

Trabajo Fin de Grado

Diseño, optimización y validación de un delineador de electrocardiogramas basado en redes neuronales profundas U-NET

Design, optimization and validation of an Electrocardiogram delineator based on deep neural networks U-NET

Autor

Rubén Fernández Rodrigo

Directores

Juan Pablo Martínez Cortés
Antonio Miguel Artiaga

Titulación del autor

Ingeniería de Tecnologías y Servicios de Telecomunicación

Escuela de Ingeniería y Arquitectura
2024

Agradecimientos

A mis tutores Juan Pablo y Antonio Miguel por ofrecerme este proyecto y resolverme cualquiera de los problemas surgidos.

Al Instituto Universitario de Investigación en Ingeniería de Aragón (I3A) por ofrecerme la oportunidad de realizar prácticas en el marco de este trabajo.

A mi familia y a Hugo por apoyarme en todo momento.

Resumen

El análisis de electrocardiogramas (ECG) es una herramienta fundamental en la medicina para la detección y diagnóstico de diversas enfermedades cardíacas. Sin embargo, la interpretación manual de las formas de onda puede ser un proceso complejo y propenso a errores, especialmente en entornos con alta carga de trabajo o en casos de señales anómalas. En este contexto, la implementación de redes neuronales para la delineación automática de ECG representa un avance significativo, ofreciendo la posibilidad de mejorar la precisión y la eficiencia en la identificación de características clave del ECG, como ondas P, complejos QRS y ondas T. Este Trabajo de Fin de Grado (TFG) se centra en explorar la importancia de estas técnicas de delineación automática, evaluando su potencial para revolucionar la práctica clínica y optimizar el diagnóstico de afecciones cardíacas a través del uso de inteligencia artificial.

En el presente trabajo, se ha propuesto la delineación de electrocardiogramas mediante la red neuronal U-NET, reconocida por su destacado desempeño en la segmentación de imágenes, siendo esta la aplicación original. Se plantea una adaptación de esta red para la delineación de electrocardiogramas, una tarea que hasta ahora no ha sido ampliamente explorada en la literatura existente, pero mostrando resultados prometedores. El objetivo principal de este proyecto se centró en superar los algoritmos de Procesamiento Digital de Señales (DSP), los cuales presentan un rendimiento elevado en estas tareas, pero suelen estar limitados a un número reducido de morfologías. Se busca, por tanto, lograr un sistema con mayor capacidad de generalización.

Inicialmente, se llevó a cabo un proceso exhaustivo de búsqueda de bases de datos, encontrando dos corpus de acceso público, aunque de tamaño relativamente limitado. Posteriormente, se procedió a la lectura y extracción de dichas bases de datos, realizando un preprocesamiento para adecuarlas como entrada a la red neuronal. Dada la escasez de datos disponibles, se planteó la tarea de aumentar el conjunto de datos mediante un generador de electrocardiogramas sintéticos y un delineador basado en algoritmos de procesamiento de señal [14] [32], empleado como preentrenamiento de la red. Adicionalmente, este generador incorporó ruido realista presente en los casos clínicos, el cual se utilizó en las grabaciones para aumentar la variabilidad de los datos de entrenamiento.

Se ha desarrollado una red U-NET básica con la cual se realizó una prueba de robustez, con el fin de corroborar que los resultados futuros fueran consistentes y no producto del azar. A continuación, se procedió a la variación de los hiperparámetros, tales como el número de capas y niveles de la red, con el propósito de identificar el modelo que presentara un mejor desempeño en la tarea de delineación. Se concluye que la red U-NET con 5 niveles y 3 capas obtiene, en promedio, los mejores resultados.

Una vez obtenidos los resultados con el modelo óptimo de U-NET, se ha evaluado la inclusión de distintos bloques en el cuello de botella de la red. De todas las configuraciones analizadas, la red U-NET complementada con un bloque LSTM bidireccional ha mostrado un rendimiento superior, tanto en comparación con otros bloques añadidos como con la red U-NET preentrenada. Por consiguiente, se ha seleccionado esta configuración como la definitiva para su comparación con otros trabajos existentes.

El modelo de red mencionado anteriormente alcanza resultados que no solo son comparables con los reportados en estudios previos revisados en este trabajo, sino que en ciertas métricas específicas los supera. Esto es especialmente destacable considerando las limitaciones de tiempo y, por ende, el número limitado de pruebas realizadas en el contexto de un Trabajo de Fin de

Grado (TFG). Estas circunstancias subrayan la eficacia y el potencial del modelo propuesto, a pesar de las restricciones inherentes al marco de esta investigación.

Se concluye que el preentrenamiento de la red proporciona mejores resultados que el entrenamiento directo con los datos de las bases de datos. Asimismo, se ha demostrado que la red desarrollada en este trabajo posee una mayor capacidad para delinear correctamente una diversidad de morfologías, en comparación con los algoritmos DSP. Por último, se han introducido combinaciones de modelos que, hasta donde alcanza el conocimiento actual, no han sido mencionadas en la literatura, lo cual abre nuevas vías para futuras investigaciones.

Abstract

The analysis of electrocardiograms (ECG) is a fundamental tool in medicine for the detection and diagnosis of various heart diseases. However, the manual interpretation of waveforms can be a complex and error-prone process, especially in high-workload settings or in cases of anomalous signals. In this context, the implementation of neural networks for the automatic delineation of ECGs represents a significant advancement, offering the potential to improve accuracy and efficiency in identifying key ECG features such as P waves, QRS complexes, and T waves. This Final Degree Project (TFG) focuses on exploring the importance of these automatic delineation techniques, evaluating their potential to revolutionize clinical practice and optimize the diagnosis of heart conditions through the use of artificial intelligence.

In the present work, the delineation of electrocardiograms has been proposed using the U-NET neural network, recognized for its outstanding performance in image segmentation, which was its original application. An adaptation of this network is proposed for ECG delineation, a task that has not been widely explored in the existing literature but shows promising results. The main objective of this project was to outperform Digital Signal Processing (DSP) algorithms, which have high performance in these tasks but are usually limited to a small number of morphologies. Therefore, the goal is to achieve a system with greater capacity.

Initially, an exhaustive search for databases was carried out, finding two publicly accessible corpora, although of relatively limited size. Subsequently, the reading and extraction of these databases were performed, with preprocessing to make them suitable as input for the neural network. Given the scarcity of available data, the task of augmenting the dataset was proposed using a synthetic electrocardiogram generator and a delineator based on signal processing algorithms [14] [32], used for pretraining the network. Additionally, this generator incorporated realistic noise present in clinical cases, which was used in the recordings to increase the variability of the training data.

A basic U-NET network was developed with which a robustness test was performed to ensure that future results were consistent and not due to chance. Next, the hyperparameters, such as the number of layers and levels of the network, were varied to identify the model that presented the best performance in the delineation task. It was concluded that the U-NET network with 5 levels and 3 layers achieved, on average, the best results.

Once the results were obtained with the optimal U-NET model, the inclusion of different blocks in the bottleneck of the network was evaluated. Among all the configurations analyzed, the U-NET network complemented with a bidirectional LSTM block showed superior performance, both compared to other added blocks and the pretrained U-NET network. Consequently, this configuration was selected as the final one for comparison with other existing works.

The aforementioned network model achieves results that are not only comparable to those reported in previous studies reviewed in this work but, in certain specific metrics, even surpasses them. This is particularly noteworthy considering the time constraints and, therefore, the limited number of tests conducted in the context of a Final Degree Project (TFG). These circumstances highlight the efficacy and potential of the proposed model, despite the inherent limitations of this research framework.

It is concluded that the pretraining of the network provides better results than direct training with the database data. Furthermore, it has been demonstrated that the network developed in this work has a greater capacity to correctly delineate a diversity of morphologies, compared to DSP

algorithms. Finally, combinations of models were introduced that, as far as current knowledge extends, have not been mentioned in the literature, which opens new avenues for future research.

Índice

Resumen	2
Abstract	4
1. INTRODUCCIÓN	8
1.1. Contexto y Motivación.....	8
1.2. Organización de la memoria.....	8
1.3. Electrocardiograma (ECG)	9
1.3.1. Fundamentos del ECG	9
1.3.2. Ondas del ECG	11
1.4. Redes Neuronales	12
1.4.1. Conceptos Básicos.....	12
1.4.2. Estado del arte	14
1.4.3. Arquitectura UNET	15
1.4.4. Modificaciones U-Net.....	17
1.5. Objetivos.....	19
1.6. Recursos.....	19
1.7. Organización del trabajo.....	20
2. MATERIALES Y MÉTODOS.....	20
2.1 Bases de datos.....	20
2.1.1 Base de datos de QT (QTDB)	20
2.1.2 Base de datos de la Universidad Lobachevsky (LUDB).....	20
2.1.3 Base de datos de ECGs sintéticos	21
2.2 Extracción de datos.....	22
2.3 Preprocesado de datos.....	22
2.3.1 Normalización Z-score.....	22
2.3.2 Interpolación	23
2.3.3 Filtrado	23
2.3.4 Adaptación de señal QT	25
2.3.5 Generación de máscara null	26
2.3.6 Aumento de datos en la base de datos LUDB	26
2.3.7 Delineación de datos sintéticos.....	27
3. ENTRENAMIENTO DE LA RED.....	28
3.1 Estrategia de entrenamiento	28
3.1.1 Entrenamiento y preentrenamiento	28
3.1.2 Validación cruzada	29
3.1.3 Inicialización de pesos	29

