenida por los experimentos llevados a cabo en [6]. En una primera instancia se seleccionó un filtro de mediana que permitiera cancelar el rango de frecuencias dado de la línea basal de las señales. El uso de este filtro conllevaba un coste computacional elevado, entre 1 y 2 segundos por señal procesada, tal y como señala la Tabla 1, derivando en un coste computacional de todo el conjunto de señales en más de 12 horas. Tras el análisis realizado, se contempló cambiar el método de análisis y se puede observar que la técnica *Wavelet Cancellation* predomina como el algoritmo con más rendimiento en varios de los factores; sin embargo, uno de los principales factores a tener en cuenta en el procesamiento realizado dentro del ámbito sanitario es el tiempo de computación. Debido a esto, se utilizó el filtro de Butterworth, siendo este predominante, en la materia abordada, con una gran diferencia, al mismo tiempo que contiene características similares en los demás campos de comparación frente al resto de técnicas de procesamiento señaladas.

Paralelamente, este factor es muy importante en el desarrollo del proyecto, aunque gracias al aporte del grupo GSDPI, se ha podido hacer uso de un servidor remoto para realizar el entrenamiento de la red convolucional.

El filtro Butterworth utilizado, filtro de paso alto para suprimir las frecuencias indeseadas correspondientes con las bajas frecuencias, entre 0.5 y 3 Hz, se componía de las siguientes características:

- Número de orden del filtro: 4
- Frecuencia de corte: 1 Hz

Con la utilización del filtro Butterworth se redujo el tiempo de cómputo en el procesamiento de las señales, entre 0.1 y 0.3 segundos de tiempo de cómputo medio por señal, mientras que el tiempo de cómputo final sería de alrededor 30 minutos.

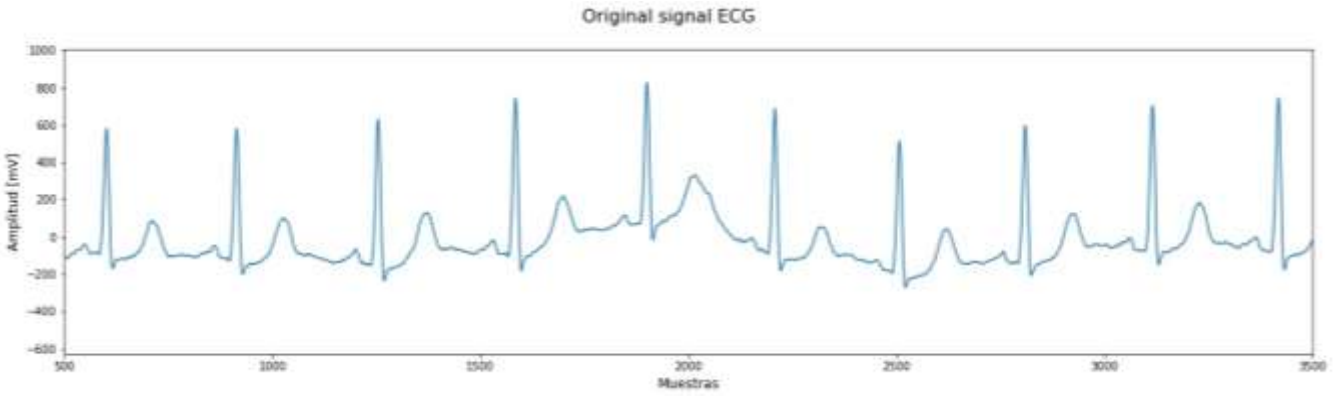

Figura 3.9. Señal de electrocardiograma con línea de interferencia base debido a fallo en la instrumentación, contracciones musculares o respiración del paciente.

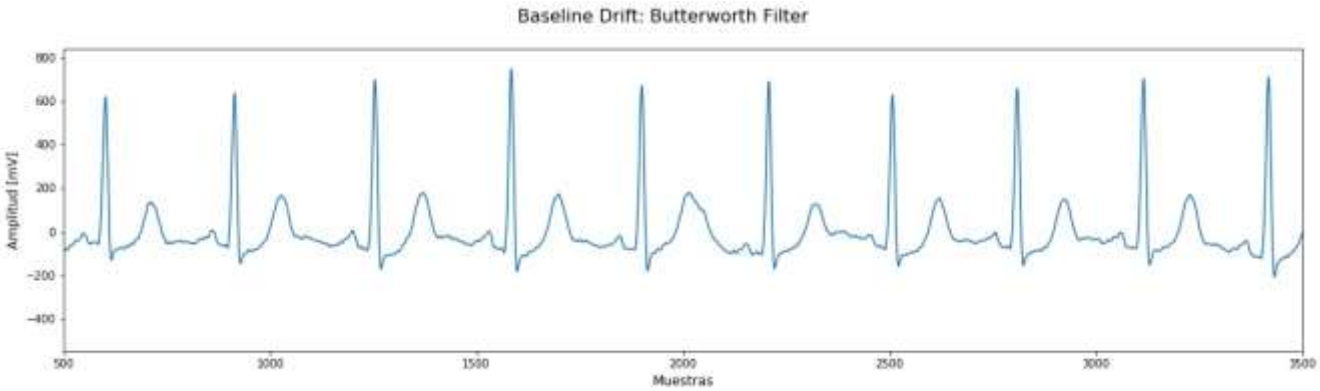

Figura 3.10. Reducción de la línea de interferencia base mediante un filtro de Butterworth.

En la figura anterior se puede observar el empleo del filtro Butterworth diseñado sobre una señal ECG, figura 3.10, con línea de interferencia basal. A pesar de que el tiempo de cómputo es mínimo con la utilización de esta técnica, hay una observación a destacar frente a la utilización del primer tipo de técnica estudiada, filtro corte banda, representado en la siguiente figura.

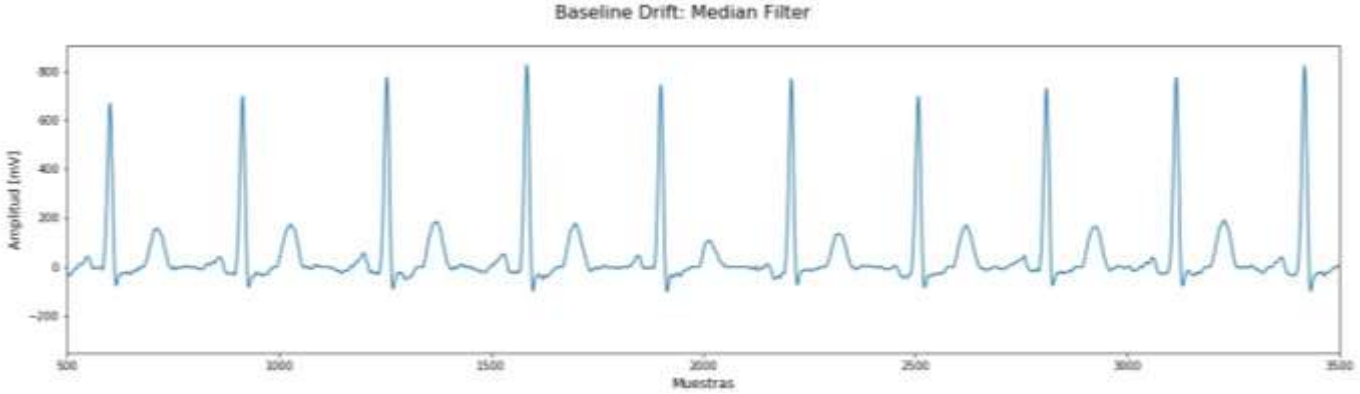

Figura 3.11. Reducción de la línea de interferencia base mediante un filtro de corte banda.

En la figura 3.11 se puede observar el resultado tras aplicar un filtro de corte banda en el rango de las bajas frecuencias mencionadas, $[0.5, 3]\ Hz$. La principal diferencia entre las señales resultantes tras la aplicación de sus respectivos filtros es la modificación del segmento ST. El filtro de corte banda realiza una supresión más efectiva de la línea de interferencia basal; sin embargo, esta técnica de filtrado también suprime ciertas características frecuenciales del segmento ST que pueden llegar a ser determinantes para el diagnóstico de patologías cardiacas relacionadas con este segmento.

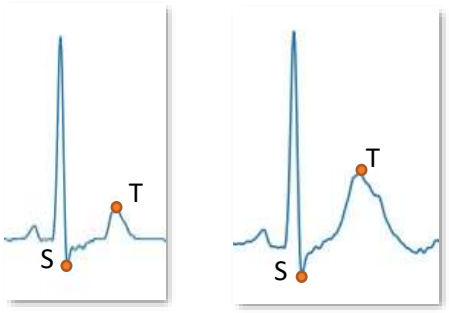

Figura 3.12. Comparación de los resultados obtenidos tras la aplicación de un filtro de corte banda (figura izquierda) y el filtro Butterworth diseñado (figura derecha) de la misma señal ECG original en relación con la modificación realizada sobre el segmento ST.

3.1.2.3. TÉCNICAS DE REDUCCIÓN DE RUIDO

Según [22] y las técnicas descritas en el apartado 2.2.3., el ruido en las señales ECG pueden llegar a afectar notoriamente en su posterior análisis, derivando en un falsa o errónea interpretación de las afecciones cardiacas.

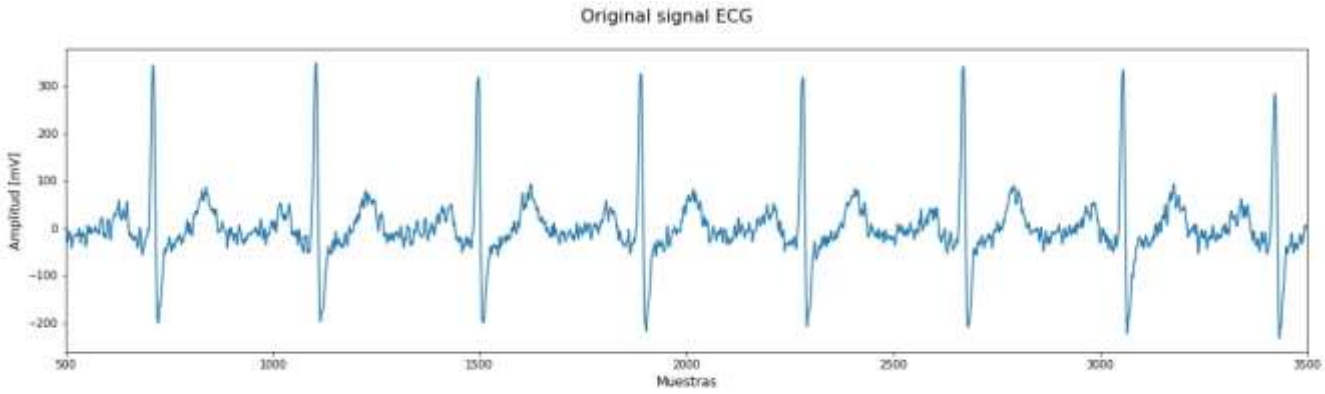

Figura 3.13. Señal de electrocardiograma tras haber eliminado la línea de interferencia basal y sin haber realizado la etapa posterior de reducción de ruido.

En esta etapa se necesita un filtro que sea adaptable al posible abanico de bandas de frecuencias y para cualquier tipo de señal. Debido a esto, la técnica que más ampliamente se utiliza, incluido en el sector sanitario para la reducción de ruido de señales ECG, es DWT. Esta técnica de procesamiento incluye tres escenarios: la descomposición de la señal, la identificación de los coeficientes de mínima energía y su supresión y, finalmente, la reconstrucción de la señal con los nuevos coeficientes.

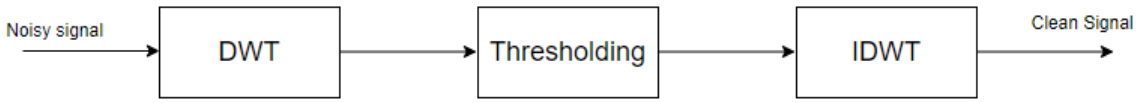

Figura 3.14. Proceso llevado a cabo para la reducción de ruido de las señales ECG [22].

Teniendo en cuenta el proceso de la figura anterior y adaptado al procesamiento de señales ECG, la señal ECG original, '$ECG_{original}$', considerada en la figura como *Noisy signal*, se puede descomponer según la siguiente relación, ECUACIÓN, en '$ECG_{filtered}$', que es la señal resultante tras haber aplicado las técnicas de filtrado y 'ECG_noise', que es el ruido que pertenece a la señal:

$$ECG_{original} = ECG_{filtered} + ECG_{noise} \tag{3.1}$$

Al tratarse de un sistema lineal, los coeficientes *wavelet* corresponden con la aplicación de la técnica DWT a cada factor de la ecuación anterior:

$$DWT(ECG_{original}) = DWT(ECG_{filtered}) + DWT(ECG_{noise}) \quad (3.2)$$

Sin embargo, las características en tiempo y frecuencia pueden variar drásticamente entre las señales ECG de distintos pacientes, por lo que, según [21], es necesario establecer un método que identifique las mejores condiciones de procesado. Para evaluar el rendimiento de la técnica implementada y de la calidad de la señal ECG filtrada que se utilizará como input de la red neuronal se utiliza *Root Mean Square Percentage (PRD)*, siendo efectivo el procesado si la medida del rendimiento es alta y poco efectivo si es baja:

$$PRD = \sqrt{\frac{\sum\left(DWT(ECG_{original}) - DWT(ECG_{filtered})\right)^2}{\sum\left(DWT(ECG_{original})^2\right)}} \quad (3.3)$$

El procedimiento general del análisis utilizando wavelets se basa en emplear una función denominada comúnmente *wavelet mother* o modelo, con la que, el análisis temporal realizado en las resoluciones de las técnicas de procesado *wavelet*, produciendo traslación y escalado sobre la señal original, se basan en la utilización de esta función. Esta función es un parámetro que hay que seleccionar entre todos los tipos de familias *wavelet* existentes. Su elección no es única y depende del tipo de funciones o datos que se van a analizar, ya que una elección adecuada de esta función desemboca en una buena efectividad de la técnica a utilizar. Por lo tanto, dado que la elección de la *mother wavelet* se escoge en función de los datos de entrada, para el análisis y procesado de señales ECG, las *Daubechian* son las recomendadas [24], especialmente *Daubechian db3* y *db4.*

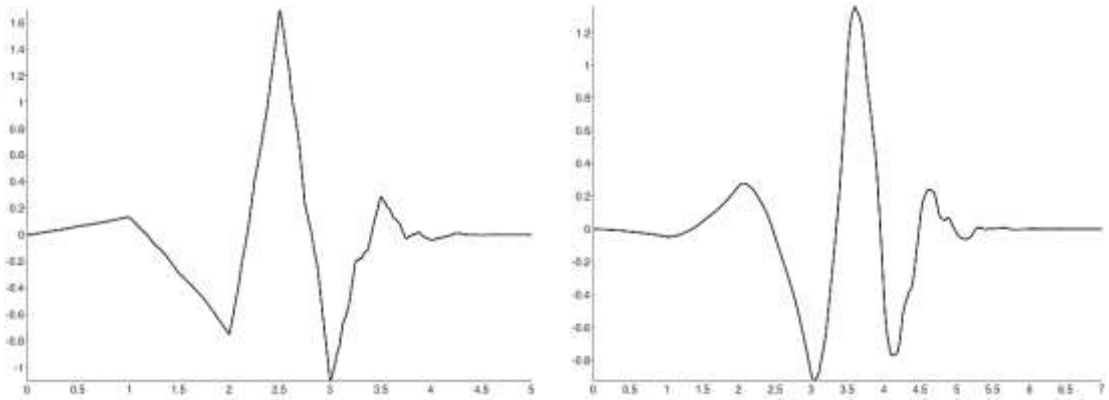

Figura 3.15. Daubechians db3 (figura izquierda) y Daubechian db4 (figura derecha) utilizadas en la aplicación de Discrete Wavelet Transform sobre señales de electrocardiograma.

Una vez que se obtienen los coeficientes proporcionados por *Discrete Wavelet Transform* se aplica un umbralizado según el *threshold* aplicado. El objetivo de establecer un umbral es identificar las señales externas que corresponden a ruido blanco mediante su energía, ya que esta es escasa con baja amplitud. Entonces los coeficientes de la señal obtenida, ECG_{noise}, con baja amplitud corresponden al ruido de la señal. Este ruido se puede eliminar con un umbral donde, los valores inferiores al límite establecido se ponen a cero.

Posteriormente, tras aplicar la técnica de umbralizado se debe reconstruir la señal hasta obtener la señal inicial por medio de las *wavelet*. Finalmente se obtiene una señal resultante, figura 3.16, a partir del filtrado de la señal ECG original, figura 3.13. La señal resultante se utilizará en los siguientes apartados de extracción de características.

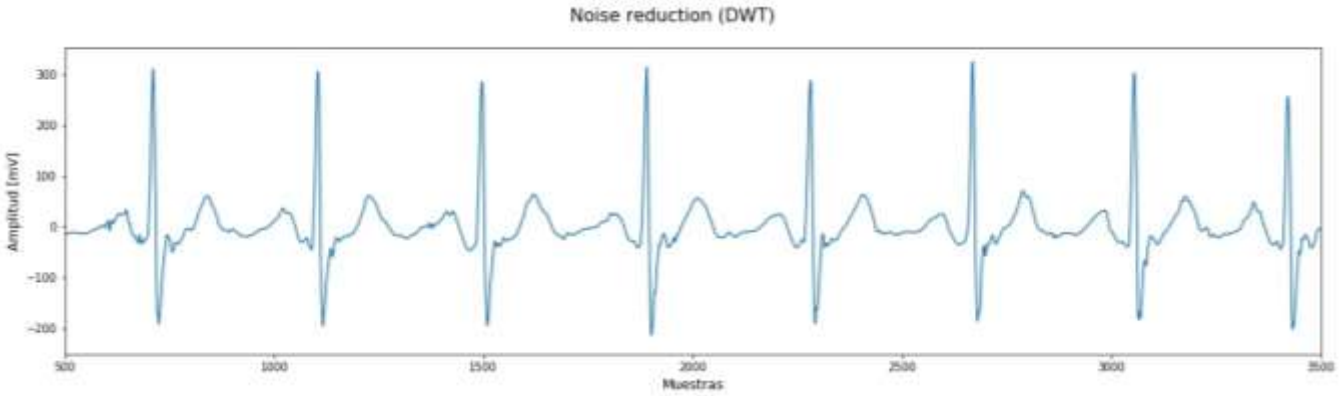

Figura 3.16. Señal de electrocardiograma de la figura 3.13 tras haber aplicado la reducción de ruido.

3.1.3.EXTRACCIÓN DE LAS CARACTERÍSTICAS

Siguiendo la línea planteada en la metodología, la siguiente etapa corresponde con la extracción de características de la señal. En concreto se definirán las técnicas de escalado llevadas a cabo y la segmentación de latidos propia del presente estudio.

Por otro lado, el aumento de datos se explicará más adelante, dentro de los análisis planteados.

3.1.3.1. TÉCNICAS DE ESCALADO

En la etapa de estudio de técnicas de escalado se mencionan la normalización y la estandarización. La normalización es una técnica que se utiliza cuando la distribución de los datos no sigue una distribución gaussiana, previo a la aplicación de algoritmos como K-NN o Redes Neuronales. Por otro lado, la estandarización puede ser útil en los casos en que los datos siguen una distribución gaussiana, no teniendo por qué ser necesariamente cierto. A diferencia de la normalización, la estandarización no tiene un rango límite. Por lo tanto, incluso si existen valores atípicos en los datos, no se verían afectados por la aplicación de esta técnica de escalado.

Siguiendo las diferencias mencionados, se hace uso de la técnica de estandarización, *Z-score*, dado que las señales ECG contienen valores atípicos, no erróneos, propios del ámbito de análisis de las señales y de las patologías analizadas. Otro distintivo por el que se hace uso de la estandarización para este tipo de aplicación es el problema que genera la utilización de técnicas de normalización con la comprensión de los datos de entrada en los límites expuestos. Este problema deriva en una ampliación del ruido propio de señal, característico por su composición de altas frecuencias e imposible de suprimir mediante técnicas de preprocesamiento sin afectar a la morfología de la señal.

3.1.3.2. SEGMENTACIÓN DE LATIDOS

La etapa más importante de la extracción de características reside en la segmentación de los latidos cardiacos de las señales. La longitud de los registros de estas señales varía en función de la duración de las grabaciones, véase la tabla 2. Utilizando el método descrito de segmentación de latidos en el apartado de extracción de características, se divide la señal en varios segmentos de latido, conociendo el intervalo RR, donde cada uno de estos segmentos se compondrá por un PQRST de la propia señal, siempre y cuando su patología o grupo de patologías correspondientes no afecten sobre la morfología de la señal.

La implementación de este método de detección de los picos R puede llevarse a cabo por una librería en Python denominada Biosppy. Esta librería admite una señal ECG como entrada a la función y la frecuencia de muestreo correspondiente, y proporciona los intervalos RR junto con los segmentos de latidos. Sin embargo, este método puede llegar a proporcionar problemas con señales previamente no procesadas, ya que se dificulta la localización del pico R. La función tratará de seleccionar los picos R con mayor resolución de la señal.

Los segmentos extraídos de las señales se compondrán del mismo número de muestras; por lo, se escogerá el mismo número de muestras para su posterior alimentación en la arquitectura de la red convolucional. De esta manera, se omite cualquier técnica de remuestreo en las señales originales y se evita la posible pérdida de información que esta conlleva. La segmentación de la señal en PQRST de un número de muestras significativamente inferior al de una señal completa ayuda, adicionalmente, de manera computacional a la tarea de clasificación de la red neuronal, además de focalizar la información de una señal propuesta en el aprendizaje de la red neuronal convolucional 1D.

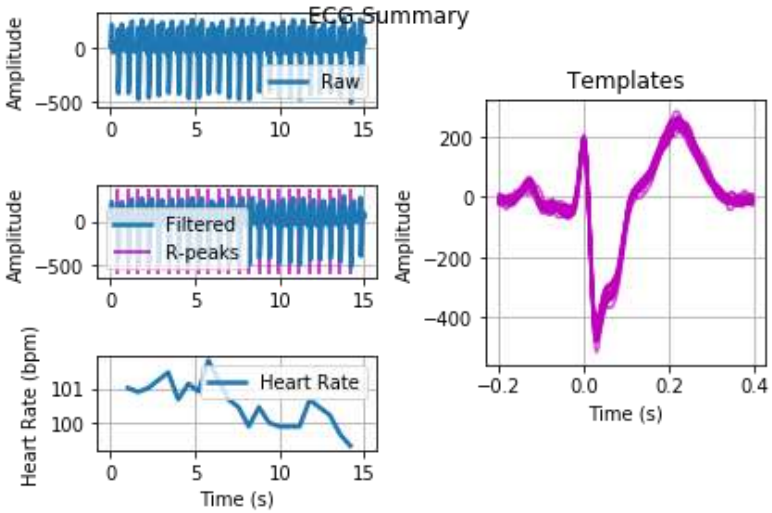

Figura 3.17. Segmentación realizada por el algoritmo de segmentación de latidos explicado (Librería Bioppsy).

La implementación del método de segmentación busca en la señal ECG previamente procesada, los picos R y sus intervalos para proceder a la segmentación de la señal en función de la frecuencia de muestreo proporcionado, tal y como se puede observar en la figura anterior. En la figura anterior, a la derecha se puede observar un ejemplo de una segmentación obtenida tras el uso de esta técnica.

3.1.4.RED CONVOLUCIONAL

Una vez que los datos se han procesado con las técnicas de preprocesamiento de señal definidas, se utilizan las señales ECG resultantes para alimentar la red neuronal convolucional. En un primer análisis se omitió la extracción de características y se utilizó una red neuronal convolucional 1D diseñada a partir de la *ResNet*. La *ResNet*, abreviatura de Redes Neuronales Residuales, son redes neuronales inspiradas en el hecho biológico de que algunas neuronas se conectan con neuronas en capas no necesariamente contiguas, saltando capas intermedias. Este tipo de red fue la red ganadora de una competición en 2015 dado el impacto generado al permitir por primera vez que se pudieran entrenar redes muy profundas, de más de 100 capas, controlando con éxito el problema del desvanecimiento de gradiente o *vanishing gradient*[4].

La arquitectura ofrecida por este tipo de red hace que sea idónea para el problema planteado, siendo necesario un entrenamiento que aproxime las señales a las características extraídas de las patologías estudiadas con el mínimo error posible. Por esto, ResNet proporciona un rendimiento superior a muchas arquitecturas de aprendizaje profundo. Su propiedad de establecer conexiones entre capas no contiguas, además, permite que el modelo sea capaz de aprender una función de identidad que garantice que la capa superior funcionará al menos tan bien como la capa inferior, y nunca peor.

La red *ResNet* viene definida por bloques repetitivos donde cada bloque está compuesto por una capa convolucional, una capa de normalización, generalmente por lotes, denominada *Batch Normalization* y una función de activación. Uno de los aspectos de utilizar una red como *ResNet* ante un conjunto de datos pequeño es el sobreajuste en el entrenamiento. Para evitar un sobreajuste de la red sobre los datos de entrenamiento se añadieron varias capas adicionales en los bloques descritos a las ya existentes:

- **Capa convolucional 1D.**

- **Batch Normalization:** La capa de normalización, definida en el bloque de la ResNet, se utiliza como una técnica de ayuda al entrenamiento, tratando de normalizar las activaciones de salida de cada capa convolucional.

- **Dropout**: Esta técnica se utiliza como un complemento a la regularización del sistema para evitar un sobreajuste en el entrenamiento. Por cada nueva entrada a la red en la fase de entrenamiento, el dropout desactivará aleatoriamente un porcentaje de las neuronas en cada capa oculta, acorde a una probabilidad establecida. No se recomienda utilizar una probabilidad de desactivación superior al 50%. En este caso, la probabilidad asignada ha sido la cifra límite, 50%, con el fin de evitar lo máximo posible el sobreajuste de la red. De esta manera se consigue evitar que ninguna neurona se capaz de memorizar parte de los datos.

- **Gaussian Noise o Ruido Gaussiano:** Esta capa es muy útil para mitigar al máximo el posible sobreajuste de la red neuronal. Este método solo se activa con los datos en el momento de entrenamiento de la red; puede verse como una forma de obtener aumento de datos de forma aleatoria, ya que introduce ruido gaussiano en las señales.

[4] *Vanishing or exploding gradient*, o desvanecimiento del gradiente, es una dificultad encontrada en el entrenamiento de redes neuronales mediante métodos de aprendizaje basados en descenso estocástico de gradientes y retro propagación, donde el gradiente decrece exponencialmente tendiendo a cero o se incrementa tendiendo a infinito.

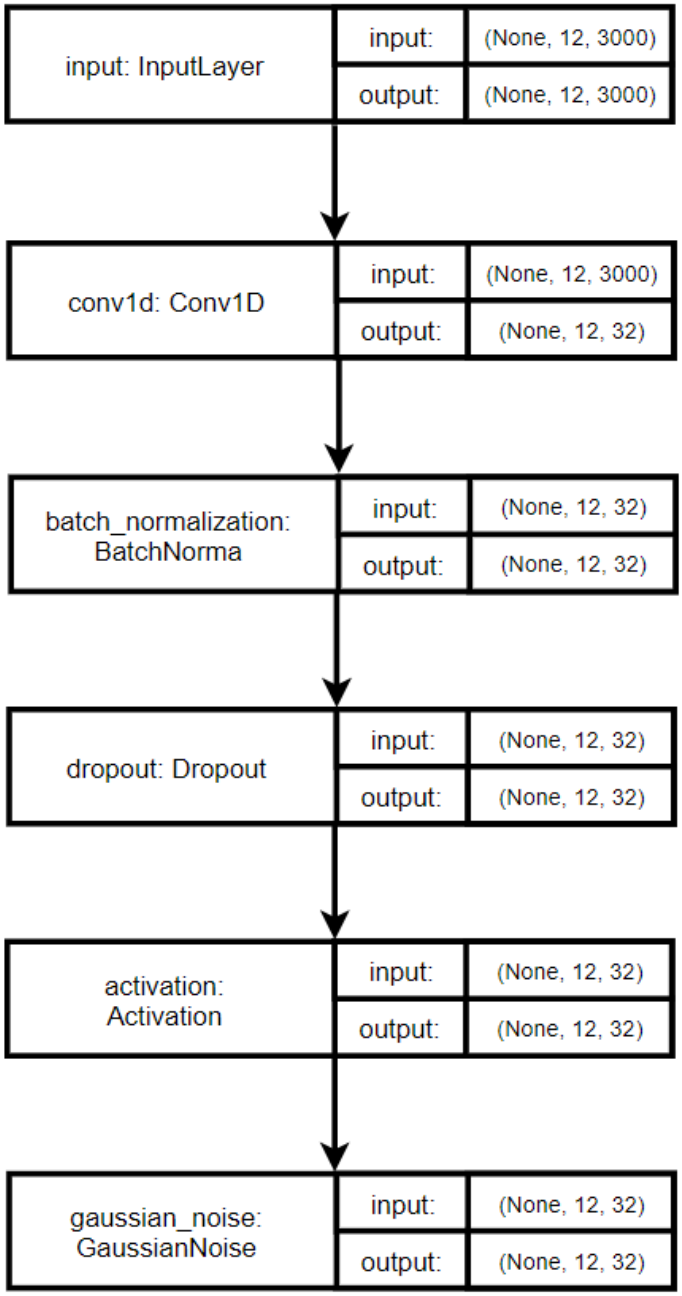

Figura 3.18. Capa de entrada de la red neuronal convolucional 1D conectada con un bloque de la red, similar al bloque de la ResNet.

Finalmente, la red utilizada se compone de 6 bloques con sus respectivas capas convolucionales y cada 2 bloques se realiza la propiedad de las ResNet de saltar la conexión entre capas profundas, tal y como se puede observar en la siguiente figura:

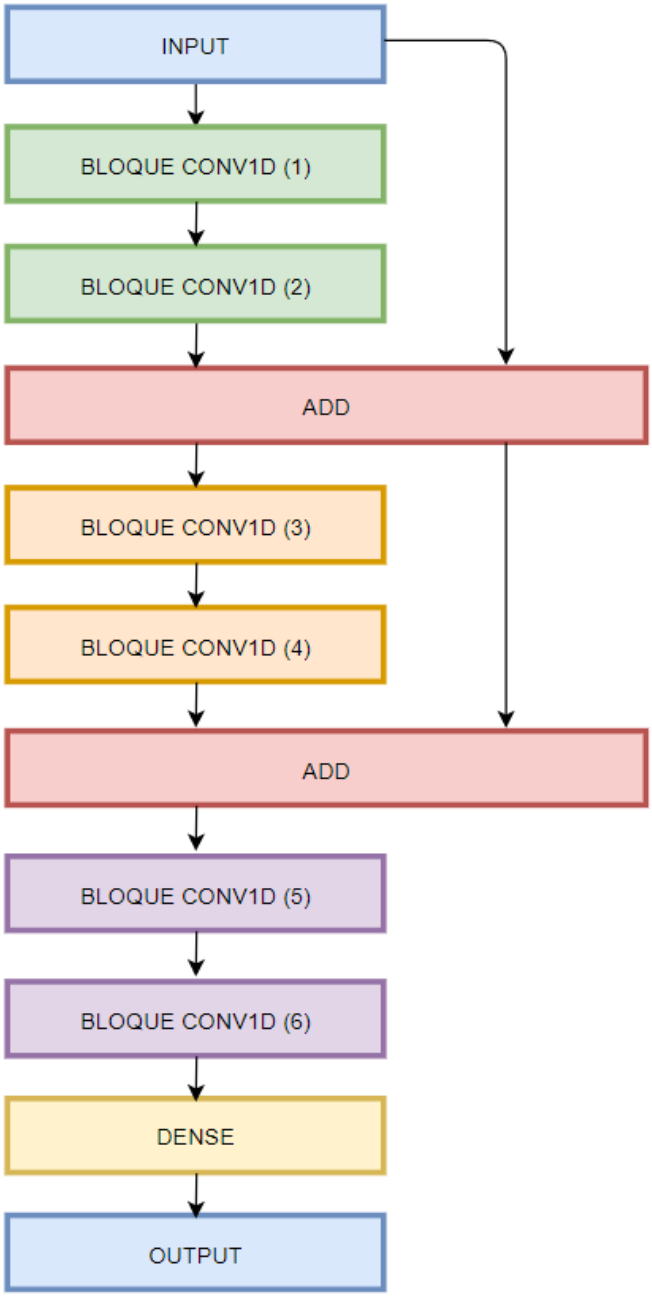

Figura 3.19. Estructura de la Red Neuronal Convolucional 1D utilizada en el primer análisis.

Al finalizar la estructura de la red convolucional, la salida del último bloque convolucional se deriva a una capa densa que establece la clasificación proporcionada por la salida de la red. La red utilizada contiene, previamente a la capa de salida con la función de activación *sigmoid*, se propuso la capa *Global Average Pooling*. La capa *Global Average Pooling* es primordial para aplicar la interpretabilidad utilizando el método de *Class Activation Maps* estudiado y evaluar la importancia de las regiones de las señales de electrocardiograma. Esta capa tiene como objetivo proyectar los pesos de la salida final con

las activaciones de la estructura convolucional. Como diferencia con el modelo presentado de CAM es la función de activación utilizada. En este caso se utiliza una sigmoide para realizar una clasificación multi-etiqueta donde las patologías tienen igual de importancia en el diagnóstico. La ecuación utilizada que sustituye la utilizada por el método, Ecuación 2.15, es:

$$P_c = \frac{1}{1 + \exp(S_c)} \tag{3.4}$$

3.2. ANÁLISIS DESARROLLADOS

Una vez que se ha abordado el desarrollo del trabajo realizado en función de los datos utilizados en el presente estudio, se presentan los resultados obtenidos distribuidos en tres análisis: el primer análisis consta del estudio de una red neuronal convolucional 1D sin una previa extracción de características, el segundo corresponde con el estudio de una red neuronal convolucional 1D con una tarea previa de extracción de características y el tercer análisis desarrollado corresponde con una optimización de la red convolucional empleada para esclarecer los resultados obtenidos.

Primeramente, para llevar a cabo los análisis mencionados se debe realizar una tarea de exploración de los datos que servirán de base para el estudio. Tras la exploración se observó que había una gran cantidad de patologías cardiacas que estaban representadas en el dataset por un pequeño número de pacientes, en el orden de 1 – 10 pacientes; esta casuística podría llegar a derivar en un problema en el entrenamiento de la red, produciendo un sobreajuste de los datos de entrenamiento indeseado. Debido a esto, en una primera instancia, para evitar errores en los análisis, se decidió suprimir aquellas patologías que no llegaran a estar lo suficiente representadas en el dataset; por lo que en el apartado de preprocesamiento de los datos se redujo el dataset a analizar hasta 60 patologías definitivas.

Posteriormente, se representa el dataset para observar la cantidad de señales ECG que pertenecen a cada una de las patologías definitivas:

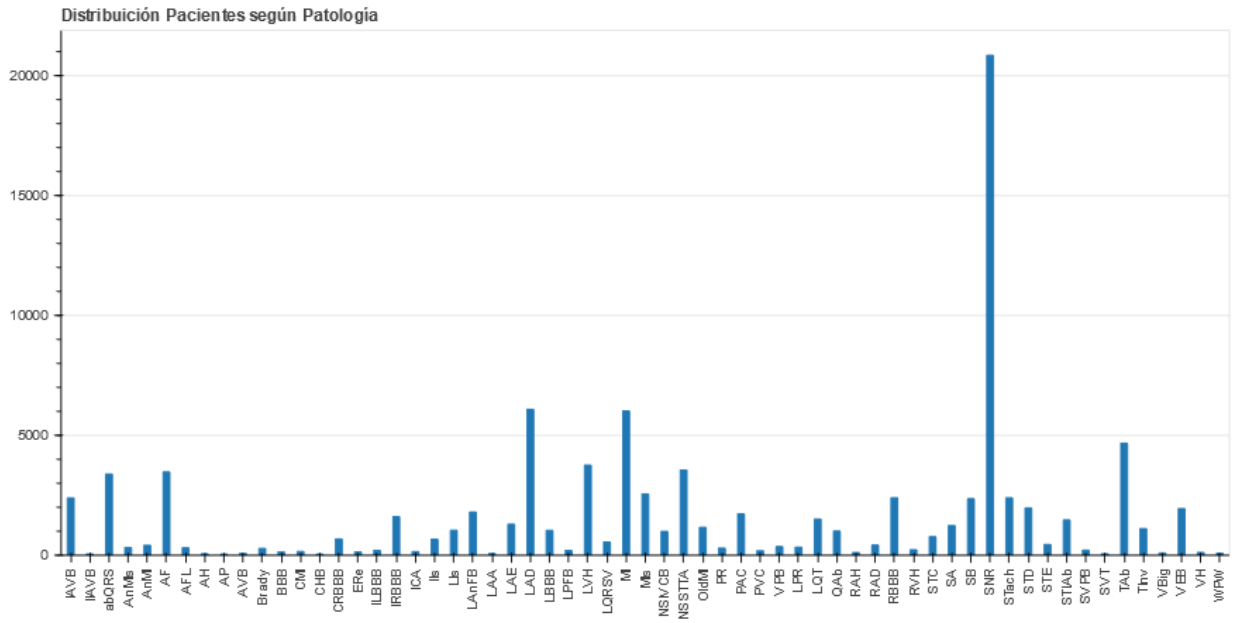

Figura 3.20. Distribución de pacientes según las patologías estudiadas. En el Anexo 1 se puede observar una descripción general de las patologías que se comprenden. Es importante observar que, al ser un problema de multi-etiqueta, hay pacientes que tendrán varias patologías y, por lo tanto, se contabilizará en cada una de ellas.

De la observación de la figura anterior, resulta evidente que el conjunto de datos se encuentra desbalanceado, siendo predominante la clase SNR, ritmo sinusal normal, en el conjunto de datos con alrededor de 20.000 pacientes, mientras que las siguientes clases se encuentran en un orden más bajo, cercano a 5.000 pacientes.

Los conjuntos de datos desbalanceados, comunes en aplicaciones reales, suponen un efecto negativo en el rendimiento de los algoritmos de Machine learning [42]. En la literatura científica se encuentran varios métodos utilizados para corregir este tipo de problemas, entre los que se encuentra el aumento de datos de las clases minoritarias como solución para prevenir el efecto negativo mencionado.

Sin embargo, observando la figura 3.20, es notoria la gran diferencia de la clase predominante, SNR, frente al resto de patologías existentes en el dataset, por lo que un aumento de datos, en estas circunstancias traería una serie de desventajas: realizar un número ilimitado de aumento de datos en un dataset sobre unos datos iniciales con el gran desbalanceo que sufre el dataset de la figura 3.20, donde las clases (patologías cardiacas) minoritarias se constituyen de un pequeño número de señales ECG en comparación con la clase mayoritaria, se corre el riesgo de obtener sobreajuste en el entrenamiento de la red, de manera que se evita la generalización sobre los datos de test.

Debido a esto, previo al desarrollo de los tres análisis llevados a cabo, se realiza una reducción de las señales de electrocardiograma empleadas en el clasificador, dado el predominio de la clase SNR en los datos iniciales, así como del número de patologías, o clases, tenidas en cuenta en el proceso. Al reducir el número de datos considerados como dataset, se obtiene un conjunto de datos más balanceado donde la patología o etiqueta SNR ya no es una clase predominante.