3.2 Introducción de datos en la red	30
3.2.1 Dimensiones	30
3.2.2 Solapamiento	30
3.2.3 Máscara (función de coste)	31
3.3 Métricas	32
3.4 Pruebas	34
3.4.1 Robustez U-NET	34
3.4.2 Profundidad y niveles U-NET	36
3.4.3 Función de coste personalizada	40
4. EVALUACIÓN	40
4.1 Método de evaluación con ventanas	40
4.2 Resultados	41
4.2.1 Red U-NET	41
4.2.2 Adaptaciones y Mejoras de la Red U-NET	42
4.2.3 Comparación de resultados	44
5. CONCLUSIONES Y PROYECCIONES	45
5.1 Conclusiones	45
5.2 Líneas futuras	50
Bibliografía	51

1. INTRODUCCIÓN

1.1. Contexto y Motivación

En España, la principal causa de mortalidad son las enfermedades del sistema circulatorio, cuya incidencia ha mostrado una tendencia al alza, principalmente debido a la pandemia de COVID-19. En concreto, 119,853 personas fallecieron por causas cardiovasculares en España en 2020, lo que representa el 24,3% del total de fallecimientos [34]. Una de las pruebas cruciales para la detección de dichas enfermedades es el electrocardiograma, prueba rápida y no invasiva que muestra la actividad eléctrica del corazón.

Sin embargo, un electrocardiograma no es útil sin un especialista que sea capaz identificar las posibles morfologías existentes en esta prueba mediante la identificación o delineación de las ondas inherentes al ciclo cardíaco. Debido a la importancia de la delineación para el diagnóstico clínico, han aparecido numerosos trabajos cuyo principal objetivo es la creación de un delineador automático para agilizar y ayudar al especialista en la detección de anomalías.

Tradicionalmente, la detección y delineación de las ondas y segmentos del ECG han sido realizadas manualmente por médicos especialistas, lo cual puede ser laborioso en caso de largos periodos de monitorización de la actividad cardíaca del corazón (Holter) y está sujeto a errores humanos y diferencias de criterio.

Actualmente, existen algoritmos de procesamiento digital de señal (DSP) que permiten alcanzar precisiones y sensibilidades elevadas. Sin embargo, estos algoritmos se basan en reglas, las cuales no abarcan la plenitud de los casos provocando que morfologías no vistas en las bases de datos utilizadas requieran una laboriosa adaptación de dichos algoritmos por técnicos especialistas.

Debido a la gran variabilidad de formas que puede adquirir una ECG se ha empezado a emplear redes neuronales para la delineación. Estas redes, actualmente, están compuestas por un gran número de capas, denominadas profundas, lo que les permite manejar de manera efectiva la complejidad inherente a las señales ECG. Específicamente, redes neuronales convolucionales (CNN) [3] [2], redes de memoria a corto y largo plazo (LSTM) [1], y redes totalmente convolucionales (FCN) [33]. Sin embargo, gran parte de los trabajos muestran todavía unos resultados peores que los de los algoritmos DSP del estado del arte.

Con base en lo mencionado anteriormente, este trabajo busca proporcionar un sistema de delineación robusto que, en un futuro, pueda llegar a aplicarse como sistema de ayuda a los profesionales médicos para la detección precoz de enfermedades cardíacas, facilitando y agilizando así el diagnóstico.

Para ello se pretende crear, entrenar y optimizar un modelo de red neuronal que alcance resultados comparables a los obtenidos con algoritmos DSP pero abarcando un número mayor de morfologías, y que sea totalmente automático, sin que requiera la intervención de ningún especialista.

1.2. Organización de la memoria

La descripción de este trabajo se estructura en dos partes principales: la primera corresponde a la memoria del proyecto, en la cual se detalla el desarrollo completo del mismo, y la segunda comprende los anexos, cuya función es complementar la información presentada en la memoria.

La memoria se inicia con un resumen y una introducción que explican los fundamentos del electrocardiograma y de las redes neuronales, dado que estos son los temas relacionados con los objetivos del proyecto. Asimismo, se proporciona una descripción de la red básica utilizada en este trabajo.

En el capítulo II, titulado "Materiales y métodos", se detalla el procedimiento de lectura de los datos, así como el formato en el que estos se almacenan. Además, se describe el proceso de preprocesamiento de los datos para prepararlos adecuadamente para la fase de entrenamiento.

Continuando con el orden cronológico de las tareas y reflejando esta secuencia en la estructura de la memoria, en el capítulo III, "Entrenamiento de la red", se describen las pautas y consideraciones tenidas en cuenta durante la etapa de entrenamiento además de las métricas utilizadas para evaluar la red. En este capítulo también se comentan las diversas pruebas realizadas para identificar el modelo de red óptimo para la tarea específica, evaluando su robustez, eligiendo el número de capas y niveles de la red y función de coste.

En el capítulo IV se expone el método de evaluación aplicado y los resultados obtenidos tanto con la arquitectura U-Net básica propuesta como con sus modificaciones. Se comparan los resultados y se selecciona el modelo más adecuado.

Finalmente, en el capítulo de conclusiones, se analizan los resultados obtenidos y se comparan con otros trabajos relacionados con la delineación. Además, se proponen posibles líneas de trabajo futuras que no han sido abordadas en este proyecto.

1.3. Electrocardiograma (ECG)

1.3.1. Fundamentos del ECG

El corazón es el órgano encargado de proporcionar oxígeno y nutrientes a todas las células del cuerpo. Este se compone principalmente de cuatro cavidades, dos aurículas (izquierda, derecha) y dos ventrículos (izquierdo, derecho).

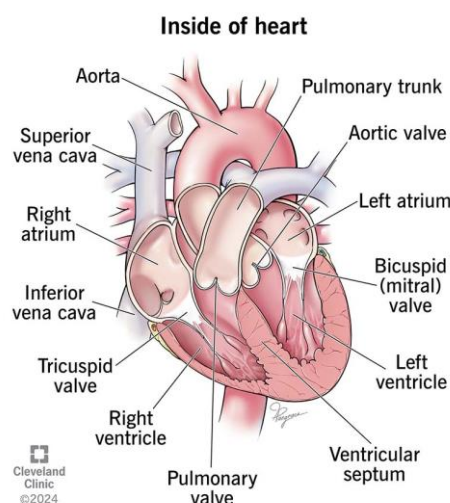


Figura 1: Representación del corazón y sus partes. Disponible en: [Heart: Anatomy & Function \(clevelandclinic.org\)](https://www.clevelandclinic.org/heart-anatomy-function)

El corazón está compuesto por células especializadas denominadas cardiomiocitos, que son responsables de la generación y transmisión de impulsos eléctricos a lo largo del tejido cardíaco, lo cual permite la contracción y relajación del músculo cardíaco. Aunque estos procesos son regulados por el sistema nervioso autónomo, la generación de impulsos eléctricos es llevada a cabo por cardiomiocitos especializados que actúan como marcapasos.

Los cardiomiocitos que funcionan como marcapasos, localizados en el nodo sinoauricular, son fundamentales para la regulación de la fuerza y la frecuencia cardíaca. Los impulsos eléctricos generados por estas células se reflejan en el electrocardiograma, proporcionando una herramienta no invasiva para evaluar el funcionamiento del corazón [37].

Para una correcta medición de los impulsos eléctricos del corazón (electrocardiograma) se requieren una serie de electrodos dispuestos en el torso y en las extremidades de una forma concreta. Concretamente, se sitúan 10 electrodos, 4 de ellos en cada una de las extremidades (RA, LA, RL y LL) y 6 electrodos situados en el precordio (parte central del pecho que queda por delante del corazón) [6], como se muestra en la figura 2.

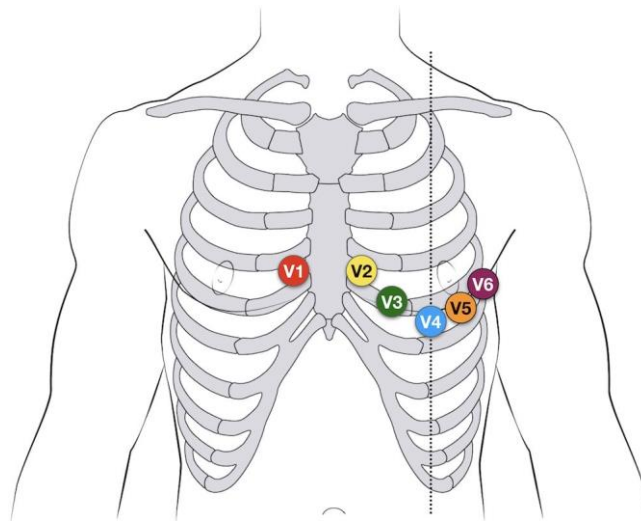


Figura 2: Disposición de los electrodos V1-V6 en el cuerpo humano. Disponible en: [ECG Lead positioning • LITFL • ECG Library Basics](#)

Con todo ello, se puede obtener una señal ECG de 12 derivaciones, que es el sistema de adquisición estándar. Las derivaciones son direcciones de propagación de los impulsos eléctricos a través del cuerpo humano. Cada derivación puede interpretarse como la proyección de la actividad eléctrica en una determinada dirección. Las derivaciones V1-V6 corresponden con la señal que miden dichos electrodos y es, por tanto, una vista en plano horizontal de la señal del corazón. El resto de derivaciones se obtiene de la combinación de las señales que miden los electrodos de las extremidades (I, II, III, aVR, aVL, aVF) que conforman la vista de la señal en el plano vertical [7].

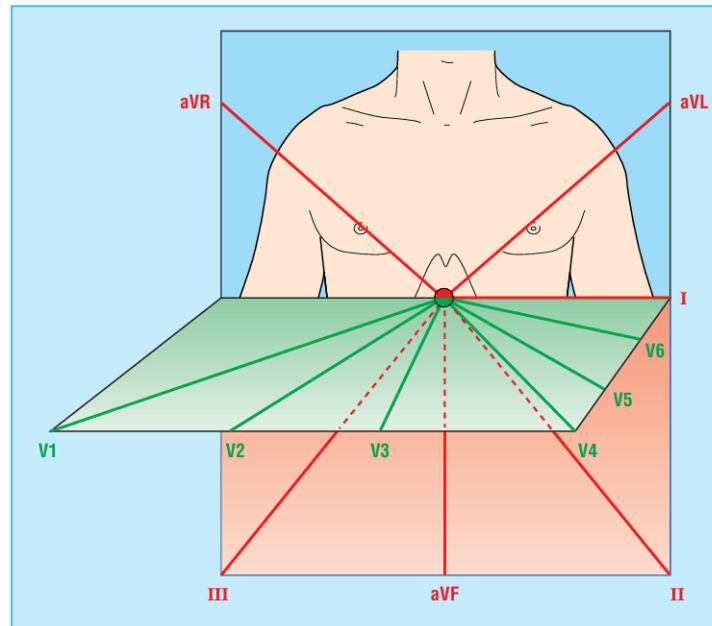


Figura 3: Vista 12 derivaciones de una ECG con respecto al cuerpo humano. Disponible en: [415.pdf \(nih.gov\)](#)

1.3.2. Ondas del ECG

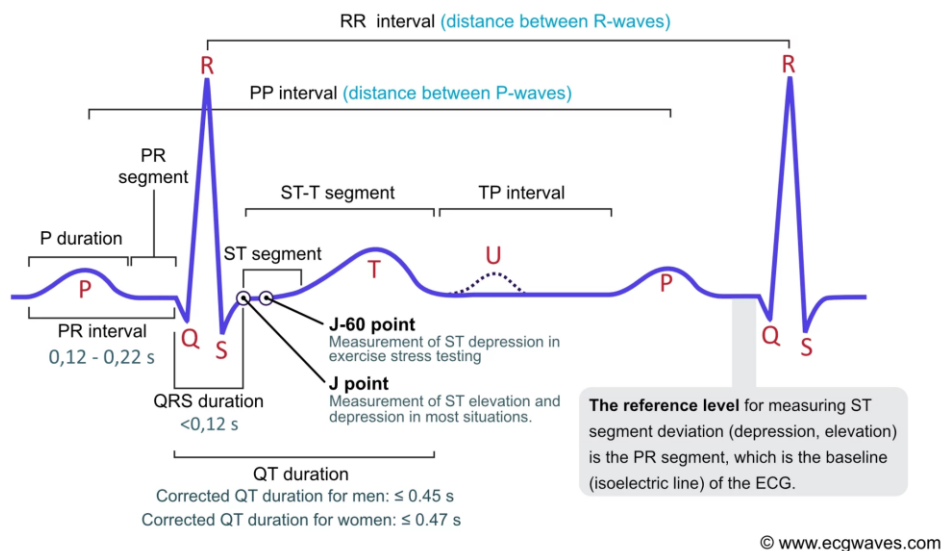


Figura 4: Componentes clave del ciclo cardíaco en el ECG. Disponible en: [ECG interpretation: Characteristics of the normal ECG \(P-wave, QRS complex, ST segment, T-wave\) – Cardiovascular Education \(ecgwaves.com\)](#)

A la hora de realizar un electrocardiograma de 12 derivaciones de un paciente con una actividad cardíaca “normal” obtenemos una señal como la de la figura 4. Se observan tres ondas principales, estas son la onda P, el complejo QRS y la onda T, que serán las delineadas en este trabajo.

La onda P representa la despolarización o activación eléctrica del músculo auricular por la cual es producido el movimiento contracción de las aurículas, que transportan la sangre a los ventrículos del corazón.

El complejo QRS registra la despolarización de los ventrículos, es decir, la contracción de los ventrículos, encargados de enviar la sangre al resto del cuerpo. Suele ser de mayor amplitud debido a que los ventrículos son de mayor tamaño y, por tanto, la contracción es mayor.

La onda T representa la repolarización del músculo ventricular, es decir, es el proceso por el que las células musculares del corazón (miocitos) recuperan su estado eléctrico normal para poder contraerse en el próximo latido. La repolarización de las aurículas sucede casi simultáneamente a la despolarización de los ventrículos, por lo que se ve enmascarado por el complejo QRS [8].

Una explicación más amplia de la relación entre la señal medida en un electrocardiograma y el funcionamiento del corazón se encuentra en el Anexo I.

1.4. Redes Neuronales

1.4.1. Conceptos Básicos

Las redes neuronales artificiales son sistemas complejos diseñados para ejecutar tareas específicas mediante la emulación del funcionamiento de las neuronas en el cerebro humano. Estas redes se implementan a través del desarrollo de sofisticados algoritmos computacionales de aprendizaje automático, los cuales no precisan de la programación explícita de reglas involucradas en la tarea. En lugar de ello, la red neuronal, basada en funciones no lineales, contiene gran variedad de grados de libertad, representados por coeficientes, que son optimizados a lo largo de un proceso iterativo conocido como entrenamiento. Este proceso de entrenamiento permite que la red aprenda y se adapte a partir de los datos, refinando su capacidad para realizar tareas con un alto grado de precisión y eficiencia. Las redes neuronales están compuestas por unidades básicas no lineales llamadas perceptrones. Mediante la agrupación de los perceptrones en formas concretas se consiguen redes capaces de aprender tareas concretas (una explicación detallada de los fundamentos de las redes neuronales se encuentra en el Anexo II).

La fase fundamental en una red neuronal es el entrenamiento. En esta fase se establecen las tareas o problemas que la red será capaz de resolver.

Como requisito fundamental, para tareas relativamente complejas, se requiere de bases de datos extensas, con ejemplos significativos, es decir, que exista una variedad suficientemente amplia de ejemplos para que la red sea capaz de generalizar el proceso para cualquier entrada.

Existen tres tipos de algoritmos para realizar el aprendizaje:

- Entrenamiento supervisado: se provee a la red un objetivo (target) por cada entrada, es decir, se indica a la red la solución que debería de proveer con la entrada que se está introduciendo. A continuación, se compara el objetivo con la salida obtenida de la red mediante una métrica de pérdida que dependerá del tipo de datos con el que se trabaje. La pérdida es una métrica que utiliza la red para actualizar los pesos (conexiones) que contiene cada perceptrón mediante algoritmos de propagación de forma que se reduzca esta pérdida y, por tanto, la salida sea más parecida al objetivo [9]. Adicionalmente, el método de optimización que se suele utilizar es el descenso por gradiente o alguna de sus variantes, esto limita a que la función de pérdida a optimizar y las funciones no lineales utilizadas en la red tienen que ser derivables.

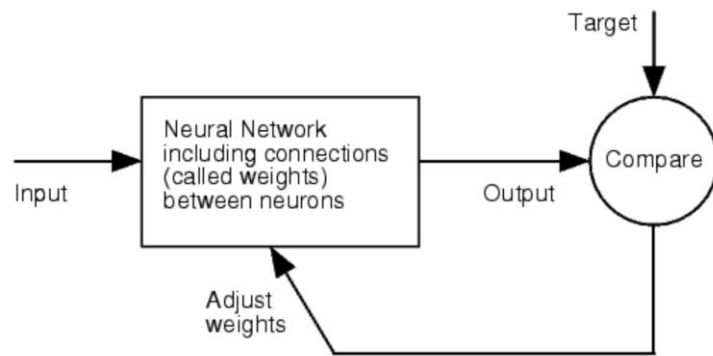


Figura 5: Esquema de entrenamiento supervisado de una red neuronal. Disponible en: [Neural Networks :: Preface \(Neural Network Toolbox\) \(izmiran.ru\)](#)

Este entrenamiento se utiliza habitualmente en tareas de regresión y clasificación.

- Entrenamiento no supervisado: la red no recibe retroalimentación del entorno. En cambio, la propia red es la encargada de identificar patrones o representaciones de los datos, los cuales no están etiquetados, agruparlos según similitudes y representar estos datos en un formato comprimido [9] [10].

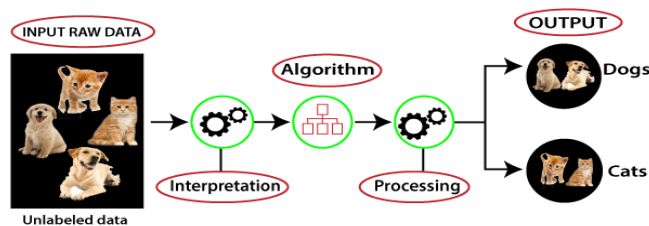


Figura 6: Esquema de entrenamiento no supervisado. Disponible en: [Unsupervised Machine learning - Javatpoint](#)

- Entrenamiento por refuerzo: es más parecido al aprendizaje supervisado. Consiste en un aprendizaje de prueba y error. Las acciones de software que trabajan para alcanzar su objetivo se refuerzan, mientras que las que se apartan del objetivo se ignoran. [9] [11].