Al ser un problema de multi-etiqueta, es decir, las señales ECG pueden tener una o varias patologías simultáneamente, es muy probable que una señal ECG o, mejor dicho, un paciente, tenga varias patologías cardiacas y una de ellas sea SNR. Si fuera así, al realizar la reducción de esta clase, el paciente no sería procesado y se suprimirían sus otras patologías del dataset. Tras realizar esta supresión de las patologías, el dataset de la figura 3.20 se transforma en el siguiente conjunto de datos:

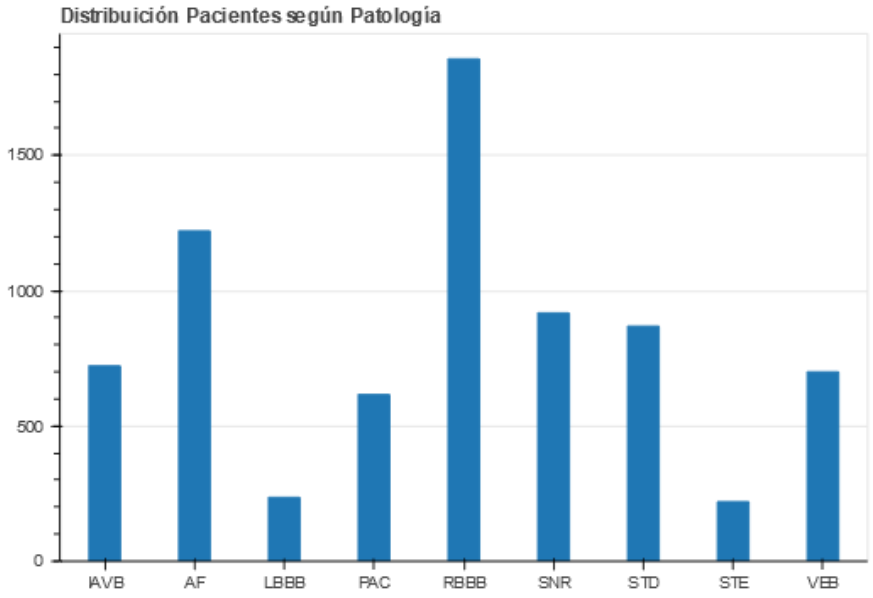

Figura 3.21. Conjunto de datos utilizado tras la reducción de la clase mayoritaria SNR.

Resulta muy interesante observar qué patologías se mantienen en el dataset frente a aquellas que se eliminaron al suprimir la clase mayoritaria SNR y por qué, tras esta transformación, la clase SNR ya no es la clase mayoritaria.

En el Anexo 1 se encuentra una tabla con una descripción general de todas las patologías que se han tenido en cuenta para realizar el presente estudio. Analizando la variedad de patologías descritas, se llega a la conclusión de que el conjunto de patologías se puede dividir en dos grupos principales: Arritmias y No Arritmias. Las arritmias son problemas cardiacos derivados del ritmo del corazón que se originan cuando los impulsos eléctricos que coordinan los latidos cardiacos no funcionan adecuadamente lo que hace que el corazón lata demasiado rápido, lento o de forma irregular. Sin embargo, estas tres características no son necesarias para que una patología cardiaca sea considerada como arritmia, es decir, una patología puede provocar bradicardia, ritmo regular lento, y no ser considerada como arritmia.

Dentro del conjunto de patologías no arrítmicas se encuentra el ritmo sinusal normal, conocido como SNR. El ritmo sinusal normal no es más que una denotación de las patologías cardiacas no arrítmicas por lo que prevalece en aquellos pacientes que carecen de patologías arrítmicas y cuyo ritmo cardiaco es regular. Hay varios criterios que constituyen el ritmo sinusal:

1. Frecuencia cardiaca entre 60 – 100 lat/min.

2. El intervalo del segmento RR debe ser constante (ritmo cardiaco regular).

3. Onda P positiva en la derivación II y negativa en la derivación avR. Seguida de un QRS.

Teniendo en cuenta los criterios mencionados se puede observar que varias de las patologías restantes del dataset pertenecen al grupo de arritmias evitando una simultaneidad de existencia con el ritmo sinusal normal:

4. I-AVB, Bloqueo de primer grado del nodo AV, contradice el criterio de la onda P del ritmo sinusal normal porque, en esta patología, la onda P se caracteriza por una conducción ralentizada entre la aurícula y el ventrículo (atrio ventricular).

5. AF, Fibrilación Auricular, es una patología cardiaca caracterizada por la ausencia de onda P y por representar los intervalos RR de forma irregular.

Por otro lado, varias de las patologías que aparecen en el nuevo dataset no pertenecen al grupo de arritmias; sin embargo, son patologías que se caracterizan porque no necesariamente deben tener un ritmo sinusal normal, aunque sí pueda darse el caso:

6. RBBB y LBBB, Bloqueo de rama derecha y bloqueo de rama izquierda, son patologías cardiacas que no necesariamente deben tener un ritmo sinusal normal, pero pueden coexistir.

7. STE y STD, elevación y disminución del segmento ST, son patologías del grupo de no arritmias; tienen un ritmo sinusal normal cuando acontecen, pero son patologías arritmogénicas, es decir, pueden llegar a generar que cierto territorio cardiaco se quede sin sangre, derivando en un tejido muerto, con ausencia de conducción eléctrica que llegue a provocar una arritmia.

Por último, el grupo que queda es PAC y VEB, complejos prematuros ventriculares y auriculares. Este tipo de patologías son patologías que, en función de su origen, mayoritariamente vienen acompañados de ritmo sinusal normal, pero periódicamente surge un latido ectópico que si es mantenido en el tiempo puede llegar a considerarse una arritmia.

Una vez comprendida la relación entre las patologías restantes del conjunto de datos que se va a analizar, se procede con el primer análisis descrito: aplicación de la red neuronal convolucional 1D sin una etapa previa de extracción de características.

3.2.1. PRIMER ANÁLISIS: CNN1D SIN EXTRACCIÓN DE CARACTERÍSTICAS

El primer análisis realizado sobre el conjunto datos para la discriminación de las patologías y una correcta clasificación se basa en el aprovechamiento de una de las principales características de la aplicación de técnicas de aprendizaje profundo, similares a la red convolucional 1D: la extracción automática de características de los datos iniciales, en este caso, señales de electrocardiograma [5]. Como consecuencia, estos algoritmos son descritos, en muchas ocasiones, como cajas negras. Muchos de los ejemplos o procesos llevados a cabo se realizan teniendo en mente la descripción de la caja negra. Teniendo esto en cuenta, se decide seguir una línea similar en el primer acercamiento. Por lo tanto, a partir del dataset de las señales de electrocardiograma obtenidas tras el balanceo, figura 3.18, se realiza la etapa de preprocesamiento de las señales ECG, se diseña la estructura de una red convolucional 1D, que, se entrena y se valida con los datos previamente procesados, manteniendo un conjunto de datos de entrenamiento del 80% y de test del 20% y, finalmente, se obtiene una clasificación de las señales ECG según sus posibles patologías cardiacas, representada por las siguientes métricas:

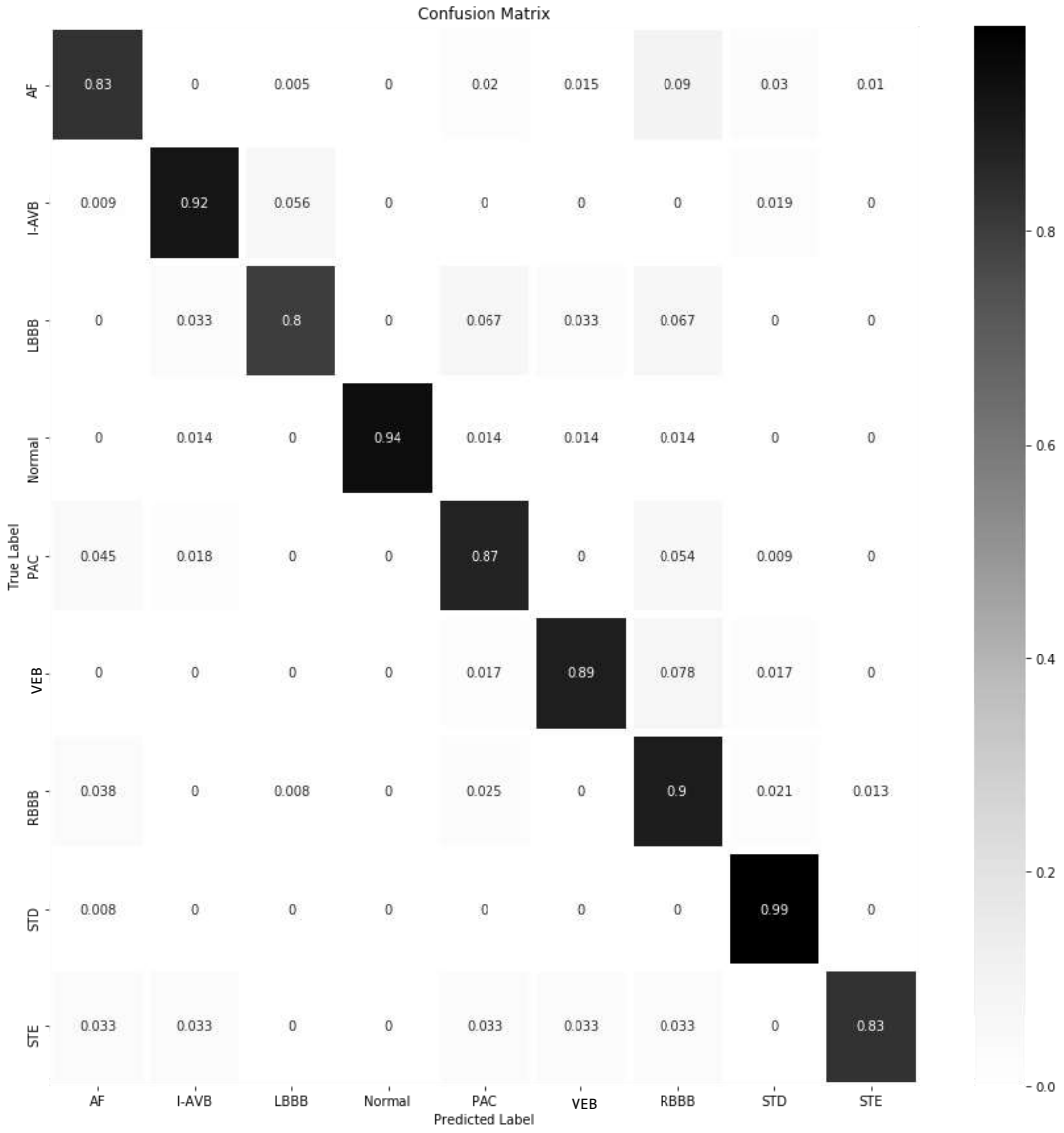

Figura 3.22. Matriz de confusión generada tras el entrenamiento de la red convolucional 1D sin una etapa previa de extracción de características.

Se puede observar que ha resultado un éxito la clasificación de las patologías cardiacas, obteniendo un alto rendimiento por parte de la red y resultando en un algoritmo que, implantado en un entorno real, podría ayudar al personal sanitario a discriminar entre las patologías estudiadas.

Tabla 3. Resultados del primer análisis desarrollado: aplicación de una red neuronal convolucional 1D sin etapa previa de extracción de características. En la tabla se describe la precisión obtenida por el modelo para los datos de entrenamiento y para los datos de test.

EVALUACIÓN DEL MODELO	PORCENTAJE
Evaluación en datos de entrenamiento	96.94%
Evaluación en datos de test	89.51%

Sin embargo, en un entorno con un alto grado de responsabilidad como es el mencionado, es preciso evitar el concepto de caja negra y esclarecer el modelo, pudiendo dar una explicación de la clasificación realizada, tal y como se ha realizado en la etapa de interpretabilidad mostrada en los resultados. Para ello se selecciona una señal ECG de forma aleatoria propia del conjunto de datos test y se comprueba la técnica de interpretabilidad CAM sobre ella para observar la interpretabilidad del modelo. Se utiliza una escala de color definida en la siguiente figura, donde los tonos orientados al blanco sugieren que la red no presta atención mientras que las zonas con tonos orientados al extremo rojo sugieren que para la red tales zonas son significativas para la extracción de información:

Figura 3.23. Escala cromática utilizada

La señal ha sido clasificada correctamente con su respectiva etiqueta como una señal con la patología cardiaca RBBB, bloque de rama derecha.

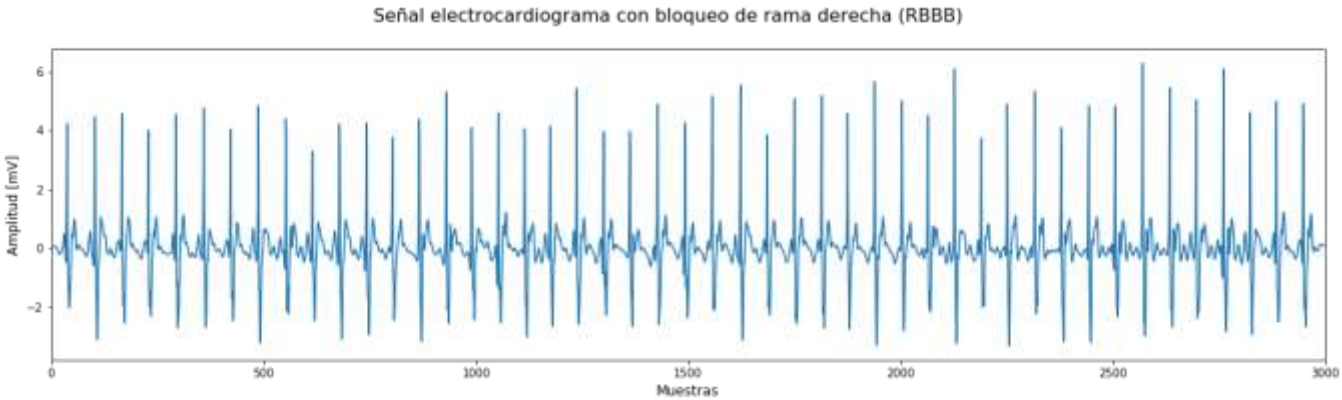

Figura 3.24. Señal ECG clasificada por la red neuronal correctamente.

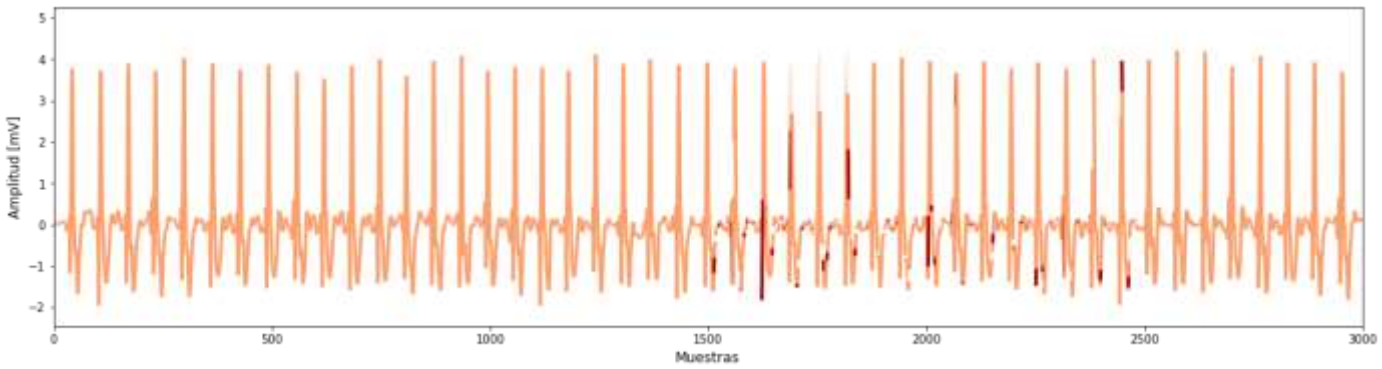

Figura 3.25. Interpretabilidad del modelo planteado sobre una señal ECG cuya patología asociada es RBBB (Bloqueo de rama derecha).

La figura anterior sería utilizada por el personal sanitario para esclarecer el proceso del modelo en su tarea de clasificación de patologías. Con los requisitos de interpretabilidad planteados, la extracción de características propuesta por el modelo no se ajusta a las características de bloqueo de rama derecha, véase en el Anexo 1. El modelo refleja una extracción de características lejos de la realizada por un personal sanitario al utilizar partes de la señal que no son significativas, desde el punto de vista tradicional clínico, para su clasificación con su respectiva patología cardiaca. El modelo se centra en un rango de la señal, en el intervalo [1500, 2000] para extraer información para establecer la clasificación de la señal cuando esta patología mostrada se caracteriza por existir en todos los latidos de la señal. No obstante, las métricas del modelo obtenidas como resultado final prometen unos magníficos resultados, lo que indica que el modelo es capaz de discriminar estas patologías, tal vez no desde un punto de vista clínico, pero si ante características ocultas en la señal que, un profesional especializado, podría evaluar.

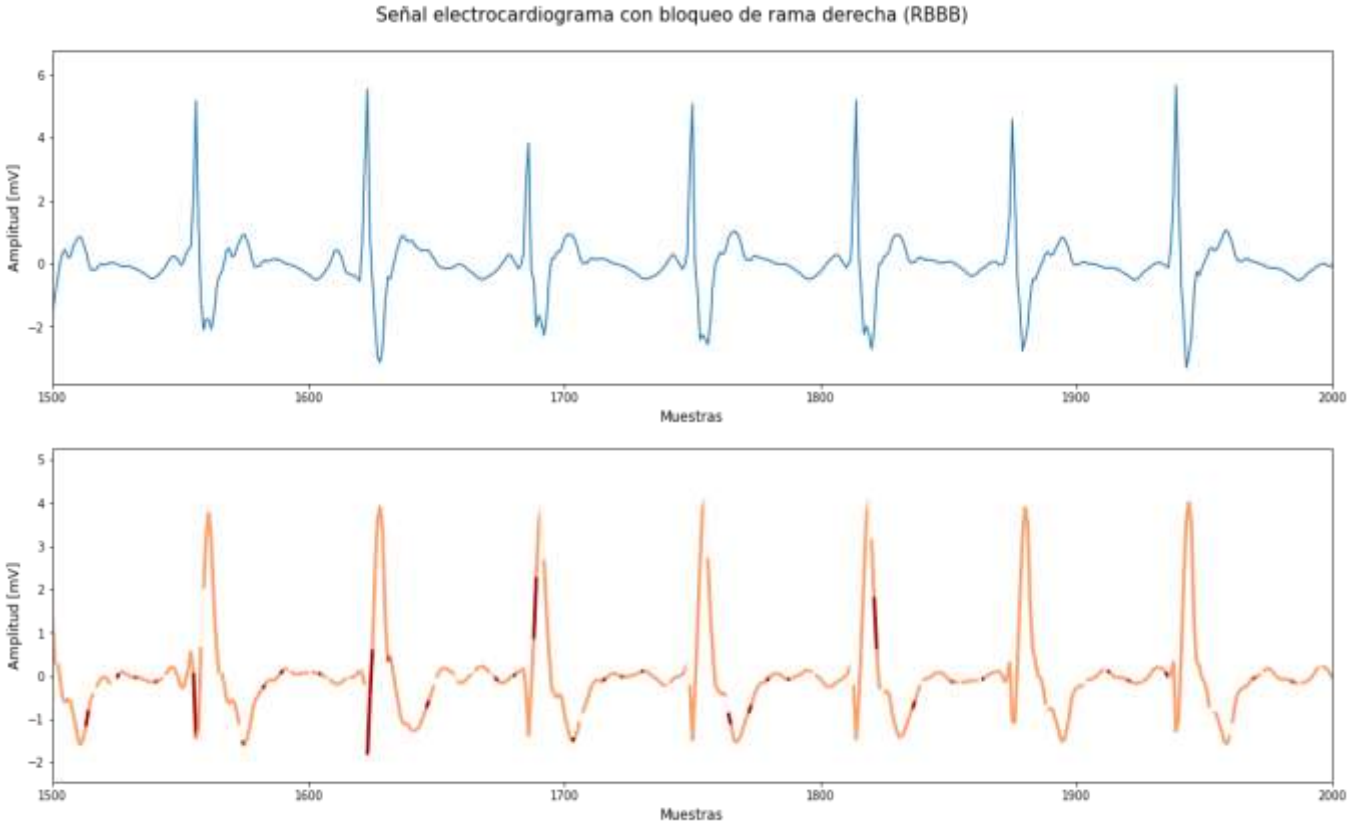

Figura 3.26. Figura obtenida tras realizar un aumento de la señal de la figura 3.19. Se visualizan una parte de la señal encontrada entre la muestra 1500 y la 2000.