En este trabajo se estudia el aprendizaje supervisado de forma que se requerirá de bases de datos con suficientes ejemplos para generalizar la tarea deseada. En el caso en que no se dispusiese de ejemplos suficientes, se puede realizar un aumento de datos. Consiste en introducir modificaciones a los ejemplos contenidos en las bases de datos disponibles y se introducen a la red como ejemplos nuevos. Esto habitualmente reduce el error de generalización del modelo [12].

Por otra parte, al escoger el modelo de red neuronal usado, es necesario establecer sus hiperparámetros. Los hiperparámetros son parámetros configurables de un modelo de aprendizaje automático que se establecen antes del proceso de entrenamiento y no varían durante este.

Se han de considerar diferentes parámetros como son el número de iteraciones del conjunto completo de datos de entrenamiento (epochs); el ratio de dropout, que hace referencia al porcentaje de neuronas que se desactivan en la red a entrenar; la tasa de aprendizaje, valor que referencia la magnitud del ajuste de los pesos en el proceso de entrenamiento; el tamaño del lote, refiriéndose al lote como un subconjunto del conjunto total de datos de entrenamiento; el número de niveles y capas, considerándose nivel como el bloque de operaciones básico cuya repetición conforma la red neuronal, entre otros. Todos ellos son parámetros variables, sin embargo, en cualquier modelo, es necesario escoger los parámetros a variar debido a la infinidad de posibilidades que existe al considerar todos ellos (una explicación más detallada de los hiperparámetros que conciernen a este trabajo se encuentra en el Anexo II).

1.4.2. Estado del arte

La delineación de electrocardiogramas (ECGs) es un área crítica en el análisis de señales biomédicas, que implica la identificación y demarcación de las características específicas del ciclo cardíaco, como las ondas P, QRS y T, así como los intervalos y segmentos correspondientes. La precisión en la delineación de estas características es fundamental para el diagnóstico de diversas condiciones cardíacas. Con el avance de las redes neuronales y el aprendizaje profundo, la delineación de ECGs ha mejorado notablemente en términos de precisión y eficiencia.

Las técnicas tradicionales para la delineación de ECGs incluyen métodos basados en la transformación de ondas, la detección de picos y enfoques basados en reglas heurísticas. Aunque estos métodos proporcionan resultados razonables, tienen limitaciones significativas, especialmente en presencia de ruido, variabilidad interindividual y diferentes condiciones de registro.

El surgimiento de las redes neuronales profundas ha transformado la delineación de ECGs al permitir la automatización y mejora de la precisión en la identificación de las características del ECG.

Las CNNs, redes con capas convolucionales las cuales procesan datos con una estructura de rejilla (como imágenes) extrayendo características, han sido ampliamente adoptadas para la delineación de ECGs debido a su capacidad para extraer características locales y su robustez frente al ruido.

Algunos trabajos han utilizado las CNNs básicas para la detección de picos y la clasificación de ondas en el ECG. La capacidad de las CNNs para aprender características jerárquicas ha permitido mejorar la identificación de las ondas P, QRS y T [2] [3].

También se han adaptado redes avanzadas como U-Net, originalmente diseñadas para segmentación de imágenes, para la delineación de señales de ECG, aprovechando su capacidad para combinar información de múltiples escalas [14][15].

Las RNNs, redes que utilizan conexiones recurrentes que permiten mantener una memoria de entradas anteriores, y en particular las LSTM y GRU, son adecuadas para la delineación de ECGs debido a su capacidad para capturar dependencias temporales en las señales.

LSTM es una variante de RNNs que ha sido utilizada para modelar la dinámica temporal de los ECGs, mejorando la precisión en la identificación de las ondas y los intervalos, al tener en cuenta el contexto temporal de la señal [1] [16].

Otros trabajos más recientes han adaptado los modelos basados en transformers, conocidos por su éxito en procesamiento de lenguaje natural, para tareas de series temporales [17], incluyendo

la delineación de ECGs. Concretamente, los transformers utilizan mecanismos de atención para capturar relaciones a largo plazo en las señales de ECG, permitiendo una delineación más precisa de características que dependen de un contexto temporal extendido.

1.4.3. Arquitectura UNET

La red U-Net es una arquitectura de red neuronal convolucional diseñada principalmente para tareas de segmentación de imágenes biomédicas, aunque su aplicación se ha extendido a otros campos que requieren segmentación precisa. Fue desarrollada por Olaf Ronneberger, Philipp Fischer y Thomas Brox en 2015 en el contexto de la segmentación de imágenes de microscopía de células [13].

Una de las ventajas más importantes es que la arquitectura U-net es ideal para preservar a su salida el tamaño inicial (es decir, el tamaño de la señal o imagen a su entrada).

1.4.3.1 Entrada y salida

La red principal utilizada en este estudio es una red U-net unidimensional (1D), diseñada específicamente para la tarea de delineación de señales de electrocardiograma (ECG). Esta arquitectura cuenta con una única entrada que corresponde al ECG que se va a analizar, la cual se introduce en la red en subconjuntos denominados ventanas (la duración de estas ventanas es un parámetro de diseño). La red produce una salida del mismo tamaño que el de la entrada con 4 posibles valores para cada muestra (ondas P, QRS, T y casos nulos), como muestra la figura 7, convirtiendo así el problema en una tarea de clasificación muestra a muestra. Cada muestra de entrada es evaluada individualmente por la red neuronal para asignarla a una de estas cuatro clases, utilizando el formato one-hot para la salida.

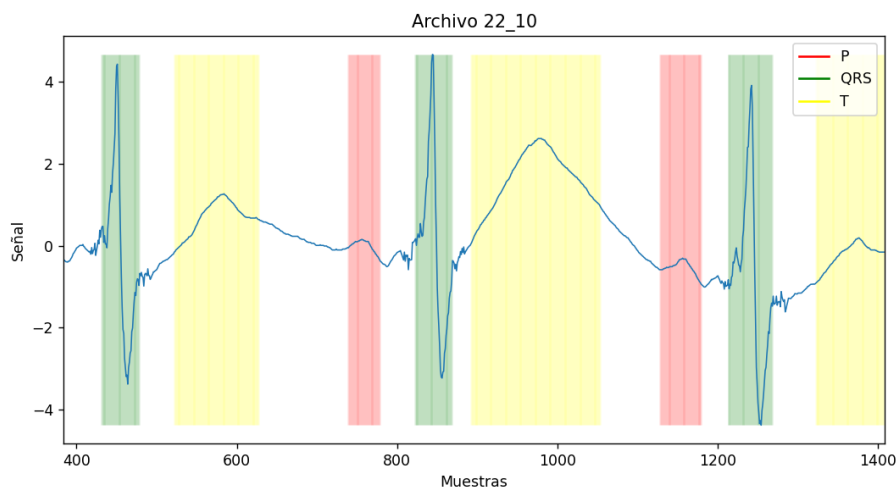


Figura 7: Salida de la red neuronal U-net, donde cada etiqueta se indica con una región coloreada diferente, junto a señal de entrada de la derivación 10 del registro 22 (Lobachevsky).

1.4.3.2 Etapas

La U-net presenta una arquitectura fundamental compuesta por dos trayectos principales representados en la figura 8. El primero, conocido como codificador o etapa de contracción,

reduce la dimensión de la entrada aumentando el número de características extraídas. De este modo se obtienen representaciones jerárquicas y abstractas, necesarias para la comprensión de las características de las señales de electrocardiograma (ECG). Estas representaciones permiten a la red capturar una variedad de aspectos de la señal a distintas escalas y niveles de detalle, como frecuencia, energía, patrones morfológicos variados y la variabilidad de los latidos cardíacos. Este proceso se facilita mediante la reducción de la dimensionalidad temporal y el aumento del número de filtros convolucionales o características.

Por otro lado, el decodificador o etapa de expansión consiste en una secuencia de operaciones de convolución traspuesta y concatenación. Esta etapa utiliza como entrada las características extraídas en la etapa de contracción, conocida como el cuello de botella. Su función es reconstruir una salida más precisa basada en la información previamente capturada.

Es importante mencionar que tanto el codificador como el decodificador están estructurados con un número equivalente de capas, lo cual configura a la U-net como una red prácticamente simétrica en términos de profundidad y complejidad arquitectónica.

Una limitación inherente a la estructura descrita anteriormente es la compresión lineal de la entrada, lo cual puede restringir la transmisión efectiva de todas las características de la señal. Para abordar este desafío, se han implementado conexiones directas entre capas del codificador y el decodificador, conocidas como "conexiones saltadas" (skipped connections). Estas conexiones facilitan la transferencia directa de información desde el codificador al decodificador, concatenando la entrada de cada nivel del decoder con la salida del nivel del encoder correspondiente, mitigando así el problema de la pérdida de detalles importantes durante el proceso de codificación y decodificación [13].

Estos elementos constituyen los pilares fundamentales de la arquitectura U-net, diseñada para tareas que requieren una comprensión profunda y precisa de señales complejas como los electrocardiogramas.

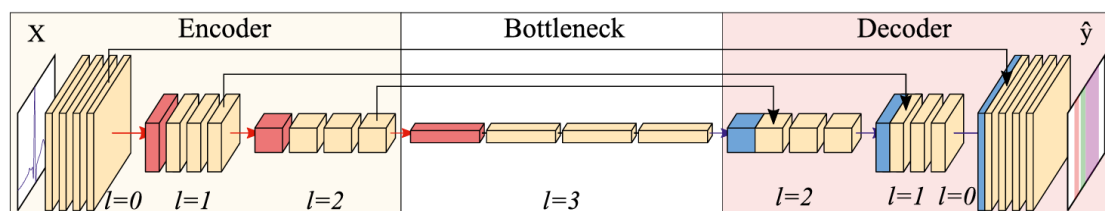


Figura 8: Representación de arquitectura U-net básica. Disponible en: [Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks - PubMed \(nih.gov\)](#)

1.4.3.3 Niveles

El codificador y el decodificador de una red neuronal U-net se estructura en diversos niveles, cada uno de los cuales realiza operaciones específicas destinadas a la manipulación de la señal de entrada. Cada nivel de la red está diseñado para realizar dos funciones principales: el diezmo o interpolación de la dimensión temporal de la señal y el aumento o reducción del número de filtros convolucionales, conocidos como canales o características.

En el codificador, cada nivel se inicia con una capa convolucional que duplica el número de filtros convolucionales de la capa anterior. Posteriormente, se aplica una normalización por lotes para

mitigar la posibilidad de sobreajuste, asegurando así que la red no memorice los resultados específicos de los casos de entrada durante el entrenamiento, y aumentar la estabilidad del modelo. Acto seguido, la salida de esta normalización se activa mediante la función RELU (explicada en el Anexo II), que introduce la no linealidad necesaria para la capacidad de aprendizaje de la red [13]. En la figura 9 se aprecian estas operaciones en el caso de una U-net para el procesamiento de imágenes, si bien en este trabajo la señal de entrada es unidimensional.

En el codificador se concatenan capas convolucionales para incrementar la profundidad y mejorar la capacidad de aprendizaje de la red neuronal. Por último, la dimensión temporal de la señal de entrada se reduce a la mitad en cada nivel mediante la técnica de maxpooling [35].

En el decodificador, se replican estas operaciones, si bien las convoluciones son sustituidas por convoluciones traspuestas. Además, el número de filtros convolucionales se reduce a la mitad en cada nivel, mientras que la dimensión temporal de la señal de entrada se duplica en cada nivel hasta alcanzar las dimensiones originales de la entrada [13].

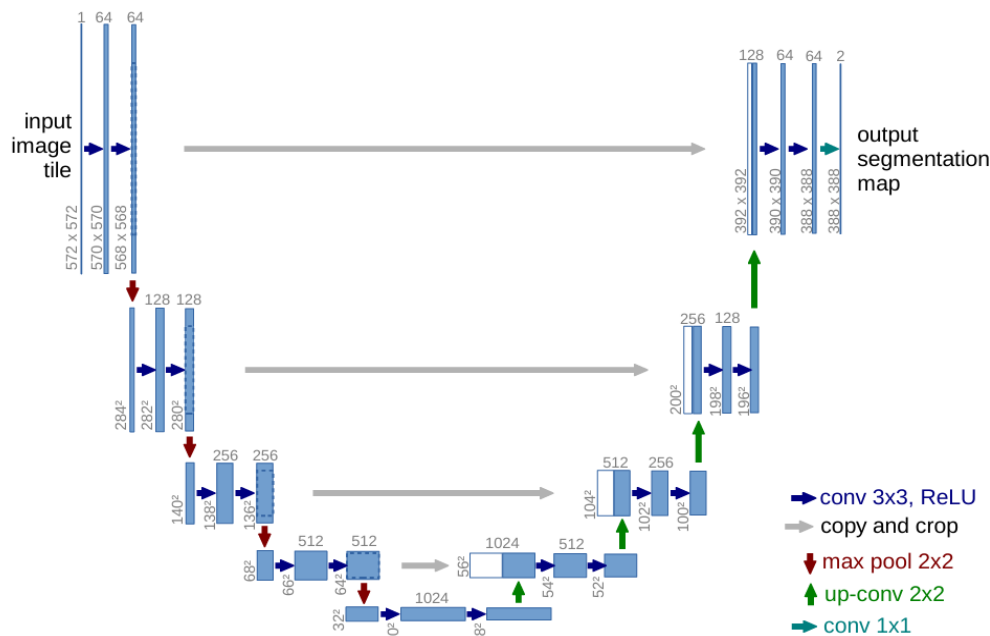


Figura 9: Representación de la red U-net 2D con las correspondientes dimensiones.

Disponible en: <https://arxiv.org/pdf/1505.04597>

1.4.4. Modificaciones U-Net

1.4.4.1 Bloques

Se han añadido una serie de bloques a la arquitectura U-NET básica para mejorar el desempeño de esta.

LSTM: La red LSTM (Long Short-Term Memory) es un tipo de red neuronal recurrente diseñada para aprender y recordar secuencias de datos a largo plazo. Utiliza una estructura de células con puertas de entrada, olvido y salida que controlan el flujo de información y permiten retener o descartar información según sea necesario.

Mecanismos de atención: Los mecanismos de atención en redes neuronales son técnicas que permiten a los modelos de aprendizaje automático centrarse en las partes más relevantes de los

datos de entrada cuando realizan una tarea específica, asignando diferentes niveles de importancia a diferentes partes de la entrada.

Una explicación más detallada de estos modelos se encuentra en el Anexo IV.

1.4.4.2 Conexión

Existen diferencias fundamentales entre la estructura de conexión del bloque LSTM y la del bloque de atención en la arquitectura propuesta. El bloque de atención se integra en cada nivel del decodificador, específicamente después de la convolución traspuesta, la cual se encarga de reducir el número de características. Esto se aprecia en la figura 10.

En contraste, el bloque LSTM se sitúa entre el codificador y el decodificador. A la salida de este bloque, se aplica una capa lineal para reducir a la mitad el número de canales, ya que el bloque LSTM, al ser bidireccional, procesa tanto el contexto pasado como el futuro, generando el doble de canales de salida. Además, el bloque LSTM se conecta al decodificador mediante una conexión residual, lo que permite preservar la información de la señal de entrada y evitar pérdidas significativas. Esta arquitectura se muestra en la figura 11.

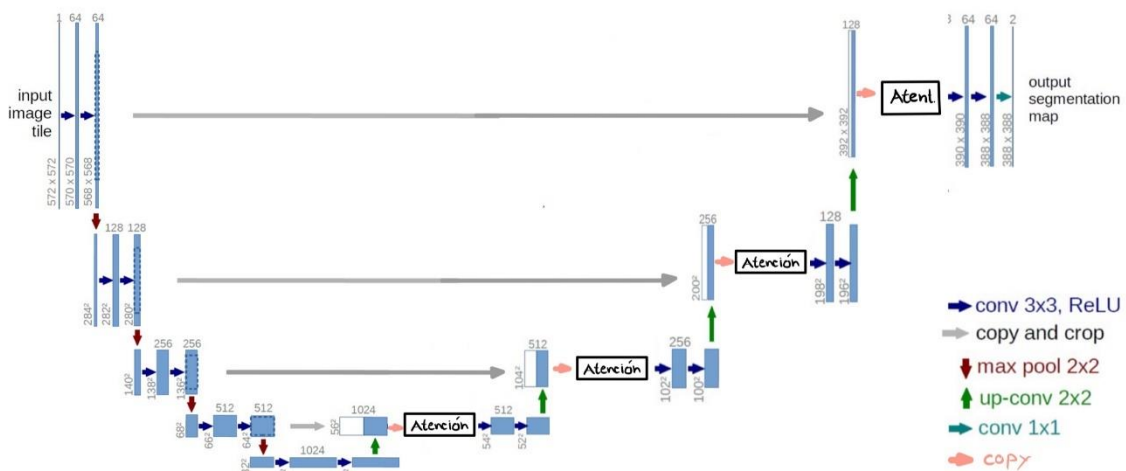


Figura 10: Representación de la red U-net 2D con la inclusión de bloque de atención.