En esta ocasión, para el problema planteado en el presente trabajo fin de máster, teniendo en cuenta que el tratamiento de las patologías cardiacas no es un proceso trivial como realizar un clasificador común, hoy en día, se ha decidido realizar un segundo análisis con una etapa de extracción de características que complemente las técnicas de aprendizaje profundo, con el fin de proporcionar a la red convolucional información más detallada de las señales ECG, evitando tratar los algoritmos como cajas negras y ayudando, desde el punto de vista de la ingeniería, al algoritmo en las fases previas. Así, se tratará cumplir con los requisitos de interpretabilidad planteados al comienzo del estudio: tratar de obtener una explicación de los algoritmos lo más acorde al punto de vista clínico que produzca una mayor aceptación en su uso.

3.2.2. SEGUNDO ANÁLISIS: EXTRACCIÓN DE CARACTERÍSTICAS

En vista de los resultados obtenidos en el anterior análisis, se decide tener en cuenta las técnicas de extracción de características. La extracción de características realizada permite que la red aprenda sobre un segmento PQRST, tal y como haría un personal sanitario para discriminar entre las patologías morfológicas de la señal. Entre las patologías finalmente estudiadas se ha observado que un gran porcentaje de ellas son representadas por todos los PQRST de las señales y en cada una de las 12 derivaciones, es decir, analizando un latido aleatorio de manera independiente de la señal se podría averiguar la patología cardiaca del paciente. Esta característica, fruto de la casualidad, no se observa en todas las patologías cardiacas no estudiadas en esta tesis. Hay patologías cardiacas que pertenecen a latidos ectópicos, como el caso de los complejos prematuros ventriculares y auriculares, o que vienen representadas morfológicamente en ciertas derivaciones mientras que las restantes pueden representar un electrocardiograma normal, o que vienen representadas por amplitud positiva o negativa de las señales, etc.

La etapa de extracción de características se caracteriza por el uso de la técnica de segmentación de latidos mencionada. Tras la aplicación de la segmentación, los datos iniciales se reducen, de tal manera que cada señal ECG se queda representada por una pequeña serie de latidos. No obstante, el uso de técnicas de aprendizaje profundo requiere de una gran cantidad de datos, por lo que resulta necesario aplicar técnicas de aumento de datos para el correcto análisis del dataset.

3.2.2.1. AUMENTO DE DATOS

Para realizar un aumento de datos se siguen las indicaciones teóricas dadas en la metodología, teniendo en cuenta que se desea evitar incidir sobre la representación morfológica de la señal o afectar en zonas concretas de la misma que puedan suprimir características frecuenciales importantes utilizadas para su clasificación. Es por ello por lo que una premisa dada en este punto es utilizar técnicas que no varíen partes concretas de la señal ni modifiquen las relaciones morfológicas que estas contienen. Una de las técnicas utilizadas es la alteración de la relación entre la librería descrita en la segmentación de latidos con la respectiva frecuencia de muestreo de la señal electrocardiograma. Las señales, dependiendo de su fuente de datos de procedencia, varían en la frecuencia de muestreo utilizada en su proceso de registro y el uso de esta frecuencia es muy importante de cara a realizar la segmentación y evitar una pérdida de información.

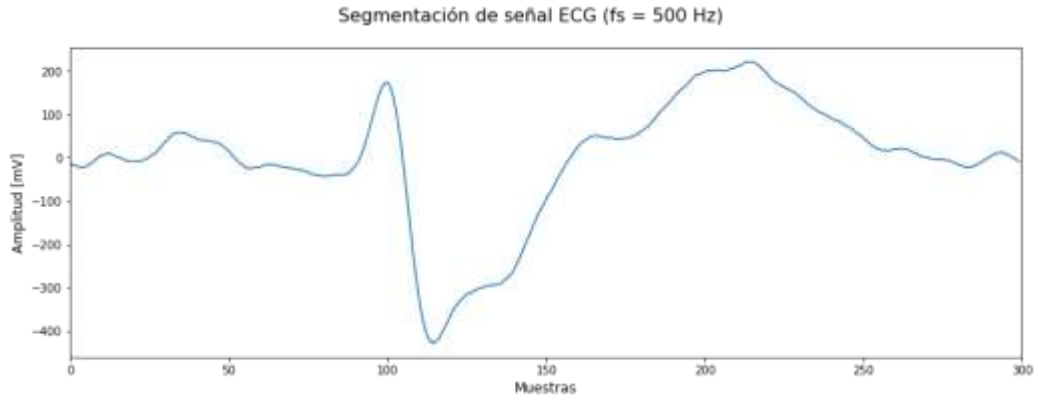

Figura 3.27. Segmentación de latido realizada a una señal ECG. El segmento muestra un único PQRST de una derivación.

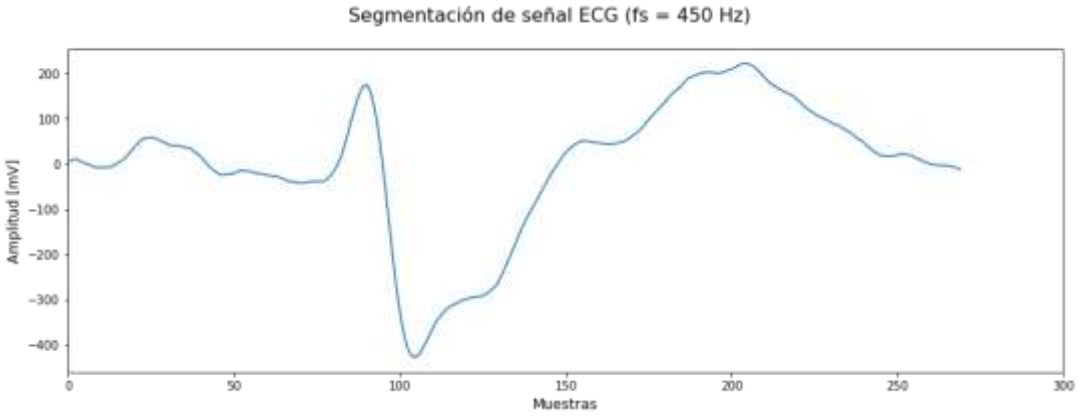

Figura 3.28. Segmentación de la señal ECG anterior alterando la frecuencia de muestreo en 50 Hz.

Esta librería, como se ha explicado, detecta los intervalos R-R entre los QRS más definidos de la señal y segmenta la señal para obtener su PQRST. Sin embargo, en la segmentación de las señales suele existir un tramo inicial (antes de la onda P) y un tramo final (posterior a la onda T) que carecen de significado relevante para la identificación de la patología cardiaca. Por lo tanto, se podría acortar la señal por los extremos hasta un valor límite sin suprimir información de la señal de cara a su posterior análisis.

Es por ello por lo que una de las técnicas mostradas aquí es la modificación de la frecuencia de muestreo en valores entre [20, 50 Hz] sobre la original; parámetro que modifica el enventanado, número de muestras, obtenido tras la segmentación de latido, como en la figura anterior.

Con el fin de obtener unos datos homogéneos en número de muestras, se realiza la técnica *'resample'* sobre el segmento de señal. La utilización de esta técnica actúa sobre la señal al completo, evitando la deformación morfológica de zonas específicas de las señales ECG.

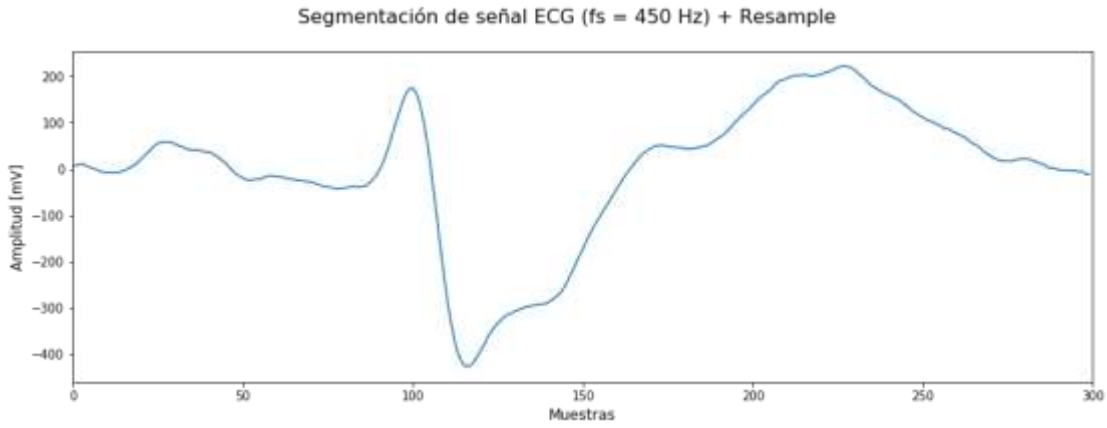

Figura 3.29. Segmentación de la señal ECG de la figura 3.18 a una frecuencia de 450 Hz + técnica de resample para obtener la señal de 300 muestras.

Otro método empleado para hacer un aumento de datos a partir de la modificación de la frecuencia de muestreo es la utilización de la técnica *'zero padding'*. Esta técnica infiere del mismo modo que la técnica de *'resample'*. Se utiliza con el fin de obtener el mismo número de muestras final en los datos del enventanado.

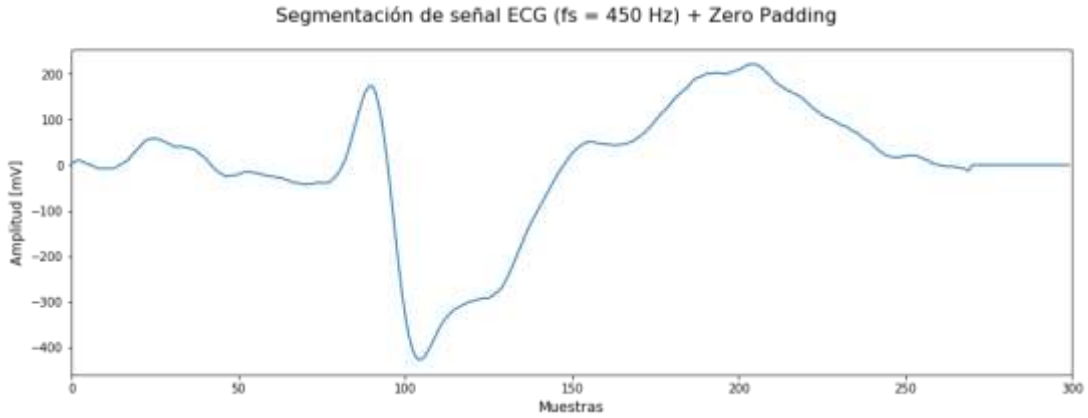

Figura 3.30. Segmentación de la señal ECG de la figura 3.18 a una frecuencia de 450 Hz + técnica de zero padding para obtener la señal de 300 muestras.

El aumento de datos no se realiza en todas las señales por igual, sino que se realiza en función del número de señales ECG correspondientes a cada categoría. Por ejemplo, teniendo en cuenta la visualización del conjunto de datos que se va a analizar, Fig. 3.18, se puede observar que la clase RBBB, Bloqueo de rama derecha, es la clase mayoritaria en el dataset. Por lo tanto, en el procesamiento de señales llevado a cabo, aquellas señales que no pertenezcan a esta patología serán utilizadas en las técnicas de aumento de datos. En el momento que otra patología llegue al mismo nivel que la clase mayoritaria, se dejará de aplicar la técnica de aumento de datos sobre ella, hasta finalmente, obtener un conjunto de datos balanceado.

Una vez que se realiza el procesamiento del dataset, se utilizan las señales resultantes para alimentar la red neuronal convolucional diseñada. Previo a su alimentación, los datos obtenidos se deben dividir en dos subconjuntos: datos de entrenamiento y datos de test.

3.2.2.2. RED CONVOLUCIONAL 1D: MULTI-ENTRADA

La arquitectura de la red neuronal, en el presente análisis correspondiente con la extracción de características, se transforma en una red neuronal convolucional 1D multi-entrada, siguiendo con el mismo estructura principal de la ResNet planteada, aspecto que permite mantener la complejidad de la red neuronal, con un gran número de capas, y que evita problemas como los derivados de *vanihing or exploding gradients* que la propia red inicial no pudiera. En [43] se observó que una red convolucional multi-entrada proponía un mayor rendimiento en el entrenamiento de la red frente a un conjunto de datos formado por varios grupos con ciertas características en común. Por lo tanto, con el fin de aislar las derivaciones de la señales con su respectivo grupo, se propone la red multi-entrada con tantas entradas como derivaciones existan en una señal ECG.

Para evitar un sobreajuste de los datos de entrenamiento se redujeron los bloques convolucionales. En la siguiente figura se puede observar la modificación realizada sobre la estructura general de la red convolucional:

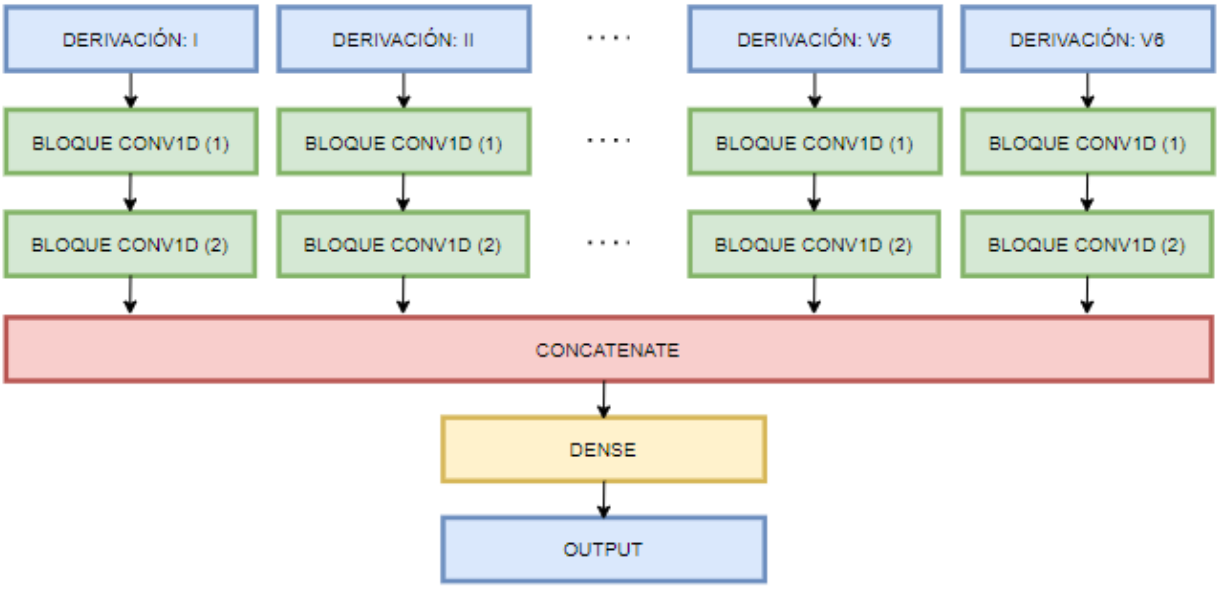

Figura 3.31. Esquema general de la red utilizada en el segundo análisis con multi-entrada.

La red se alimenta por grupos de latidos de cada derivación que conforman una señal ECG. Los datos de entrenamiento, utilizados para realizar el entrenamiento de la red, comprenden el 80% del conjunto de datos total. Estos datos se componen de señales ECG procesadas entre los que se encuentran todos los aumentos de datos obtenidos. Por otro lado, los datos de test corresponden con el 20% del conjunto de datos total y se comprende de señales originales procesadas. Esta diferenciación se establece con el objetivo de testear señales de la forma más parecida posible a como se haría con las señales que se procesarían en un entorno sanitario real.

3.2.2.3. RESULTADOS OBTENIDOS DEL SEGUNDO ANÁLISIS

Posteriormente, en la siguiente figura se muestra la matriz de confusión obtenida como resultado del modelo. La métrica obtenida del modelo viene recogida en la tabla 4. Los datos ofrecen una evaluación del modelo lejos del rendimiento proporcionado por el análisis anterior, con apenas un 70% de precisión en los datos de test. Sin embargo, la estructura del modelo da buenos resultados, evitando un sobreajuste de los datos de entrenamiento frente a los datos de test.

Tabla 4. Resultados del segundo análisis desarrollado: aplicación de una red neuronal convolucional 1D con una etapa previa de extracción de características. En la tabla se describe la precisión obtenida por el modelo para los datos de entrenamiento y para los datos de test.

EVALUACIÓN DEL MODELO	PORCENTAJE
Evaluación en datos de entrenamiento	70.18%
Evaluación en datos de test	67.22%

Particularmente, la matriz de confusión muestra una situación característica: la red convolucional es incapaz de generalizar para dos patologías en particular, PAC, complejo prematuro auricular, y PVC, complejo prematuro ventricular; sin embargo, para el resto de las patologías parece haber entrenado correctamente, llegando a generalizar los datos y estableciendo diferencias entre ellas.

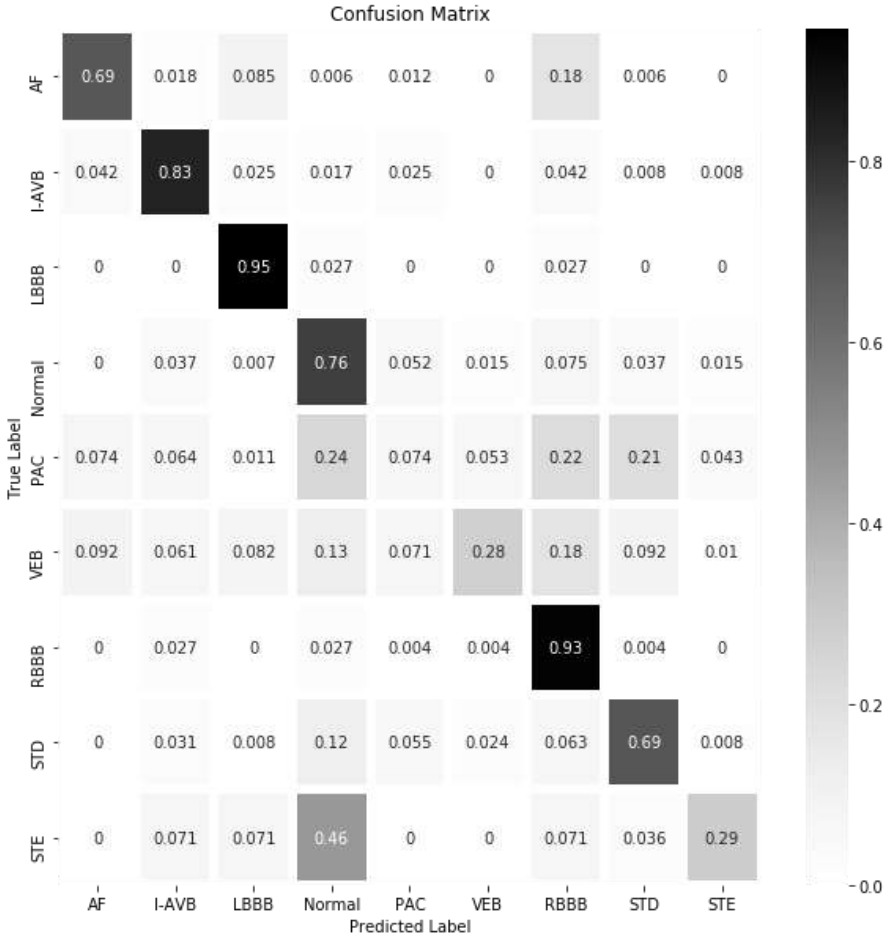

Figura 3.32. Matriz de confusión obtenida de los resultados de la red neuronal convolucional 1D multi-entrada.

Esta particularidad se debe a una característica importante de los complejos prematuros ventriculares y auriculares, VEB y PAC respectivamente, citada anteriormente: los complejos prematuros ventriculares y auriculares vienen dados por latidos ectópicos. Los latidos ectópicos son latidos que aparecen una vez cada cierto periodo de tiempo. Es por ello por lo que, al estar constituidas por este tipo de latidos, se manifiesta que el resto de la señal ECG registrada mostrará una señal con ritmo sinusal normal o, en su defecto, con otro tipo de patología cardiaca.

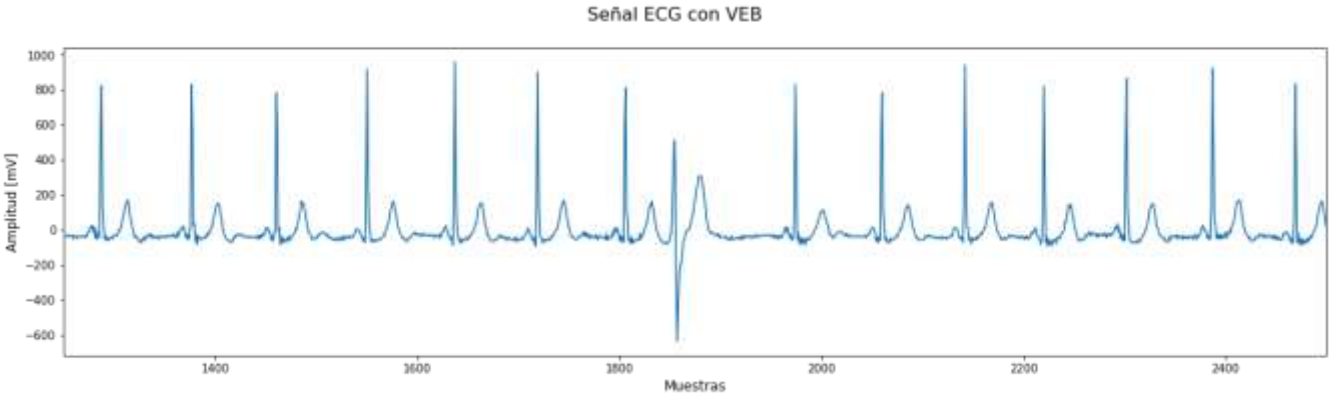

Figura 3.33. Derivación I de una señal ECG cuyo paciente tiene complejo ventricular prematuro (VEB). El latido ectópico se puede observar entre la muestra 1000-1500.

En la etapa de segmentación de latidos se localizan los picos R más fiables de la señal, realizando un enventanado en función de la frecuencia de muestreo del PQRST de la señal, correspondiente con un latido. Si ninguno de los latidos segmentados contiene un latido ectópico propio de las patologías descritas, la información resultante de la señal no será significativa para su posterior análisis, derivando en una clasificación errónea por parte de la red convolucional.

3.2.3. TERCER ANÁLISIS: SUPRESIÓN DE PAC Y PVC

En el apartado de extracción de características del procesamiento de señales ECG, se realiza una segmentación de latidos para analizar las patologías cardiacas en función de cada latido cardiaco.

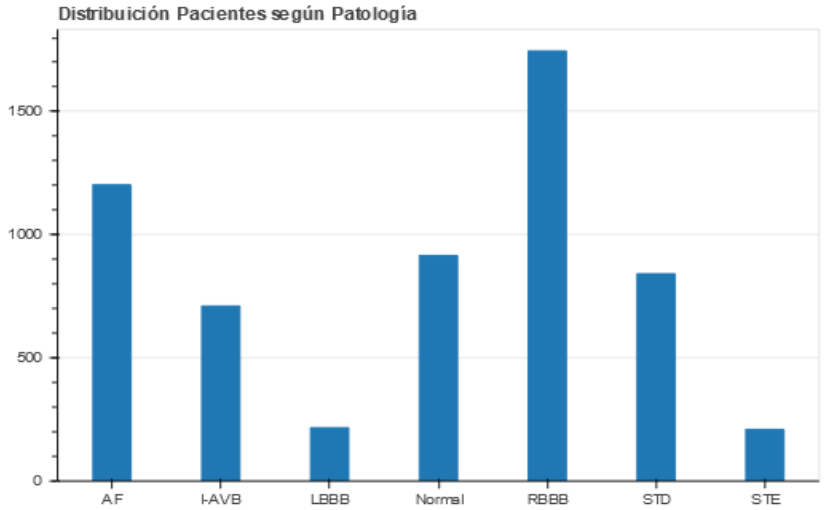

Figura 3.34. Conjunto de datos resultante tras la eliminación de las señales ECG con las patologías cardiacas PAC y VEB.

No obstante, tal y como se ha explicado, las patologías PAC y VEB pertenecen a un conjunto de patologías compuestas por latidos ectópicos, resultando en una contradicción con la etapa de segmentación. Debido a esto, se suprimen las señales ECG de estas patologías cardiacas, resultando en el conjunto de datos mostrado en la anterior figura. El nuevo dataset se somete al proceso llevado a cabo en el análisis anterior: se realiza una etapa de preprocesamiento, una etapa de extracción de características y se utiliza la misma red neuronal convolucional 1D multi-entrada diseñada.

3.2.3.1. RESULTADOS OBTENIDOS DEL TERCER ANÁLISIS

En el tercer análisis, los resultados arrojan un grado de cierto interés sobre el problema planteado. Se puede observar en las métricas de la tabla siguiente que la red convolucional se comporta muy bien frente a la compleja tarea de clasificación de los latidos cardiacos en función de su patología cardiaca con un rendimiento superior al 80%.

Tabla 5. Resultados del tercer análisis desarrollado: red neuronal convolucional 1D con una etapa previa de extracción de características sobre el conjunto de datos inicial sin PAC ni VEB. En la tabla se describe la precisión obtenida por el modelo para los datos de entrenamiento y para los datos de test.

EVALUACIÓN DEL MODELO	PORCENTAJE
Evaluación en datos de entrenamiento	83.90%
Evaluación en datos de test	82.33%

El modelo refleja que es capaz de generalizar las características extraídas de los datos para su correcta clasificación sobre los datos de test, reduciendo al máximo el posible sobreajuste. La matriz de confusión, representada en la siguiente figura, permite observar que el modelo es capaz de discriminar de forma notoria frente a las patologías analizadas en este tercer análisis.

El proceso de detección y clasificación automática de enfermedades cardiacas no es una tarea trivial y la matriz de confusión revela que el modelo tiene ciertas dificultades para la discriminación del segmento ST. Esto se debe a la etapa de preprocesamiento, específicamente, la etapa encargada de la reducción de la línea de interferencia base. La etapa de la reducción de la línea basal es una etapa clave del procesamiento de las señales. Tal y como se refleja en las conclusiones extraídas en [6], ninguno de los métodos analizados en la Tabla 1 es capaz de reconstruir satisfactoriamente una señal ECG original tras la eliminación de la línea basal sin modificar el segmento ST, por lo que, si el algoritmo fuera desplegado en un ambiente sanitario real, el personal especializado en cardiología debería mostrar especial atención en el segmento ST para complementar la clasificación de la señal.

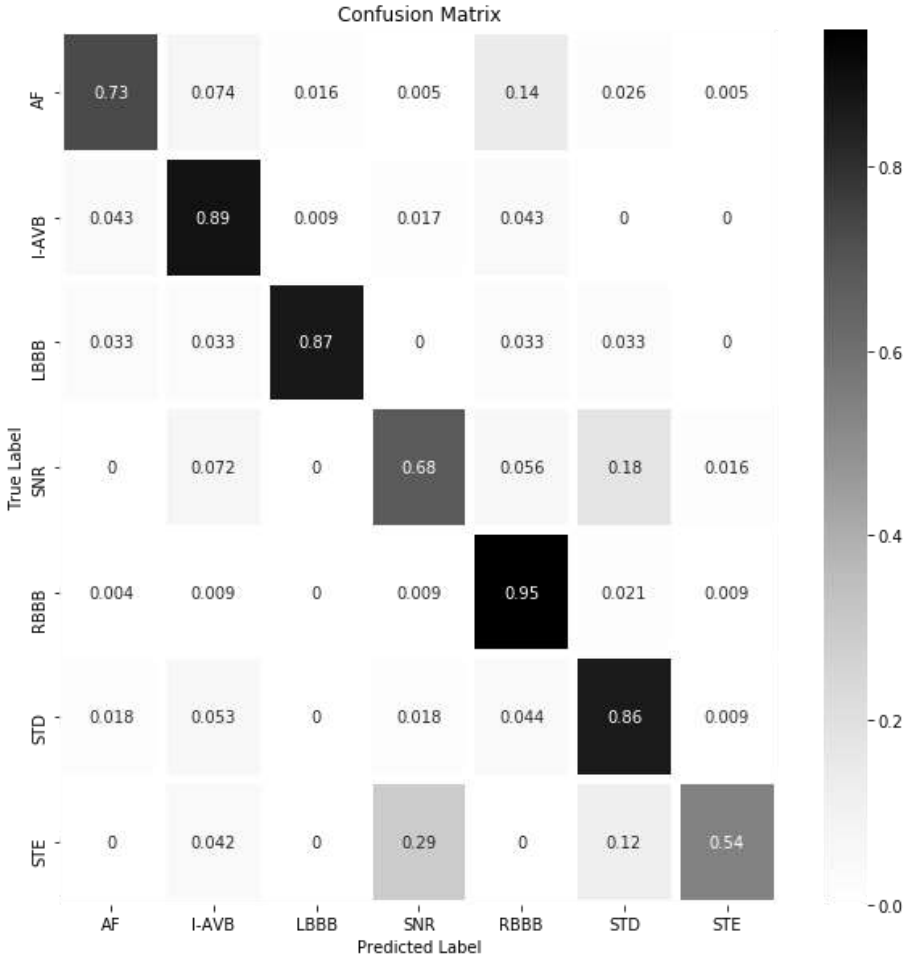

Figura 3.35. Matriz de confusión obtenida de los resultados de la red neuronal convolucional 1D multi-entrada sin las patologías cardiacas PAC y VEB.

Simultáneamente, se puede extraer una segunda conclusión de la clasificación realizada: el modelo tiene problemas en la diferenciación entre la patología de elevación del segmento ST y una señal ECG con ritmo sinusal normal, SNR. Esto se debe a que las patologías cardiacas que infieren sobre el segmento ST, STE y STD, no son patologías arrítmicas (como el resto de las patologías del dataset); por lo que, al verse modificado el segmento ST en la etapa de filtrado, la señal ECG resultante, si no contiene ninguna otra patología, estará compuesta de latidos regulares característicos de una señal ECG con un ritmo sinusal normal.

3.3. ESTUDIO DE LA INTERPRETABILIDAD

Una vez que se han obtenido los resultados de las técnicas de aprendizaje profundo se requiere arrojar luz sobre el desempeño de los algoritmos. La interpretabilidad es una cuestión importante cuando en el ámbito de aplicación de la técnica de aprendizaje profundo se requiere un alto grado de responsabilidad, tal y como se da en un entorno sanitario. Es complicado que una profesional médico/a especializado en cardiología crea en la resolución de un algoritmo para determinar la patología cardiaca de un paciente en concreto. Por lo tanto, una característica muy importante de este estudio es poder proporcionar una explicación de la clasificación realizada. Con este objetivo, en el apartado de la metodología, se proponen unas propiedades de explicación de los métodos y modelos que pueden ser utilizadas en este apartado de cara a formalizar el término de interpretabilidad.

- *PRECISIÓN DEL MODELO*

A pesar de todo, la precisión del modelo debe ser un parámetro importante a tener en cuenta en el desarrollo e implementación de un algoritmo de machine learning. Un modelo con un rendimiento bajo implica una mala clasificación de los datos por lo que su interpretabilidad, es decir, las características extraídas que derivan en la clasificación del modelo no son significativas; sin embargo, dado el caso contrario, un modelo con una alta precisión en los datos de test no implica que se las características extraídas sean significativas, desde el punto de vista del campo en el que se realiza, para su clasificación.

En este caso, la evaluación de la interpretabilidad se centra sobre la precisión del último modelo debido a sus resultados gratamente satisfactorios, precisión reflejada en la tabla 5, permite concluir que el modelo es capaz de generalizar en los datos de test con una precisión del 82.33%.

- *CONSISTENCIA*

En el estudio de la interpretabilidad se mencionan dos métodos distintos: *Class Activation Maps* y SHAP. La consistencia entre los modelos utilizados para obtener la interpretabilidad de la red neuronal dice mucho del desarrollo de la red, así como su tarea de clasificación sobre los datos analizados. Para ello, se analizan los resultados obtenidos en la utilización de los dos métodos mencionados sobre un segmento de latido dado. Con el fin de facilitar la comprensión de los resultados, se infiere sobre la visualización utilizando la escala cromática del blanco al rojo, véase Fig. 3.23., donde las características más importantes para la clasificación del modelo se encontrarán más cercanas al rojo mientras que las menos significativas se orientan al extremo contrario, el blanco.

Se puede observar del siguiente ejemplo que los modelos extraen información de regiones similares de las señales ECG, existiendo alguna diferencia.

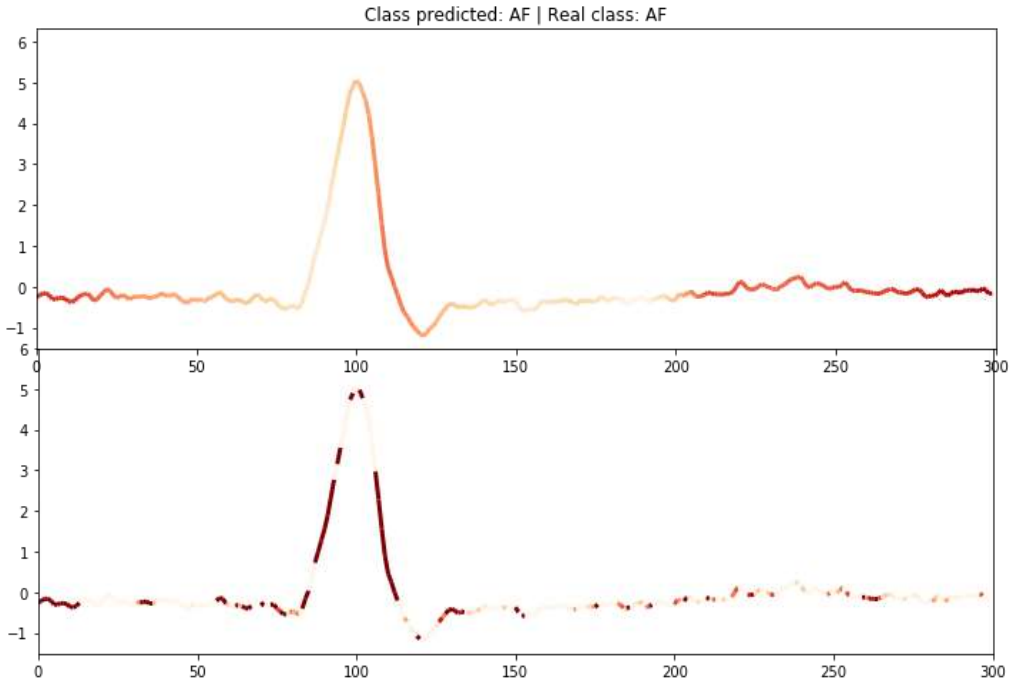

Figura 3.36. Interpretabilidad obtenida de ambos modelos, arriba la señal obtenida tras aplicar CAM; abajo, la señal obtenida tras aplicar SHAP

A pesar de existir alguna diferencia, se puede observar como las características obtenidas por ambos modelos son muy similares:

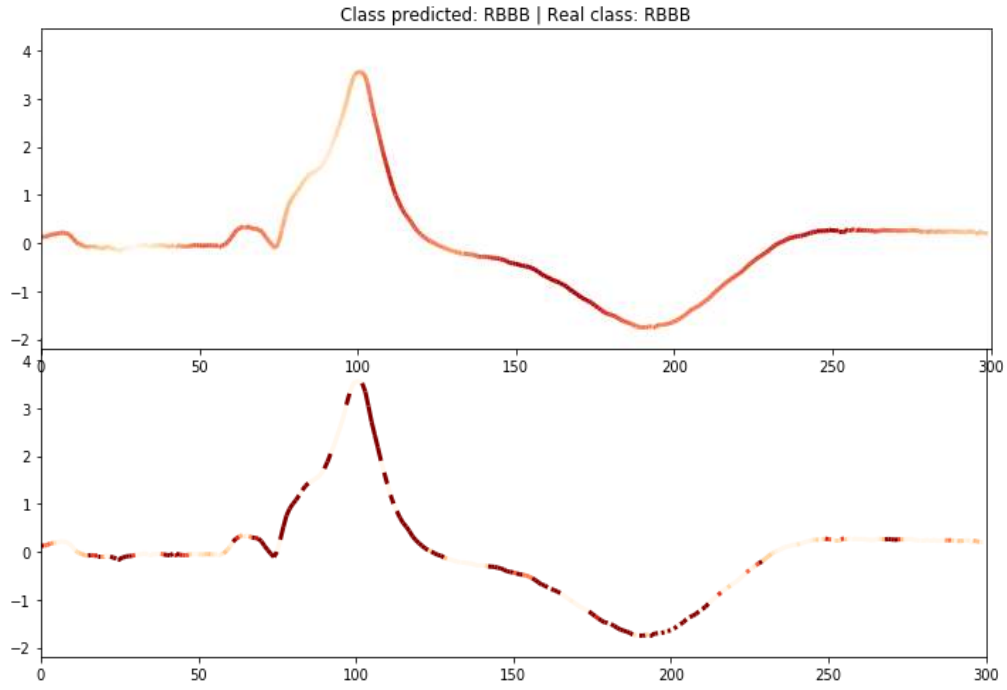

Figura 3.37. Segundo ejemplo de interpretabilidad obtenida en ambos modelos.

Se puede observar que en dos de los ejemplos expuestos las características extraídas de la señal son muy similares y significativas para su diagnóstico. Sin embargo, el método SHAP presenta varios

inconvenientes en referencia al otro método estudiado, CAM. En la siguiente tabla se muestra un método de evaluación de los modelos de interpretabilidad presentados, y se puede observar que la técnica SHAP tiene un coste computacional superior al método CAM, además de no poder ejecutarse en tiempo real. Por otro lado, la técnica CAM, gracias a su desarrollo teniendo en cuenta la capa de la red *Global Average Pooling*, permite su ejecución en tiempo real y con un coste computacional mínimo.

Tabla 6. Comparación de los métodos estudiados de interpretabilidad SHAP y CAM en base a características importantes en el ámbito sanitario: el tiempo de cómputo y su ejecución en tiempo real.

	Tiempo de Cómputo [s]	**Tiempo Real**
CAM	0.024	SI
SHAP	5.69	NO

A raíz de los resultados, se observa que para un ámbito sanitario es preciso el uso de CAM como técnica de interpretabilidad y, por lo tanto, la evaluación continua con este método; no obstante, sin desacreditar los resultados obtenidos con SHAP.

- *ESTABILIDAD*

La estabilidad es la propiedad que compara las explicaciones dadas por un modelo fijo sobre fuentes de datos distintas. En los análisis desarrollados se ha tenido en cuenta este mismo caso, ya que entre el análisis 2 y el análisis 3 la diferencia radica en el cambio del conjunto de datos a analizar, teniendo en cuenta la aplicación de la misma red neuronal convolucional. Como resultado, las matrices de confusión y el rendimiento alcanzado por ambos análisis siguen una tendencia muy similar, sin tener en cuenta estas patologías; en el análisis 2 se podría llegar a decir que casi idénticas (véase figura 3.32 y 3.35).