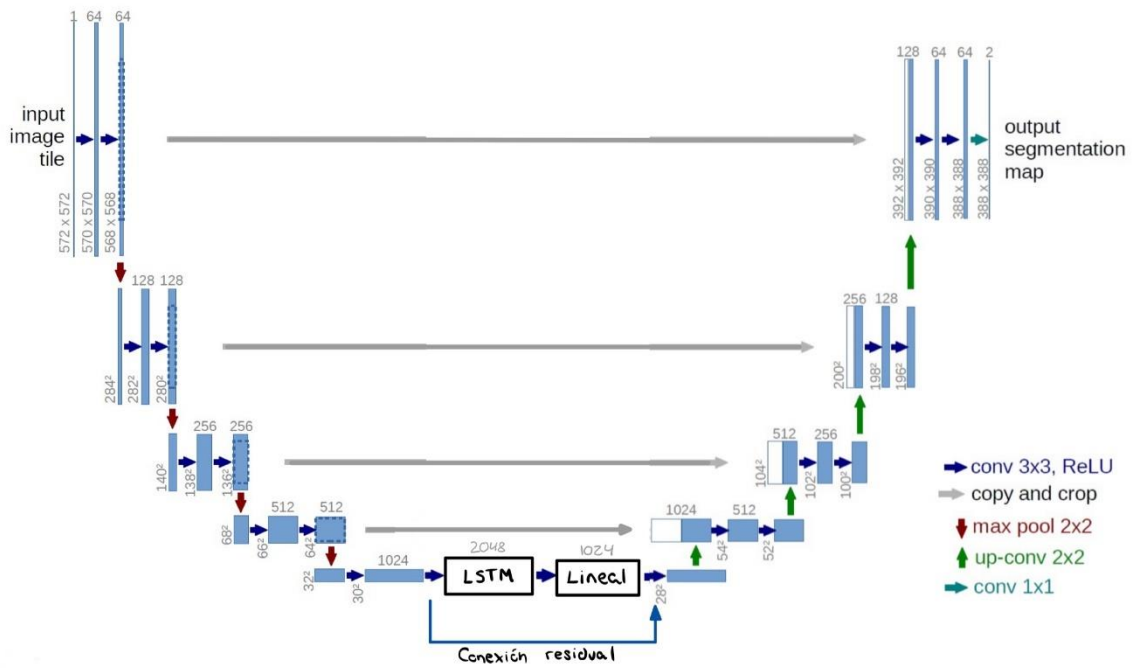


Figura 11: Representación de la red U-net 2D con la inclusión de bloque LSTM.

1.5. Objetivos

El planteamiento de este trabajo de fin de grado se realizó en base al objetivo de crear una red neuronal que fuese capaz de realizar delineaciones precisas y fiables de ECGs con el propósito de proporcionar un instrumento de apoyo a especialistas, en la detección de enfermedades cardiovasculares.

Adicionalmente, se plantea como objetivo secundario la comparación de los resultados obtenidos mediante este proyecto con aquellos obtenidos mediante métodos convencionales, tales como el delineador basado en la transformada Wavelet [14]. Se pretende evaluar si los resultados del presente proyecto muestran mejoras en la delineación de diferentes morfologías presentes en las ECGs con respecto al trabajo mencionado anteriormente, intentando crear un sistema robusto que permita mayor generalización que los algoritmos basados en el procesamiento de señal.

Finalmente, se persigue desarrollar un delineador que sea comparable a los resultados alcanzados en proyectos de delineación de electrocardiogramas basados en redes neuronales, indicando así unas prestaciones comparables a las de la literatura actual y sirviendo como base de futuros proyectos.

1.6. Recursos

Tanto el entrenamiento como el procesamiento y extracción de datos de la red se ha realizado gracias a las GPUs Nvidia RTX 3090 del clúster Hermes del Instituto de Investigación en Ingeniería de Aragón (I3A) utilizando el entorno de programación de Visual Studio, y con la colaboración del grupo Vivolab.

1.7. Organización del trabajo

La realización del trabajo se ha representado en la figura 12.

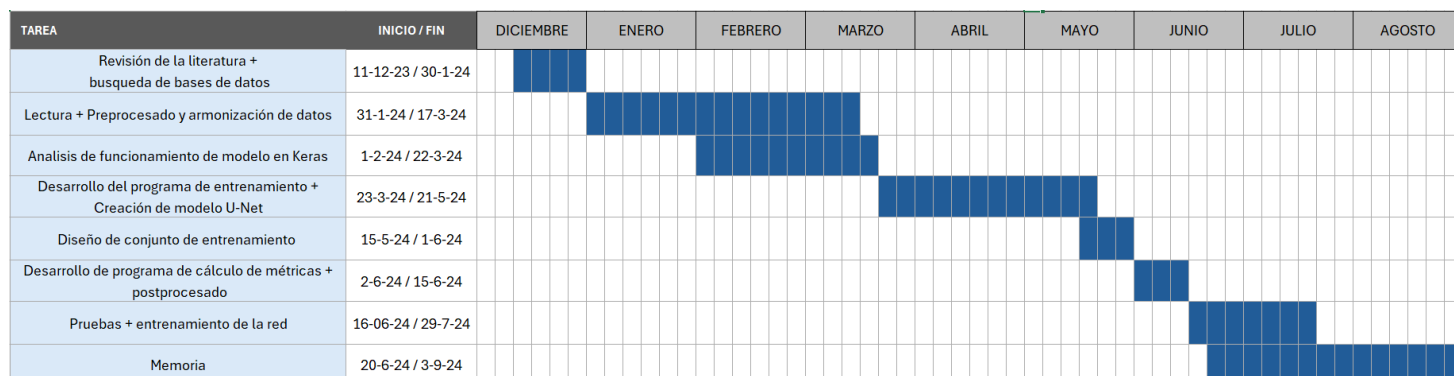


Figura 12: Diagrama de Gantt de la realización del trabajo.

2. MATERIALES Y MÉTODOS

2.1 Bases de datos

Las bases de datos elegidas contienen gran variabilidad de morfologías, permitiendo así que la red aprenda y sea capaz de generalizar a otros casos.

2.1.1 Base de datos de QT (QTDB)

Esta base de datos contiene 105 grabaciones de 15 minutos de 2 derivaciones muestreadas a 250 Hz con una variedad de patologías como arritmia, episodios de ST isquémicos y no isquémicos, deriva lenta del nivel de ST, depresión transitoria del ST y muerte cardíaca súbita. Todas las grabaciones contienen marcas de inicio, final y pico de las ondas P, QRS y T marcadas por un delineador automático y por dos cardiólogos distintos, sin embargo, solo contienen entre 30 a 100 latidos anotados debido a que la variabilidad en tiempo de un mismo paciente durante 15 minutos no es significativa. Si existen cambios en la morfología de los latidos a lo largo del registro ECG, estos latidos sí que se anotan [18] [19].

En total, la base de datos contiene 3622 latidos anotados por cardiólogos. Puede encontrarse más información sobre los latidos anotados y sus ondas puede acceder a: <https://archive.physionet.org/physiobank/database/qtddb/doc/node4.html>

Realmente se ha descargado una base de datos corregida, donde se han reanotado 112 latidos por ser extrasistólicos, los cuales no se detectaban ni delineaban, además de descartar algunos registros por contener anotaciones incompletas [14].

2.1.2 Base de datos de la Universidad Lobachevsky (LUDB)

La base de datos de Electrocardiografía de la Universidad Lobachevsky (LUDB) está formada por 200 registros de señales ECG de 10 segundos y 12 derivaciones (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) a 500 Hz, representando diversas morfologías de la señal de ECG. Los ECG fueron

obtenidos de voluntarios sanos y pacientes del Hospital No 5 de la Ciudad de Nizhny Novgorod entre 2017 y 2018. Los pacientes presentaban diversas enfermedades cardiovasculares, y algunos de ellos tenían marcapasos [20] [21] [22].

Los límites de las ondas P, T y complejos QRS fueron anotados manualmente por cardiólogos en los 200 registros. Además, cada registro está anotado con el diagnóstico correspondiente.

Concretamente, en total, la base de datos contiene 16.797 ondas P, 21.966 complejos QRS y 19.666 ondas T, lo que suma un total de 58,429 ondas anotadas. La edad de todos los voluntarios osciló entre los 11 años como mínimo y más de 89 años como máximo, con una edad promedio de 52 años. La distribución por género fue de 85 mujeres y 115 hombres.

Para más información sobre las diferentes morfologías que esta base de datos contiene acceda a: [Lobachevsky University Electrocardiography Database v1.0.1 \(physionet.org\)](http://Lobachevsky%20University%20Electrocardiography%20Database%20v1.0.1%20(physionet.org))

2.1.3 Base de datos de ECGs sintéticos

Como medida para compensar la insuficiencia de bases de datos accesibles, se ha adoptado la estrategia de incorporar electrocardiogramas sintéticos como parte del preentrenamiento de la red neuronal. Estos ECGs sintéticos se generan utilizando un simulador diseñado principalmente para la modelización de arritmias y ruido. Este generador se basa en cadenas de Markov discretas que posibilitan la simulación de ECGs con diversas patologías cardíacas, los cuales han sido validados por tres profesionales médicos distintos como representaciones prácticamente indistinguibles de ECGs reales. Por consiguiente, estos datos son sumamente útiles para el entrenamiento [23].

Específicamente, este estudio ha incluido la simulación de cuatro tipos diferentes de condiciones cardíacas. Entre estas, se encuentra el ritmo sinusal, ritmo sinusal con latidos ectópicos, fibrilación auricular y fibrilación auricular con latidos ectópicos (Ver anexo I para una explicación de estas afecciones).

De cada una de estas patologías se han simulado entre 500 y 3500 ECG diferentes, con distintas morfologías que corresponderían a la simulación de diferentes pacientes. Para cada uno de ellos, se generan 10 versiones con la misma señal ECG a la que se añaden distintas combinaciones y niveles de ruido, organizados como muestra la tabla 2 e identificando los ruidos como se muestra en la tabla 1 (véase el anexo I para una explicación detallada de los tipos de ruido).

3	6	8
Variaciones de línea de base	Ruido muscular	Artefactos de movimiento

Tabla 1: Tipos de ruido presentes en los electrocardiogramas sintéticos junto con un número de identificación.