La alta estabilidad ofrecida en el análisis indica que hay un grado mínimo de varianza, es decir, el modelo desarrollado es capaz de reconocer correctamente las características oportunas para la clasificación de las patología cardiacas.

Esta característica de la evaluación de la interpretabilidad puede ser mostrada en función de la siguiente serie de figuras. No es suficiente con mostrar en qué se está fijando la red y correlacionarlo con el punto de vista clínico, sino tratar de observar la red frente a datos nuevos con las mismas patologías para saber si es capaz de generalizar.

Para llevar a cabo la demostración, se escogieron varias señales ECG que compartían patologías y se observó cómo la red se enfrentaba a ellas y en qué partes de los latidos cardiacos se centraba para realizar la clasificación con las características extraídas. A continuación, se ponen varios ejemplos de alguna de la patología mencionada:

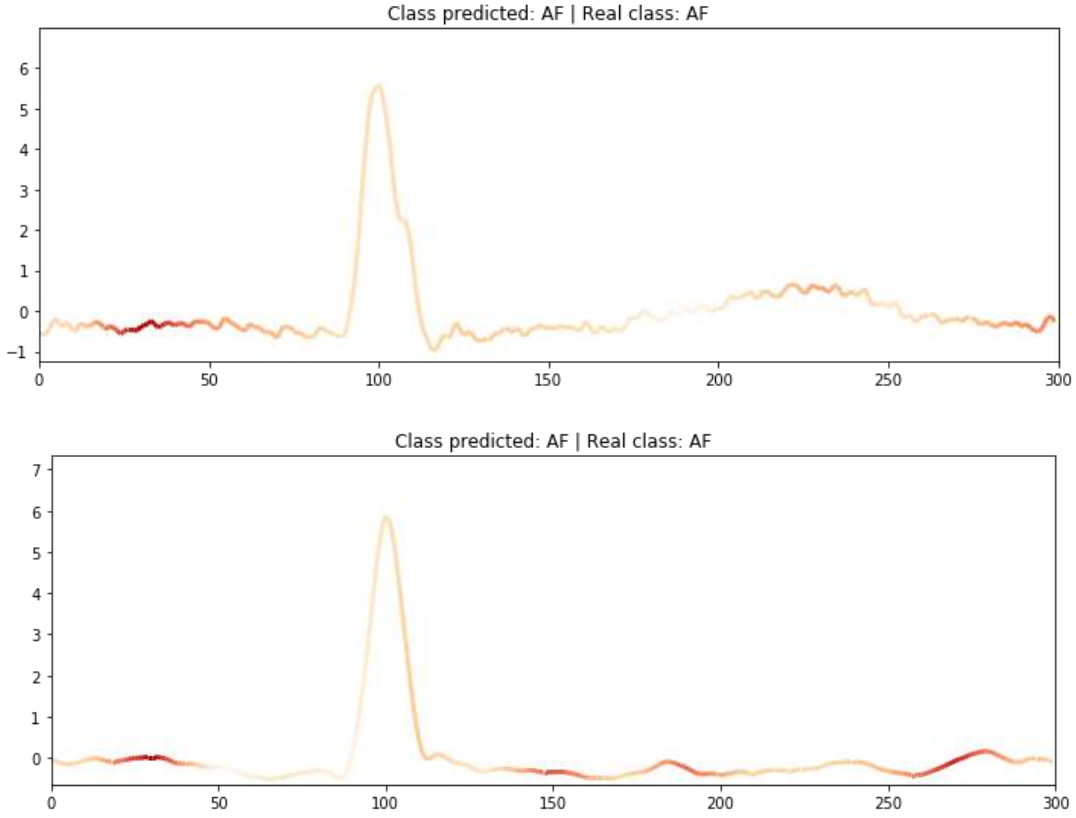

Figura 3.38. Dos latidos cardiacos de distintas señales ECG con Fibrilación Auricular. Se puede observar que la red se centra en propiedades similares de la señal.

Por otro lado, se escogió otra patología distinta que se refleja sobre la señal ECG de distintos pacientes de forma diferente. Sin embargo, lo más llamativo es el segmento RQ ancho, así como la "oreja de conejo" que se destaca en la subida al pico R (señalada en la señal ECG).

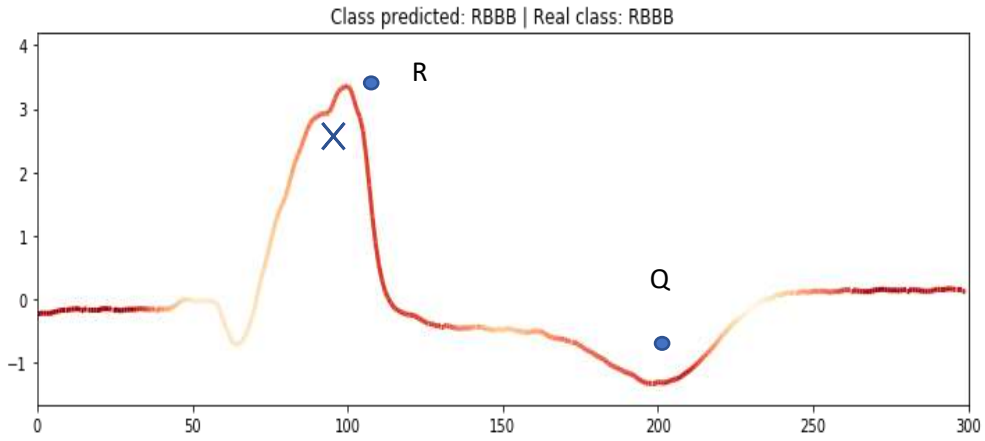

Figura 3.39. Latido cardiaco de una señal aleatoria (1) con RBBB.

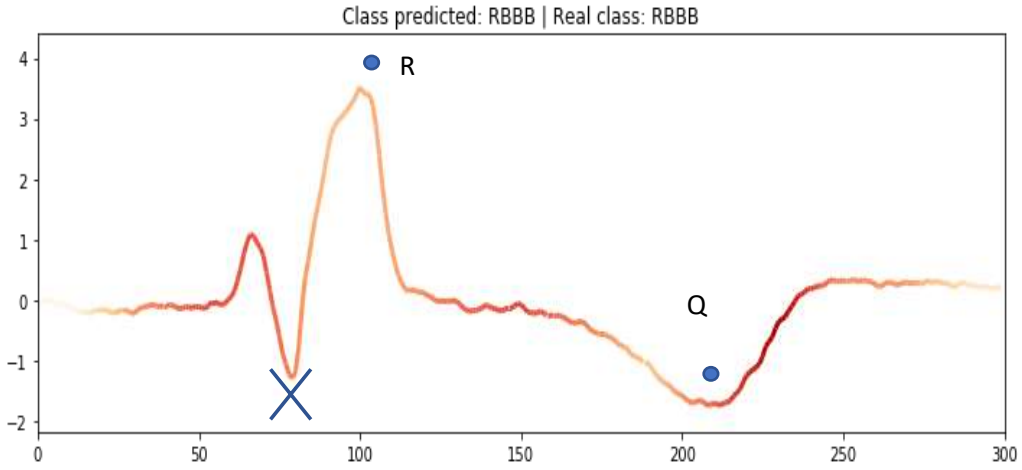

Figura 3.41. Latido cardiaco de una señal aleatoria (2) con RBBB.

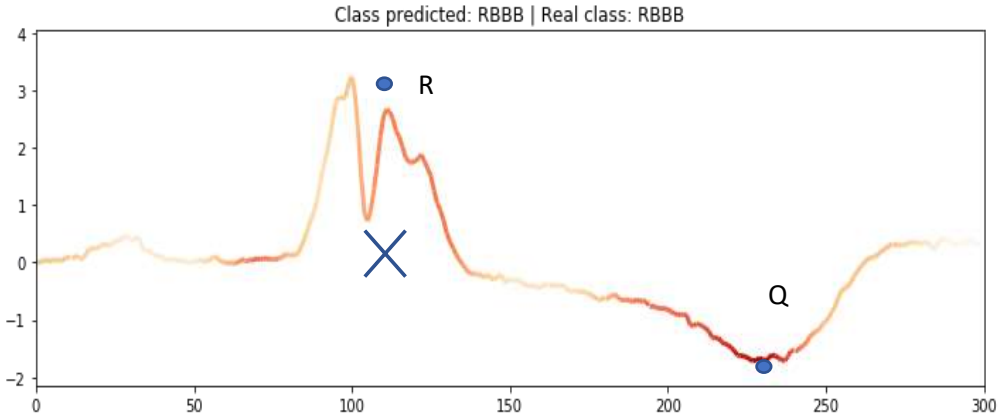

Figura 3.42. Latido cardiaco de una señal aleatoria (3) con RBBB.

- *COMPRENSIBILIDAD*

La siguiente característica que permite definir la interpretabilidad se refiere a medir qué tan bien entienden los humanos ajenos al desarrollo del algoritmo las explicaciones proporcionadas por el algoritmo de interpretabilidad empleado. En esta parte, se vuelve importante el algoritmo de interpretabilidad escogido: CAM.

CAM permite tener unos resultados en tiempo real ya que las operaciones internas que requiere se van ejecutando a medida que se desarrolla el entrenamiento del modelo. Para analizar este punto, se exponen muestras de la interpretabilidad que se obtiene al aplicarla en los datos analizados. Una explicación más detallada de los tipos de patologías cardiacas estudiadas se puede ver en el Anexo 1.

- *Fibrilación Auricular (AF)*

La Fibrilación auricular es una arritmia y se puede localizar observando el ritmo cardiaco, donde este será irregular y muy acelerado. Una señal ECG con fibrilación auricular carece de ondas P y en su defecto pueden encontrarse ondas F.

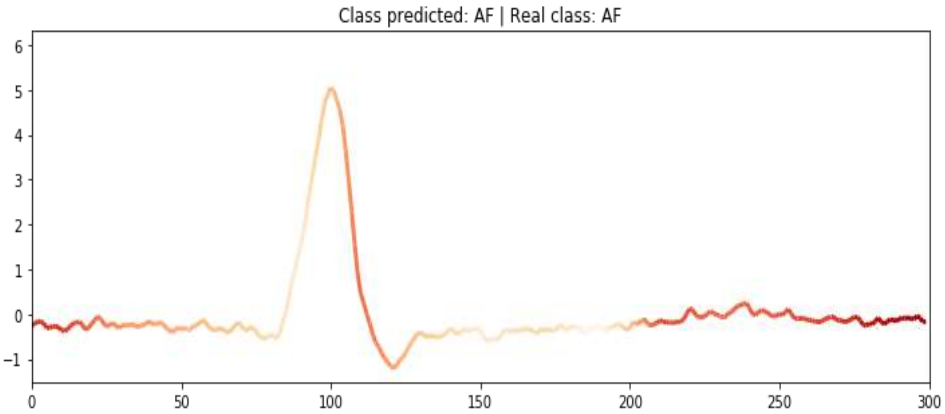

Figura 3.43. Interpretabilidad obtenida sobre una señal ECG con Fibrilación auricular.

Siguiendo la escala cromática establecida, se puede observar que la red se centra en el comienzo y en el final del segmento del latido principalmente, localización definida para la existencia de ondas P. En este caso, la red se está centrando en esta parte porque hay una ausencia de estas ondas, característica principal para la identificación de la fibrilación auricular. Por otra parte, otro punto de interés para la red son las muestras alrededor del pico R. Esto se debe a que otra característica de la fibrilación auricular es un ritmo irregular de los latidos cardiacos; ese ritmo se conoce por la separación de los intervalos RR de las señales cardiacas [45].

- *Bloqueo de primer grado del nodo AV (I-AVB)*

El boqueo de primer grado del nodo AV normalmente se puede observar por una separación de $125\ ms$ entre la onda P y el QRS. Si el intervalo PQ es superior a este tiempo, se denomina bloqueo de primer grado.

El aspecto más importante desde el punto de vista clínico es el segmento PQ, cuyo inicio se encuentra en la muestra 0 y su extremo final en la muestra 80; por lo tanto, la distancia entre muestras, denominada, N_{PQ}, corresponde con $N_{PQ} = 80 - 0 = 80$. Finalmente, el ancho total del segmento PQ es:

$$Ancho_{PQ} = \frac{80}{500} = 160\ ms$$

En definitiva, la clasificación de esta patología se debe a que el ancho del segmento PQ es superior al límite de $125\ ms$. En la siguiente figura, se puede observar como uno de los puntos que se centra la red para clasificar en este patología es a distancia entre la onda P, que se encuentra en la muestra 0 y el QRS en su totalidad.

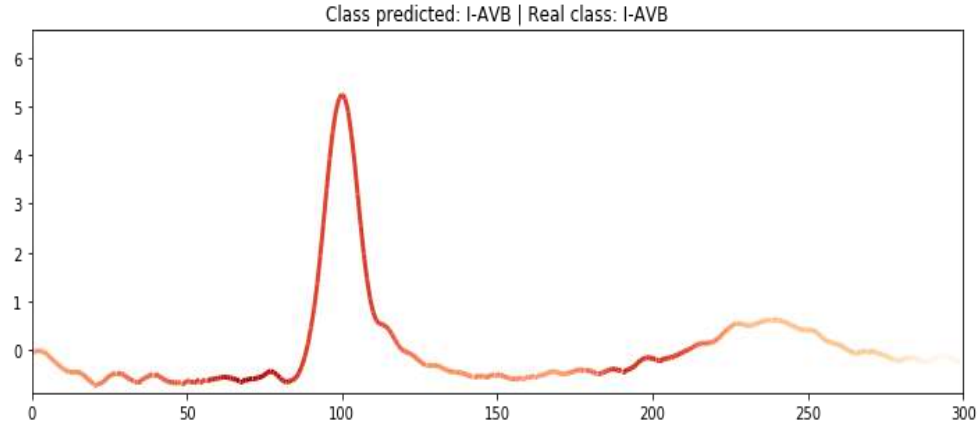

Figura 3.44. Interpretabilidad obtenida sobre una señal ECG con Bloqueo de primer grado del nodo AV.

- *Bloqueo de rama derecha (RBBB)*

El bloqueo de rama derecha (incompleto o no) se caracteriza por tener un bloqueo ventricular, definido en el electrocardiograma por mostrar un QRS más ancho de lo habitual. Si el QRS es superior a 120 ms podría considerarse que existe un bloqueo ventricular. A raíz de la siguiente figura se obtiene el número de muestras comprendidas en el QRS, $N_{QRS} = 200 - 100 = 100$, que, junto con la frecuencia de muestreo del conjunto de datos se conoce el ancho del QRS en ms:

$$Ancho_{QRS} = \frac{N}{f_s} = \frac{100}{500\ Hz} = 200\ ms$$

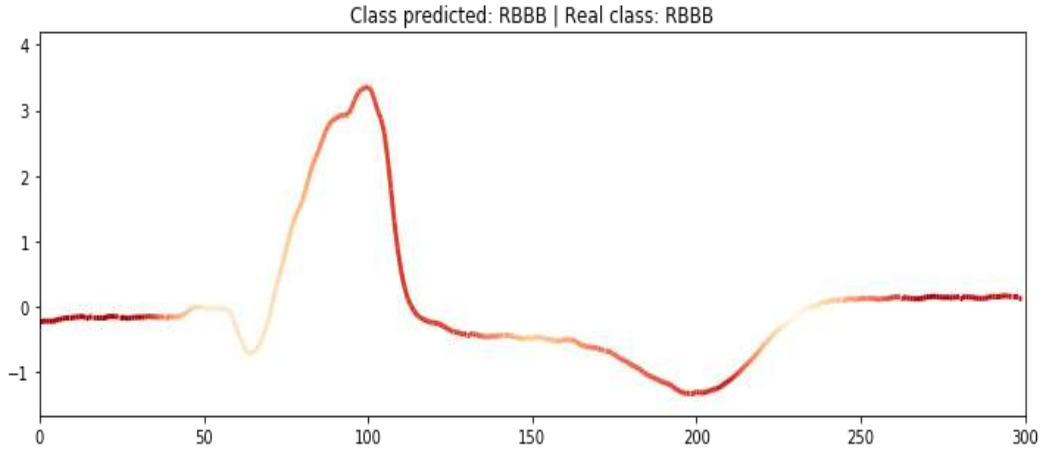

Figura 3.45. Interpretabilidad obtenida sobre una señal ECG con Bloqueo de rama derecha.

En la figura anterior se puede comprobar como la red neuronal convolucional se está centrando en el ancho del QRS, mayoritariamente, para su clasificación en esta patología. Otro aspecto particular de la morfología de las señales ECG con RBBB es la deformación que sufre el pico R (alrededor de la muestra 100) en la que parece que el modelo también presta atención.

- *Bloqueo de rama izquierda (LBBB)*

El bloqueo de rama izquierda se denota en las señales de electrocardiograma de la misma manera que el bloqueo de rama derecha. El QRS de estas patologías es más ancho de lo habitual, superior a $120\ ms$. Se realizan los cálculos anteriores para comprobar la duración del QRS, teniendo en cuenta que el QRS empieza en la muestra 100 y finaliza alrededor de la muestra 175; por lo tanto, el número de muestras del QRS es $N_{QRS} = 75$.

$$Ancho_{QRS} = \frac{N}{f_s} = \frac{75}{500\ Hz} = 150\ ms$$

Se puede observar que el QRS es de $150\ ms > 120\ ms$. Esto indica que existe un bloqueo ventricular.

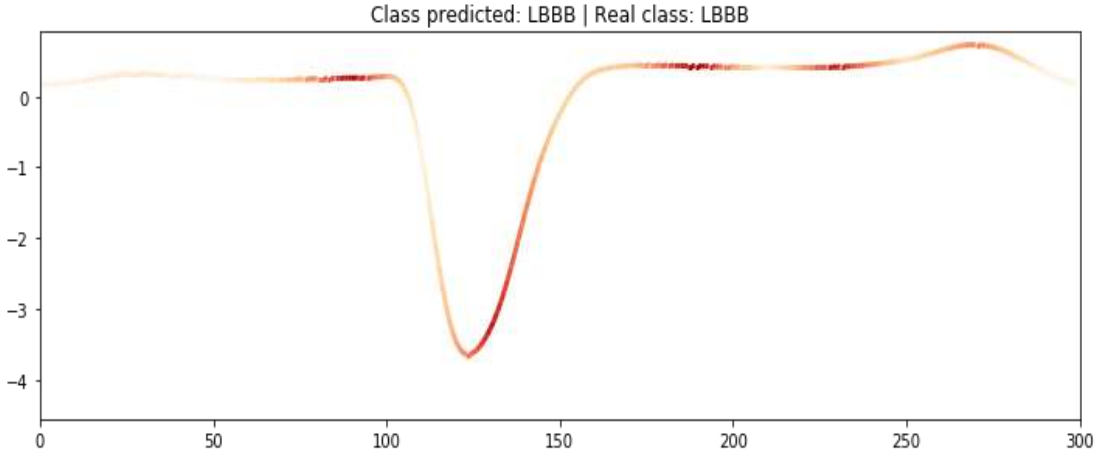

Figura 3.46. Interpretabilidad obtenida sobre una señal ECG con Bloqueo de rama izquierda.

Para reconocer que el bloqueo ventricular pertenece a la rama izquierda se debe observar la morfología de la figura. La red convolucional se está centrando en el comienzo del QRS, en su final, lo que indica que una característica que observa es el ancho del QRS. Además, se centra en el pico inferior de la figura. Este pico inferior define el pico R invertido característico de los bloqueos de rama izquierda.

- *Ritmo sinusal normal (SNR)*

El ritmo sinusal normal es una patología que se centra en mantener un ritmo regular entre los QRS de la señal ECG. La red neuronal se centra en todas las ondas pertenecientes al electrocardiograma y, principalmente, en el pico R, característico del ritmo.