Ruido	3	6	8	3, 6	3, 8	6, 8	3, 6, 8	3, 6, 8	3, 6, 8	0
RMS (mv)	0.065	0.007	0.25	0.085, 0.003	0.07, 0.15	0.0065, 0.35	0.055, 0.0035, 0.1	0.075, 0.05, 0.2	0.09, 0.009, 0.4	-

Tabla 2: Ruidos y combinaciones con sus correspondientes RMS utilizados en cada archivo de la base de datos sintética.

2.2 Extracción de datos

Ambas bases de datos se encuentran en formato WFDB, el cual es un estándar para almacenar señales biomédicas. En este trabajo se ha utilizado la librería de WFDB en Python para extraer los datos necesarios de las bases de datos.

Para leer los datos de las bases elegidas primero se tiene que leer la cabecera, la cual contiene los diferentes parámetros de los datos. Una vez leída se obtienen los datos y se guardan en formato CSV, tanto las anotaciones como los registros. En el Anexo III se explica detalladamente la lectura y escritura de datos.

Cabe mencionar que los objetivos usados para el entrenamiento de la red no van a ser las anotaciones de los cardiólogos, sino máscaras binarias que representan las muestras donde existe una onda. Se creará un vector de 1s y 0s para cada onda (máscara binaria) [14]. Inicialmente, se generará un vector compuesto por 0s cuya duración coincidirá con la duración de la señal ECG. Posteriormente, se leerán los inicios y finales de cada una de las ondas, y se asignarán 1s a las componentes del vector que se encuentra entre el inicio y final de cada onda concreta. De esta forma, se obtendrán vectores para cada una de las ondas (onda P, complejo QRS, onda T), indicando las muestras en las que estas están presentes, como se muestra en la figura 13.

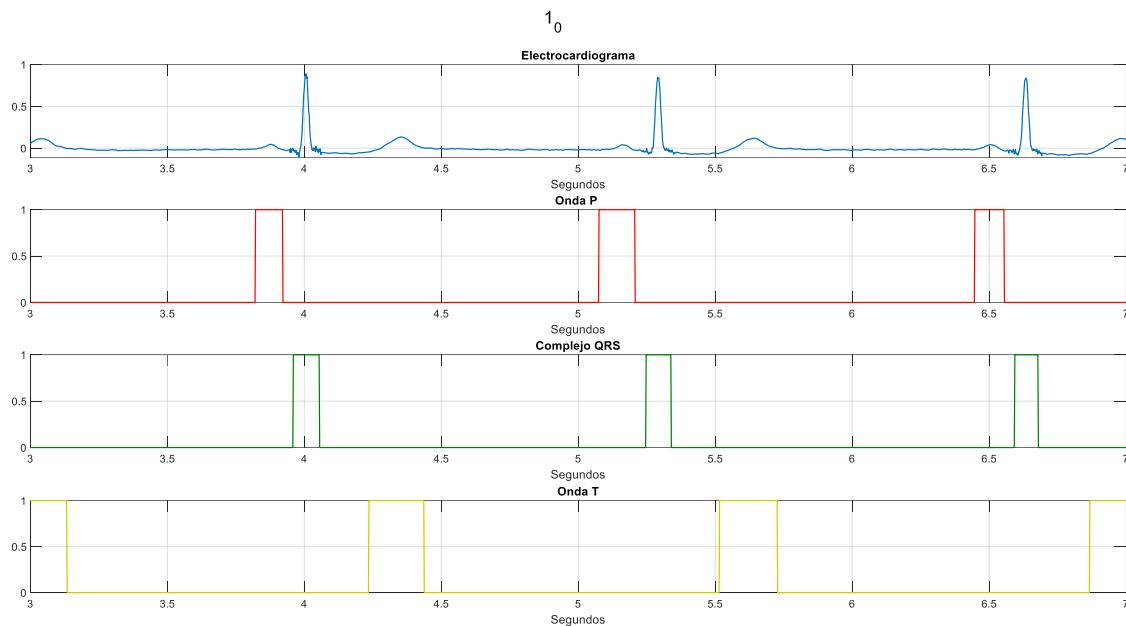


Figura 13: Ejemplo de máscaras binarias creadas a partir de las marcas de cardiólogos en la primera derivación del registro 1

2.3 Preprocesado de datos

2.3.1 Normalización Z-score

Adicionalmente, se incluye una normalización, llamada estandarización o normalización Z-score, para asegurar que no existan valores extremos en las bases de datos que repercutan en el entrenamiento de la ECG. En este caso no se usan los valores dados, sino que se normaliza mediante la media y la varianza de los datos de la base de datos de la siguiente forma:

$$x_n = \frac{x - \mu}{\sigma} \quad (2)$$

2.3.2 Interpolación

Una de las frecuencias de muestreo más comunes en registros de ECG es de 500 Hz, motivo por el cual se ha decidido que todas las señales de ECG utilizadas en este trabajo se ajustarán a dicha frecuencia.

La base de datos QT está muestreada a 250 Hz, por lo que será necesario interpolar la señal con un factor de 2. Para ello, se crea un vector de ceros cuya longitud es el doble de la señal original, donde se insertará la señal original cada dos muestras. Posteriormente, en las posiciones ocupadas por los ceros, se realiza un promediado entre las muestras adyacentes anterior y posterior, estableciendo dicho promedio en lugar del cero. Este procedimiento constituye una interpolación lineal.

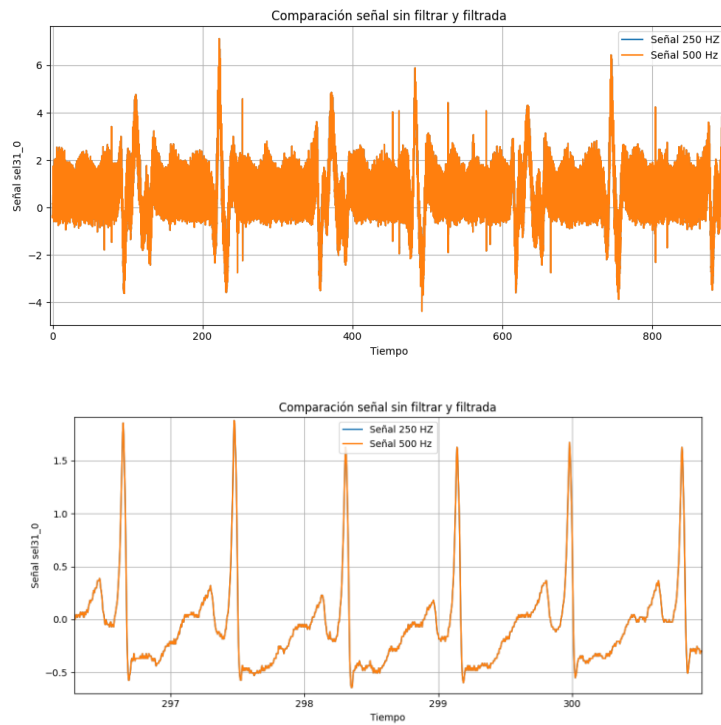


Figura 14, 15: Señal de la derivación 0 de la grabación sel_31, de la base de datos QT, a 250 Hz y a 500 Hz. Figura 14 señal completa, figura 15 señal ampliada.

Como se muestra en las figuras 14 y 15, al interpolar no se observan diferencias respecto a la señal original debido a que la interpolación realizada ha sido lineal.

2.3.3 Filtrado

Cuando se realiza un electrocardiograma siempre aparecen en la medición señales externas a las del corazón que se consideran ruido, además de artefactos o cambios por factores externo.

En el caso del entrenamiento de redes neuronales es interesante contemplar el ruido como forma de aleatorización de datos, permitiendo reducir la sobreadaptación de la red al no existir nunca un pulso exactamente igual al anterior. Adicionalmente, el ruido se puede introducir sobre la señal original como una estrategia de aumento de datos por lo que, en este trabajo, durante el entrenamiento de la red, no se realiza un filtrado de los datos.

El filtrado se añade en la fase de predicción y resultados, donde eliminar el ruido puede mejorar los resultados de la red. Para ello se aplican dos filtros distintos, uno paso bajo y otro paso alto, con especial cuidado en no eliminar factores relevantes de la señal.

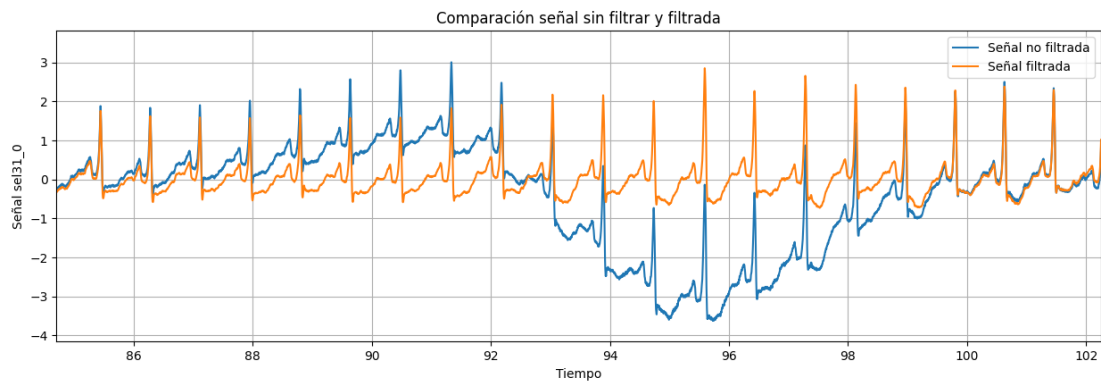


Figura 16: Representación de la derivación 0 de la grabación sel_31, de la base de datos QT, sin filtrar y filtrada.