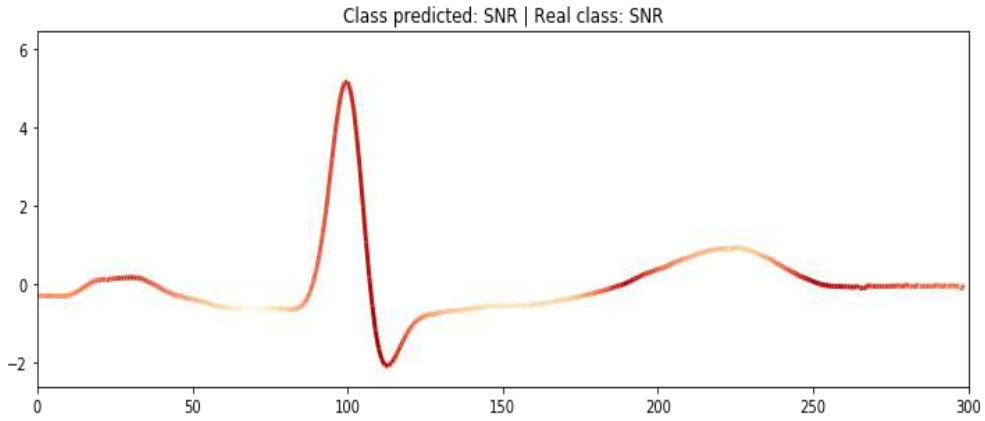

Figura 3.47. Interpretabilidad obtenida sobre una señal ECG con Ritmo sinusal normal.

- *Elevación del ST (STE)*

Los cambios en el ST son complejos de diagnosticar; sin embargo, la red neuronal parece tratar de extraer la información correcta desde el punto de vista clínico. Un cambio en el ST, en este caso, elevación del segmento ST se caracteriza por una subida constante hasta amplitudes negativas previas a la línea basal, y, a partir de este punto, una subida lenta hasta el pico T, siendo este pico más pronunciado de lo normal.

La red convolucional se centra en la subida del segmento ST, característica fundamental para diferenciar esta patología, así como en el pico T, aunque en menor medida.

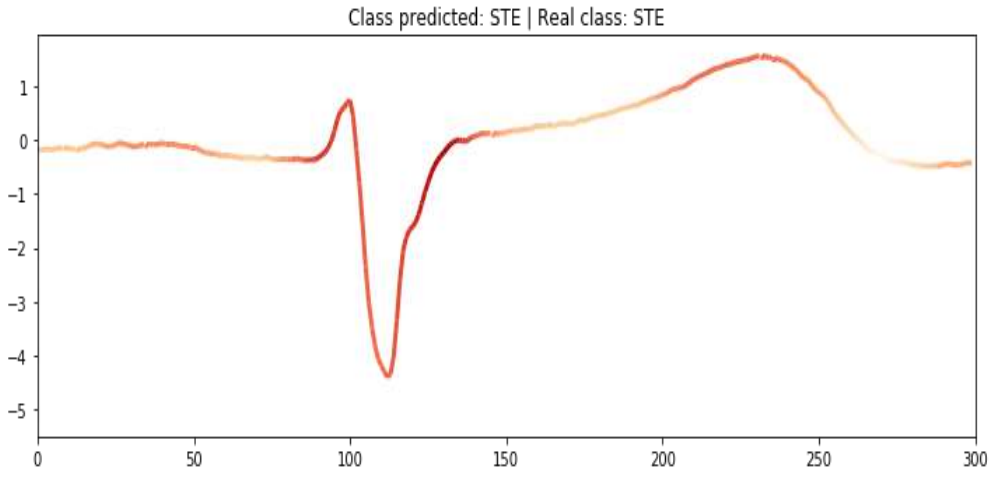

Figura 3.48. Interpretabilidad obtenida sobre una señal ECG con Elevación del ST.

- *Disminución del ST (STD)*

Por otro lado, la disminución del ST viene dada por otro tipo de subida característica del segmento ST o por contener una T negativa, como es el caso del siguiente latido. Se puede observar que la red se centra mayoritariamente en la bajada del R, poco significativo para predecir esta clase, pero, la T invertida capta su atención para proceder a su clasificación.

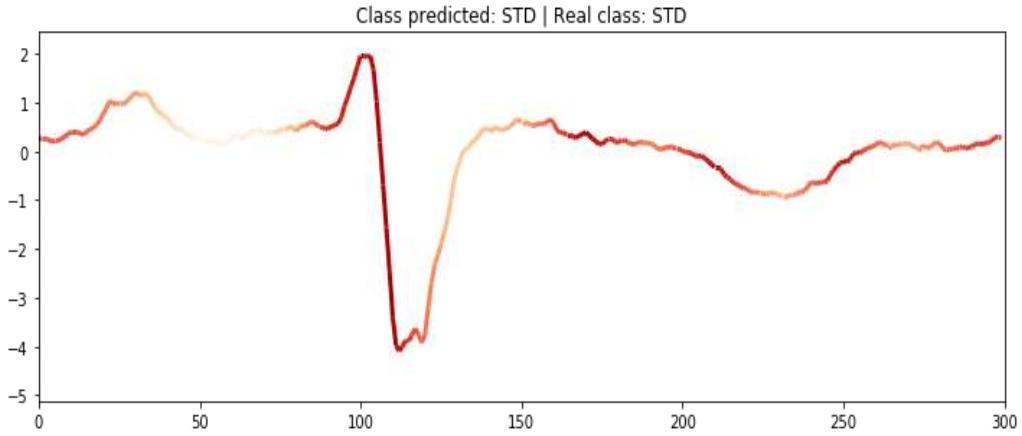

Figura 3.49. Interpretabilidad obtenida sobre una señal ECG con Disminución del ST.

4. CONCLUSIONES

En este capítulo final, se exponen las conclusiones obtenidas, las aportaciones del trabajo desarrollado y, por último, las posibles líneas de futuro de trabajo en base a la experiencia obtenida durante el trascurso del presente estudio.

4.1. DISCUSIÓN GENERAL

En el presente estudio se ha propuesto la investigación de técnicas de aprendizaje profundo para la detección, clasificación e interpretabilidad de patologías cardiacas dadas unas señales de electrocardiograma. Al comienzo del estudio se mencionó una de las principales características de la aplicación de técnicas de aprendizaje profundo: la extracción automática de características de los datos, al mismo tiempo que este proceso es llevado a cabo sin una explicación interpretable, siendo esta la causa de que estos algoritmos sean considerados como caja negra.

Esta investigación se ha centrado en realizar una clasificación de señales de electrocardiograma de pacientes con distintas patologías cardiacas, así como de mostrar la interpretabilidad del modelo en base a la extracción de características realizadas.

En el proceso llevado a cabo, se ha observado que la fase previa de procesamiento tradicional de las señales de electrocardiograma desempeña un papel crítico previamente al uso de algoritmos inteligentes, dada la importancia de la morfología de la señal para la detección de la patología cardiaca. Dentro de este ámbito se puede destacar el segmento ST de las señales (véase Fig. 1.1., para su identificación), al ser esta región una zona perjudicada de la aplicación de técnicas de filtrado para eliminar el ruido de las señales.

En el análisis posterior de las señales, se ha observado la importancia de una fase de extracción de características sobre las señales de electrocardiograma para extraer la información de un segmento PQRST, tal y como haría un personal sanitario especializado en cardiología, reduciendo la información necesaria a un latido cardiaco en cada una de las derivaciones. Este punto es importante y se observa tras realizar el primer análisis, ya que la red ofrecía buenos resultados; sin embargo, a la hora de aplicar los métodos de interpretabilidad se comprueba que las características extraídas por la red convolucional no tienen el valor esperado desde el punto de vista clínico para proceder a la correcta discriminación de las patologías cardiacas.

También es fundamental resaltar el conocimiento sobre el tipo de patologías estudiadas, con el objetivo de afrontar los problemas encontrados, tal y como se realizó en el segundo análisis con los fallos de la red al clasificar patologías de latidos ectópicos en un grupo de patologías que se observaban en todos los latidos cardiacos y, que la propia etapa de extracción de características suprimía. No todas las patologías cardiacas estudiadas y no estudiadas aquí se componen de las misma característica definida por su aparición en todos los latidos de todas las derivaciones. Hay patologías cardiacas que pertenecen a latidos ectópicos, como es el caso previamente mencionado de los complejos prematuros ventriculares y auriculares, otras que vienen representadas morfológicamente en ciertas derivaciones mientras que las restantes pueden representar un electrocardiograma normal o que pueden venir representadas por ondas positivas o negativas de las señales, etc.

Es por ello por lo que este campo resulta complejo de analizar, entender y abstraer las ideas para obtener una correcta clasificación, así como proporcionar una correcta explicación del modelo y de los resultados obtenidos en relación con el punto de vista clínico, cuestión que resultaría en un aumento de confianza por el personal sanitario y que abriría las puertas a su validación en un entorno real.

Por otra parte, los resultados obtenidos del tercer análisis arrojan una resolución satisfactoria de los objetivos del presente estudio, una clasificación con más de un 80% de acierto en cada latido cardiaco para una gran variedad de patologías, todas ellas pertenecientes al sistema de conducción cardiaco: Fibrilación auricular, bloqueos de rama izquierda y derecha, bloqueo de primer grado del nodo AV, ritmo sinusal normal y cambios en el ST. La evaluación de la interpretabilidad ha resultado satisfactoria en base a los métodos señalados en la metodología, así como la utilización de ambas técnicas CAM y SHAP, a pesar de haber descartado SHAP por su ineficiencia en necesidad de tiempo real. En un entorno de alta responsabilidad, como es el entorno sanitario, es preciso evitar el concepto de caja negra y esclarecer el modelo, pudiendo dar una explicación de la clasificación realizada, tal y como se ha realizado. Este factor propone que se pueda llevar los métodos a un entorno de validación y que el propio personal sanitario sea quien compruebe y fidelice la interpretabilidad obtenida.

4.2. APORTACIONES REALIZADAS

Esta investigación se ha centrado en la consecución de los objetivos planteados al comienzo de la misma, que consistía en poder detectar, clasificar e interpretar patologías cardiacas en base a unas señales de electrocardiograma iniciales:

1. Se propuso un estado del arte en técnicas de preprocesamiento de señales, enfocado a las señales de electrocardiograma. Se realizó un estudio de las características ofrecidas por distintas técnicas de reducción de la línea de interferencia base para su posterior elección en el desarrollo del trabajo. En cuanto a las técnicas de reducción de ruido se observó la eficacia de utilizar *Wavelet Transform* sobre este tipo de señales.

2. Se comprobó con éxito el efecto de utilizar una etapa de extracción de características previo al uso de algoritmos de aprendizaje profundo como medio de ayuda para su entrenamiento y posterior extracción de características. Debido a esto se muestra que es realmente importante el conocimiento aplicado al ámbito de estudio para el desarrollo de estos algoritmos.
3. Se propuso el desarrollo de dos técnicas de interpretabilidad y se señalaron sus características para conocer cuál de ellas es más acorde al entorno sanitario.

4. El uso de técnicas de interpretabilidad en este ámbito ha resultado uno de los aspectos más importantes, sugiriendo este tipo de técnicas para su mayor aceptación dentro de entornos con gran responsabilidad como es el sanitario.

4.3. PROPUESTAS PARA TRABAJO FUTURO

En vista de los resultados obtenidos al aplicar técnicas de procesamiento de señal y analítica de datos en un campo tan extenso, profundo, complejo y en muchas ocasiones, desconocido como es el ámbito sanitario, se puede concluir señalando que una gran parte del verdadero valor derivado del presente estudio es el aprendizaje y la propiedad intelectual generada.

El campo de la medicina no es un proceso trivial; para explicar esto solo hace falta mostrar tres señales de electrocardiogramas de pacientes distintos, correspondientes con la misma derivación, y lanzar una pregunta al aire: ¿Estas señales corresponden con la misma patología cardiaca?

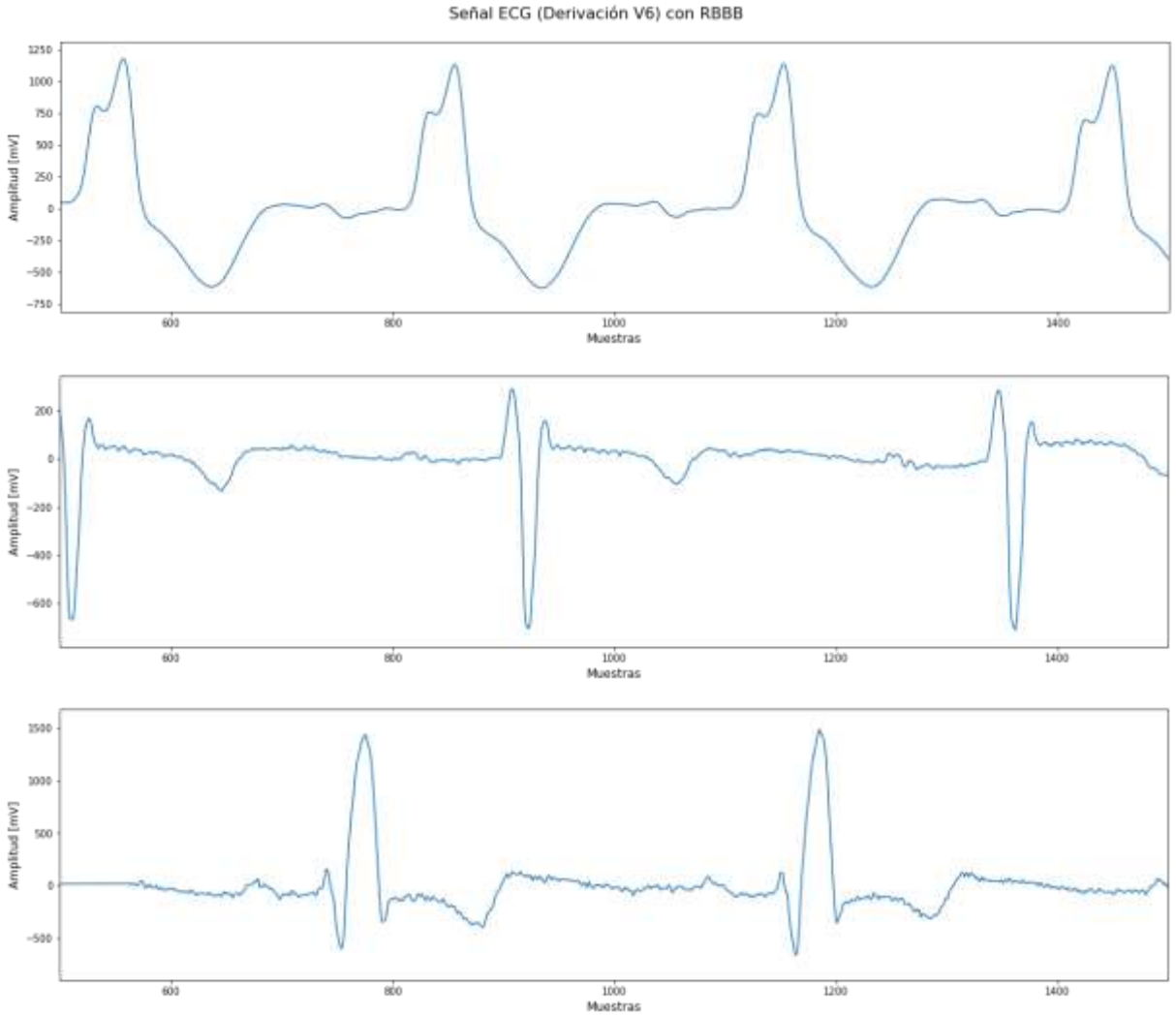

Figura 4.1. Señales de electrocardiograma correspondientes con la derivación V1 de tres pacientes distintos. La patología descrita es RBBB, patología que se observa principalmente sobre el bloqueo ventricular de esta derivación.

Sin poseer ningún concepto de medicina especializada en cardiología es complicado responder, pero si se basa la respuesta en la intuición, incluso yo, como autor, después de haber realizado un extenso estudio del campo, respondería a la pregunta anterior de forma negativa.

Las técnicas de aprendizaje profundo han demostrado su éxito en una gran cantidad de aplicaciones, resultando beneficiosas en tareas como la extracción de características automática del conjunto de datos. No obstante, a raíz de las conclusiones observadas donde se demuestra que uno de los conceptos más importantes es el conocimiento del campo estudiado, se propone utilizar un método de extracción de características tradicional a partir de la descomposición *wavelet* de las señales ECG. El estudio de las patologías, desde el punto de vista sanitario, se basa en estudiar la distancia entre las ondas o una mera observación del tipo de ondas existentes o, en su respectivo caso, de inexistencia de las mismas. Estos parámetros, a pesar de poder ser obtenidos por medio de técnicas de aprendizaje profundo como la expuesta, se pueden obtener por medios tradicionales de la mano de la implementación de *Continuous Wavelet Transform*, método idóneo para la extracción de características de las señales. En el siguiente artículo se expone este método y se plantea para realizar el estudio de este método [44].

Otra línea que se plantea es la utilización de técnicas de **Procesamiento Natural del Lenguaje** con **modelos de atención**, métodos que han resultado prometedores en series temporales y que pueden ser complementarios o estudiados para comprobar su eficacia y obtener una evaluación de la interpretabilidad todavía más robusta de la planteada al proponer una interpretabilidad intrínseca al propio entrenamiento del modelo.

4. ANEXOS

4.1. ANEXO 1: DESCRIPCIÓN DE LAS PATOLOGÍAS CARDIACAS.

Fibrilación Auricular (AF): La fibrilación auricular es una arritmia. Se puede localizar observando el ritmo cardiaco, siendo este muy acelerado (contracción auricular de unos 600 por minuto y una contracción ventricular de 200 por minuto) y será arrítmico, lo que significa que el espacio entre QRS será distinto en cada latido. Se puede ver en un latido ya que la fibrilación auricular carece de ondas P y en su defecto se encuentras ondas F.

La fibrilación auricular es un tipo de arritmia caracterizada por ser una de las más frecuencias, con una incidencia del 1.5-3% en la población general, aunque con mayor prevalencia en la población de edad avanzada y con patologías previas como hipertensión, obesidad y/o cardiopatía estructural. Se trata de una taquiarritmia supraventricular paroxística o permanente, es decir, la consecución de latidos anormalmente rápidos que ocurre por el establecimiento de conexiones eléctricas defectuosas del corazón que llegan a provocar latidos tempranos en las aurículas activándose de manera descoordinada sin contracción efectiva y, por lo tanto, reemplazando la actividad sinusal normal.

El diagnóstico de la fibrilación auricular se realiza a través de la observación de la señal ECG del paciente de 12 derivaciones, que muestra un patrón típico derivado en una irregularidad en los intervalos R-R de la señal, derivando en una frecuencia cardiaca que puede oscilar entre 110-350 lat./min. Además, las señales ECG con fibrilación auricular carecen de ondas P o no se encuentran bien definidas, y en su defecto, se encuentran ondas F, ondas que son respuestas rápidas e irregulares de diferente morfología.

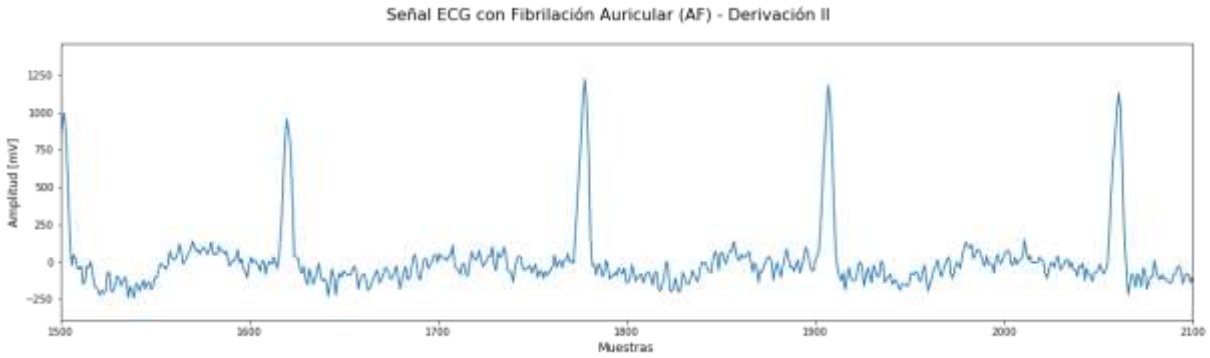

Figura 4.1. Señal ECG con Fibrilación Auricular (ausencia de ondas P)

La fibrilación auricular es una patología cardiaca que puede observarse en todos los latidos del paciente; se muestra en todas las derivaciones y, en definitiva, es analizable en cada parte de la señal.

Bloque auriculoventricular de primer grado (I-AVB): Los bloqueos auriculoventriculares son un conjunto de trastornos del sistema de conducción que provocan que el estímulo eléctrico generado en las aurículas sea conducido con retraso o no sea conducido a los ventrículos. Principalmente este tipo de bloqueos son producidos por una alteración en el nodo auriculoventricular o en el haz de His, aunque puede ser causado por fallos en otras estructuras cardiacas o por alteraciones metabólicas

El sistema de conducción cardiaco se conforma de las estructuras desde donde se producen y se transmiten los estímulos eléctricos que permiten la contracción del corazón. Sus principales elementos son el nodo sinusal, el nodo auriculoventricular (AV) y el sistema His-Purkinje, formando las fibras de Purkinje el haz de His. El sistema de conducción comienza en el nodo sinusal que es el encargado de iniciar o formar un impulso eléctrico cardiaco controlado a una frecuencia de 60-100 lat./min que deriva en la creación del estímulo rítmico de autoexcitación. El nodo auriculoventricular es el componente siguiente en la estructura del sistema de conducción cardiaco y su función principal radica en la transmisión de los estímulos de las aurículas a los ventrículos, ya que es la única conexión entre ambas estructuras. Además, este nodo realiza una gestión del ritmo cardiaco, retrasando o limitando la cantidad de estímulos que llegan a los ventrículos con el fin de evitar arritmias auriculares.

Por su parte, el sistema His-Purkinje se encarga de penetrar en el cuerpo fibroso central y provocar la despolarización de los ventrículos, transmitiendo la activación eléctrica que se originó en el nodo sinusal.

Este sistema de conducción se caracteriza por ser un sistema de autoexcitación, es decir, que en el caso de que el origen del impulso eléctrico, el nodo sinusal, fallara, los demás componentes del sistema asumirían esta función, aunque, con ciertas limitaciones. El nodo auriculoventricular asumiría la función de generar el impulso eléctrico con una frecuencia de 40-60 lat./min, más lenta que la generada por el nodo sinusal. En el caso de un fallo en el nodo AV en la generación de estos impulsos eléctricos, el sistema His-Purkinje asumiría esta función, generando un latido más débil con una frecuencia media de 30 lat./min.

Los bloqueos auriculoventriculares se clasifican en tres grados dependiendo de su severidad y del origen de las patologías. En el estudio, se analizan varias patologías debidas al sistema de conducción. Entre ellas se encuentra el bloqueo de primer grado, originado por un retraso del nodo sinusal en la generación del impulso eléctrico, derivando en una aparición tardía del complejo QRS de la señal ECG y, consecuentemente, produciendo una elongación del segmento P-Q. Esta alteración característica en la visualización del ECG se da por un intervalo superior a 200 ms y proporciona un QRS estrecho en el caso de no existir otra alteración cardiaca.

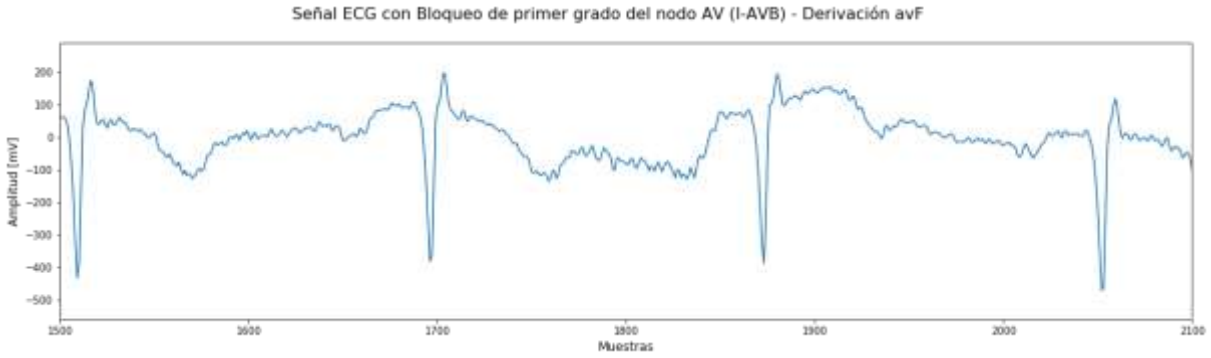

Figura 4.2. Bloqueo de primer grado del nodo AV (I-AVB) con un QRS estrecho en la derivación avF.

Los bloqueos de rama son trastornos propios de la conducción eléctrica provocados por el sistema de conducción eléctrico, específicamente distales al haz de His, generando cambios en la forma en que ambos ventrículos del corazón se despolarizan.

El haz de His, compuesto por las fibras de Purkinje, se divide en dos ramas: la rama derecha encargada de estimular el ventrículo derecho y al tercio derecho del septo interventricular, mientras que la rama izquierda estimula el ventrículo izquierdo y a los dos tercios izquierdos del septo interventricular. A su

vez, la rama izquierda se divide en dos ramas o, también considerados, fascículos: fascículo anterior que transmite el impulso eléctrico a la región antero-superior del ventrículo izquierdo y el fascículo posterior, que transmite el impulso a la región postero-inferior del ventrículo izquierdo. Una alteración en cualquiera de los componentes que conforman el haz de His será reflejado en la señal del electrocardiograma.

Los bloqueos de rama se producen cuando hay una obstrucción total del impulso de las ramas, derecha o izquierda previo a la división en fascículos.

Bloqueo de Rama Derecha (RBBB): El bloqueo de rama derecha se produce cuando la rama no es capaz de conducir el estímulo eléctrico, por lo que la despolarización de ambos ventrículos se realiza por la rama izquierda. Este retraso en la despolarización provoca un ensanchamiento del complejo QRS y cambios en su morfología, reflejo de las alteraciones de la conducción intraventricular.

Figura 4.3. Bloqueo de Rama derecha (RBBB) sobre la derivación V1.

Bloqueo de Rama Izquierda (LBBB): El bloqueo de rama izquierda se produce cuando el estímulo originado en el sistema de conducción no se transmite por la rama izquierda del haz del His. El bloqueo de la rama es completo si esta interrupción del impulso eléctrico se ocasiona previo a la subdivisión en los fascículos. Esta alteración impide que el estímulo eléctrico despolarice al ventrículo izquierdo de forma normal. Esto repercute en que aumenta el tiempo de despolarización e los ventrículos, como en el bloqueo de rama derecha, derivando en un ensanchamiento del complejo QRS y generando cambios en la morfología de la señal debido a las alteraciones propias del sistema de conducción intraventricular.

Figura 4.4. Bloqueo de rama izquierda (LBBB) sobre la derivación V1.

Ambos bloqueos muestran alteraciones sobre el complejo QRS, provocando que sea mayor de 120 ms. Estas patologías de la conducción se muestran definidas en todos los latidos a partir de su origen, pero su observación se centra en las derivaciones V1 y V6 a intervalos regulares.

El segmento ST, posterior al complejo QRS, figura 6.5, corresponde con la distancia entre el final del complejo y la onda siguiente, T, onda que muestra la recuperación eléctrica de las células denominado, repolarización. Este segmento en condiciones normales es plano o isoeléctrico, aunque puede presentar pequeñas variaciones. Para valorar su desplazamiento se utiliza como referencia este segmento y el segmento entre la onda T del latido previo y la onda P del latido analizado.

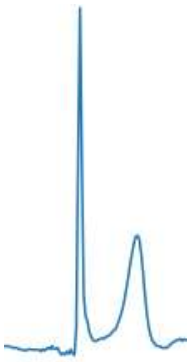

Figura 4.5. Segmento ST. El primer pico de la izquierda es el pico R; en su caída se encuentra el pico S (valle) y el segundo pico corresponde con la onda T.

En determinadas ocasiones se pueden observar variaciones del segmento ST sin que esto signifique alteración cardiológica. La elevación del ST (STE) o la disminución del ST (STD) en ligeros rangos de amplitud no tiene por qué derivar en una patología cardiaca, dificultando por tanto la deducción de alteraciones en el segmento ST.

Elevación del ST (STE): La elevación del segmento ST en alteraciones de este segmento debidas a cardiopatías isquémicas corresponde con uno de los signos tempranos más frecuentes del infarto agudo de miocardio y generalmente se puede relacionar con una oclusión aguda y completa de una arteria coronaria. El ascenso del segmento ST debe ser un ascenso persistente y en al menos dos derivaciones contiguas. Su visualización en las señales ECG se encuentra presente en todas aquellas derivaciones donde haya ocurrido la isquemia.

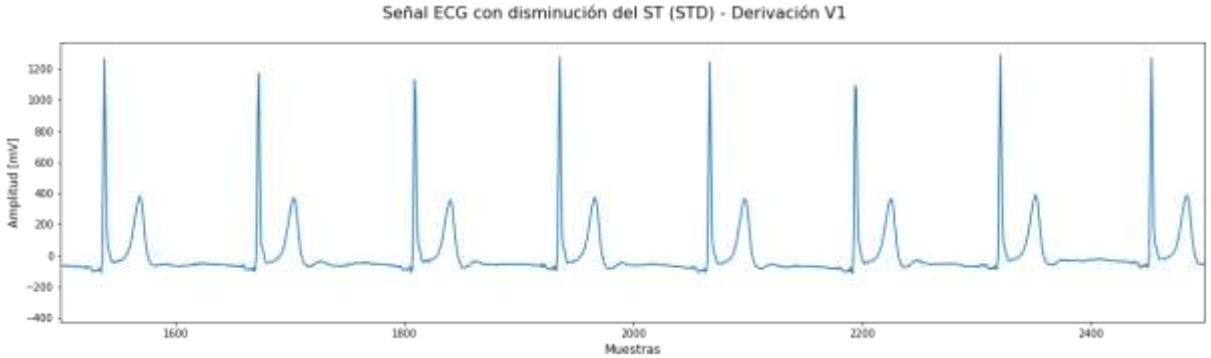

Figura 4.6. Elevación del ST (STE) en derivación V1.

Descenso del ST (STD): La depresión o descenso del segmento ST de forma aguda es un signo de isquemia al igual que la elevación del segmento ST. Esta patología cardiaca se suele correlacionar con una oclusión incompleta de la arteria coronaria. Su visualización debe de ser igual que el ascenso ST, presentándose en al menos dos derivaciones contiguas. Puede ser transitorio o persistente.

Figura 4.7. Disminución del ST (STD) en derivación V1.

Las patologías descritas, junto con el ritmo sinusal normal (SNR), congregan las patologías cardiacas que se encuentran en los datos obtenidos y que se utilizarán en el diseño y aplicación de técnicas inteligentes para tratar de predecir la patología que tiene un paciente en función de su señal ECG.

Figura 4.8. Señal ECG con ritmo sinusal normal (Intervalos RR con picos regulares)

5. BIBLIOGRAFÍA

[1] Paul Kligfield, MD, FAHA, FACC, Leonard S. Gettes, MD, FAHA, FACC, James J. Bailey, MD, Rory Childers, MD, Barbara J. Deal, MD, FACC, E. William Hancock, MD, FACC, Gerard van Herpen, MD, PhD, Jan A. Kors, PhD, Peter Macfarlane, DSc, David M. Mirvis, MD, FAHA, Olle Pahlm, MD, PhD, Pentti Rautaharju, MD, PhD and Galen S. Wagner, MD, "Recommendations for the Standardization and Interpretation of the Electrocardiogram", Journal of the American College of Cardiology Volume 49, Issue 10, March 2007 [Online].

[2] Tsai-Min Chen, Chih-Han Huang, Edward S.C. Shih, Yu-Feng Hu, Ming-Jing Hwang, "Detection and Classification of Cardiac Arrhythmmias by a Challenge-Best Deep learning Neural Network Model", iScience, March 27, 2020.

[3] Plataforma de datos estadísticos: Búsqueda de causas de muerte [Online]

[4] Harold P. AdamsJr, Gregory del Zoppo, Mark J. Alberts, Deepak L. Bhatt, Lawrence Brass, Anthony Furlan, Robert L. Grubb, Randall T. Higashida, Edward C. Jauch, Chelsea Kidwell, Patrick D. Lyden, Lewis B. Morgenstern, Adnan I. Qureshi, Robert H. Rosenwasser, Phillip A. Scott, and Eelco F.M. Wijdicks, "Guidelines for the Early Management of Adults With Ischemic Stroke", A Guideline From the American Heart Association, April 12, 2007 [Online].

[5] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao and Jimeng Sun, "Opportunities and Challenges of Deep learning Methods for Electrocardiogram Data: A Systematic Review".

[6] Akash Kumar Bhoi (Sikkim Manipal Institute of Technology (SMIT), India), Karma Sonam Sherpa (Sikkim Manipal Institute of Technology (SMIT), India) and Bidita Khandelwal (Central Referral Hospital and SMIMS, India), "Baseline Drift Removal of ECG Signal: Comparative Analysis of Filtering Techniques".

[7] Mr. Hrishikesh Limaye, Mrs. V.V. Deshmukh, "ECG Noise Sources and Various Noise Removal Techniques: A Survey", International Journal of Application or Innovation in Engineering & Management (IJAIEM), February 2016 [Online]

[8] Daqrouq, K. 2005, "ECG Baseline Wandering Reduction Using Discrete Wavelet Transform", Asian Journal of Information Technology4.issue 11, pp 989-995.

[9] Zhang, D. 2005, "Wavelet Approach for ECG BaselineWander Correction and Noise Reduction",Proceedings ofthe IEEE on Engineering in Medicine and Biology 27thAnnual Conference, pp 1212-1215.

[10] JL Rodriguez-Sotelo, D Cuesta-Frau, G Castellanos-Dominguez, "An Improved Method for Unsupervised Analysis of ECG Beats Based on WT Features and J-Means Clustering", Universidad Nacional de Colombia, Colombia, Universidad Politécnica de Valencia, España.

[11] Hongen Liao, Simone Balocco, Guijin Wang, Feng Zhang, Yongpan Liu, Zijian Ding, Luc Duong, Renzo Phellan, Guillaume Zahnd, Katharina Breininger, Shadi Albarqouni, Stefano Moriconi, Su-Lin Lee, Stefanie Demirci, "Machine learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting", Springer, 2019 [Online].

[12] Arthur Le Guennec, Simon Malinowski, Romain Tavenard, "Data Augmentation for Time Series Classification using Convolutional Neural Networks", HAL.

[13] Jianwei Zheng, Jianming Zhang, Sidy Danioko, Hai Yao, Hangyuan Guo & Cyril Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients [Online].

[14] Rahul Kher (2019) Signal Processing Techniques for Removing Noise from ECG Signals. J Biomed Eng 1: 1-9. [Online].

[15] Manpreet kaur Aneja, Birmohan Singh, January 2011, "Comparison of different approaches for removal of Baseline wander from ECG signal" [Online].

[16] Steven W. Smith, "The Scientist and Engineer's Guide to Digital Signal Processing", Second Edition, California Technical Publishing,1999 [Online].

[17] Mahesh Chavan, Mahadev Dattatraya Uplane, "Suppression of Baseline Wander And Power Line Interference in ECG Using Digital IIR Filter", January 2008 [Online].

[18] Davis Montenegro, Javier Gonzalez-Barajas, Edwin Francisco Forero-Garcia, "Procesamiento Digital de Perturbaciones de Calidad de Potencia Eléctrica", December 2012 [Online].

[19] Yansong Wang, Weiwei Wu, Qiang Zhu and Gongqui Shen, “Discrete Wavelet Transform for Nonstationary Signal Processing”, April 4, 2011 [Online].
[20] Maxime Yochum, Charlotte Renaud, Sabir Jacquir, “Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT”, 8 Jun, 2016 [Online].
[21] Petr Klapetek, David Necas, Christopher Anderson, “Gwyddion user guide”, 2004-2007, 2009-2019 [Online].
[22] Juuso Olkkonen, “Discrete Wavelet Transforms – Theory and Applications”, First published March, 2011.
[23] Dora M. Ballesteros, Andrés E. Gaona, Luis F. Pedraza, “Discrete Wavelet Transform in Compression and Filtering of Biomedical Signals”, Univeristy Military Nueva Granada, University Francisco José de Caldas, Colombia.
[24] Jose Rodrigo González, Ricardo López y Álvaro Jaramillo, “Wavelets in the analysis of EKG”, Facultad de Ciencias Básicas, Universidad Tecnológica de Pereira, Pereira, Risaralda, 2016 [Online].
[25] Wang Hua, Zhou Lijuan, Ma Cuiqin, “A brief Review of Machine learning and its Application”, Information Engineering Institude, Beijing, China, 2009.
[26] Tim Miller, “Explanation in Artificial Intelligence: Insights from the Social Sciences”, School of Computing and Information Systems, University of Melbourne, Melbourne, Australia on 15 Aug 2018.
[27] Jo, Jun-Mo. "Effectiveness of normalization pre-processing of big data to the machine learning performance." The Journal of the Korea institute of electronic communication sciences 14.3 (2019): 547-552.
[28] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." arXiv preprint arXiv:1702.08608 (2017). [Online]
[29] Christoph Molnar, “Interpretable Machine learning - A guide for Making Black Box Models Explainable”. [Online]
[30] Robnik-Šikonja, Marko, and Marko Bohanec. "Perturbation-based explanations of prediction models." Human and machine learning. Springer, Cham, 2018. 159-175. [Online]
[31] Zhou, Bolei, et al. "Learning deep features for discriminative localization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. [Online]
[32] Oquab, Maxime, et al. "Is object localization for free?-weakly-supervised learning with convolutional neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. [Online]
[33] Scott M. Lundberg, Su-In Lee, “A Unified Approach to Interpretating Model Predictions”, Paul G. Allen School of Computer Science, University of Washington, 2017 [Online].
[34] Christov, I.I. Real time electrocardiogram QRS detection using combined adaptive threshold. BioMed Eng OnLine 3, 28 (2004) [Online].
[35] Keiron O’Shea and Ryan Nash, “An Introduction to Convolutional Neural Networks”, [Online].
[36] Ana Gonzalez Muñiz, “Aplicaciones de técnicas de inteligencia artificial basadas en aprendizaje profundo (Deep learning) al análisis y mejora de la eficiencia de procesos industriales”, 2018. [Online].
[37] Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In Competition and cooperation in neural nets (pp. 267-285). Springer, Berlin, Heidelberg. [Online].
[38] Towards Data Science, “A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way”, [Online].
[39] Medium, “Understanding the Structure of a CNN”, [Online].
[40] Xie Chen, “Deep Convolutional Neural Network for Mapping Smallholder Agriculture Using High Spatial Resolution Satellite Image”, 2019, [Online].
[41] Repositorio Explicación CNN Layers [Online].
[42] Vaishali Ganganwar, “An overview of classification algorithms for imbalanced datasets”, 2012. [Online].
[43] Yu Sun,1 Lin Zhu,1 Guan Wang,1 and Fang Zhao1, “Multi-Input Convolutional Neural Network for Flower Grading”, [Online].
[44] Maxime Yochum, Charlotte Renaud, Sabir Jacquir. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. Biomedical Signal Processing and Control, Elsevier, 2016, [Online].
[45] González-Muñiz, A., Díaz, I., & Cuadrado, A. A. (2020). DCNN for condition monitoring and fault detection in rotating machines and its contribution to the understanding of machine nature. Heliyon, 6(2), e03395.

[46] Zachi I Attia, Peter A Noseworthy, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, Rickey E Carter, Xiaoxi Yao, Alejandro A Rabinstein, Brad J Erickson, Suraj Kapa, Paul A Friedman, An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction, The Lancet, Volume 394, Issue 10201, 2019, Pages 861-867, ISSN 0140-6736 [Online].