La figura 16 representa un periodo de la derivación 0 de la grabación sel31 sin filtrar y filtrada. En esta figura se aprecia un movimiento de la línea de base que es debido a la respiración del paciente de esta grabación. El filtro paso bajo elimina estas oscilaciones, estableciendo una línea de base aproximadamente constante entorno al valor 0.

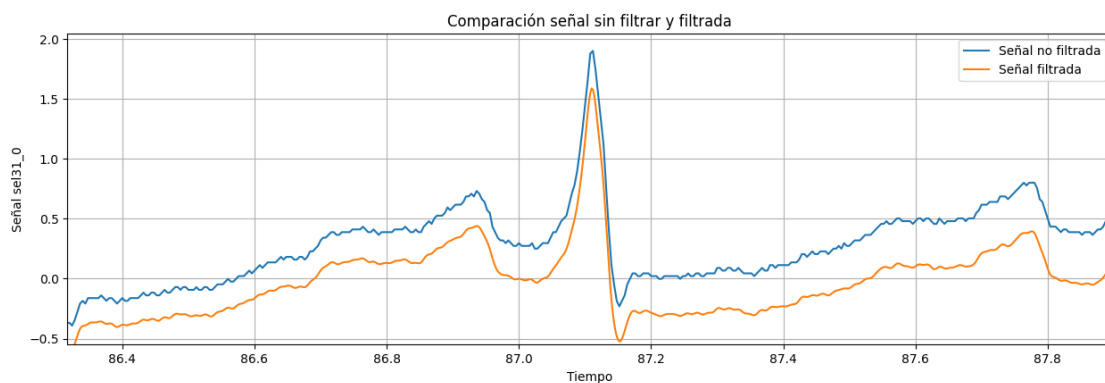


Figura 17: Representación de la señal ampliada de la derivación 0 de la grabación sel_31, de la base de datos QT, sin filtrar y filtrada.

La figura 17 es una ampliación de la figura 16, en donde se observa un ruido de alta frecuencia. Este ruido se elimina parcialmente mediante el filtro paso alto, sin embargo, el remanente del ruido es debido a que la frecuencia de corte no es lo suficientemente baja. Esto es a causa de que

si se reduce más la frecuencia de corte se suavizaría la señal, sobre todo los complejos QRS, perdiendo información clínica de la señal.

En el Anexo I se explica la procedencia del ruido del ECG y la elección de la frecuencia de corte.

2.3.4 Adaptación de señal QT

Como ya se ha mencionado anteriormente, la base de datos QT no está delineada completamente, sino que las marcas están concentradas en una parte de la señal pudiéndose delinear algún latido concreto, como muestra la figura 18.

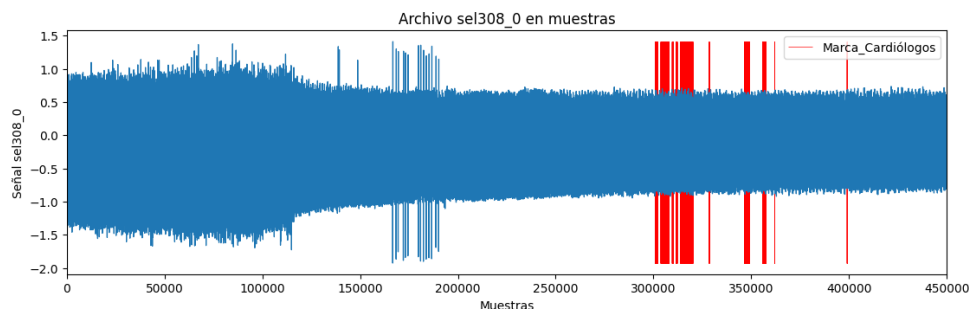


Figura 18: Representación de la derivación 0 de la grabación sel_31, de la base de datos QT, con las ondas anotadas por el primer cardiólogo.

Para solucionar este problema, se optó por aprovechar el formato de los datos utilizados durante el entrenamiento. Estos datos se presentan en forma de tensor, donde cada elemento corresponde a una agrupación que incluye el nombre del archivo junto con el inicio de la ventana de ese archivo a procesar (una explicación más detallada se encuentra en el Anexo IV). Esta estructura se justifica dado que los datos se introducen en la red neuronal en ventanas cuya duración es un parámetro de diseño. La elección de este parámetro se desarrolla en el apartado 3.2.1.

Se desarrolló un programa diseñado para recorrer las ventanas que se ingresarán como datos de entrada a la red, verificando que no contengan más de 1000 ceros consecutivos.

La elección de 0s consecutivos se fundamenta en el siguiente razonamiento: se estima que la frecuencia cardíaca mínima en reposo es de aproximadamente 40 latidos por minuto, considerando casos excepcionales. En consecuencia, teniendo en cuenta esta frecuencia, cada latido debería estar separado por un máximo de 750 muestras, dejando un margen adicional de otras 250 muestras debido a la presencia de enfermedades que pueden provocar la ausencia de alguna onda.

Por lo tanto, la ausencia de ondas durante 1000 ceros consecutivos es un indicativo de la falta de delineación de al menos un pulso en ese intervalo. En tal caso, se procederá a eliminar la pareja [nombre del archivo, inicio de la ventana] del tensor, lo cual implica que la red neuronal no considerará las ventanas que no han sido adecuadamente delineadas.

Este proceso se aplica únicamente en la delineación por cardiólogos en la base de datos QT debido a la distribución irregular de las marcas de esta base de datos.

2.3.5 Generación de máscara null

Debido al planteamiento de la red neuronal como clasificador, es necesario generar una máscara adicional que identifique las secciones de la señal que no contienen ninguna onda. Para ello, se verificó la inexistencia de solapamientos entre las tres ondas en todos los conjuntos de datos, ya que la presencia de solapamientos indicaría la existencia de más de una onda en una misma muestra, lo cual no es posible.

Una vez confirmada la ausencia de solapamientos, para generar la máscara null se realiza una operación OR con las máscaras P, QRS y T, obteniendo 0s en las muestras que no contienen ninguna onda. Posteriormente, al invertir el vector binario resultante, se obtiene la máscara null. Este proceso se representa en la figura 19.

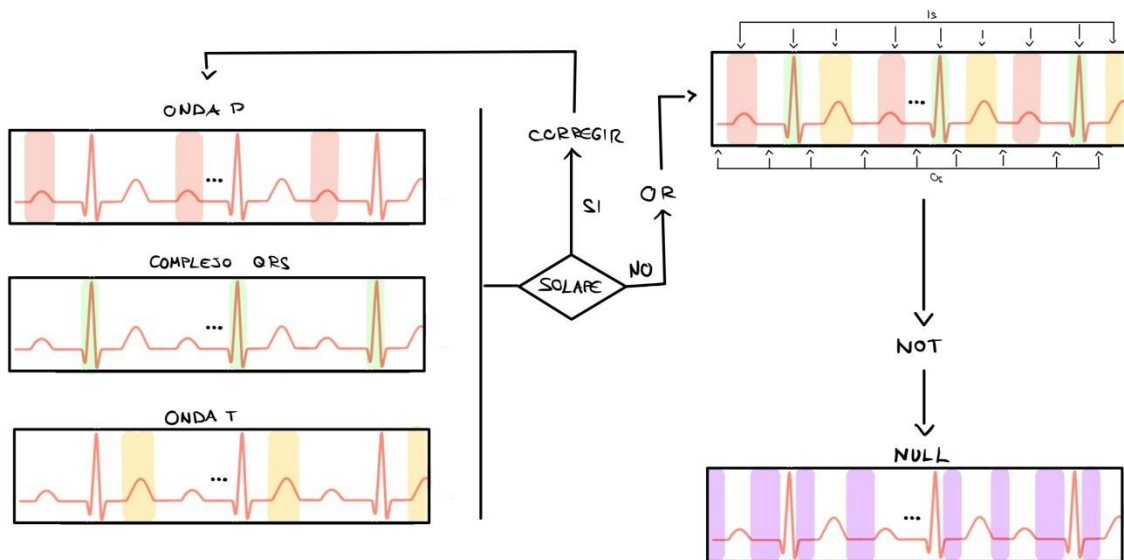


Figura 19: Esquema de generación de máscara null

2.3.6 Aumento de datos en la base de datos LUDB

La base de datos sintética obtenida del generador de ECGs proporciona la señal de ruido de cada ECG de manera separada, lo que permite la extracción del ruido sin señal. Este ruido es similar al que podría presentarse durante la realización de un ECG en un paciente real, lo cual resulta muy útil para el entrenamiento de la red neuronal.

Dado que la longitud en muestras de la señal de ruido extraída es de 7500 muestras, se ha decidido modificar únicamente la base de datos Lobachevsky, cuyas señales tienen una longitud de 5000 muestras. La base de datos QT contiene grabaciones de 15 minutos de duración, las cuales superan considerablemente la longitud de las señales de ruido simuladas, impidiendo su utilización. En la figura 20 se muestra el proceso de elección y adición de ruido.

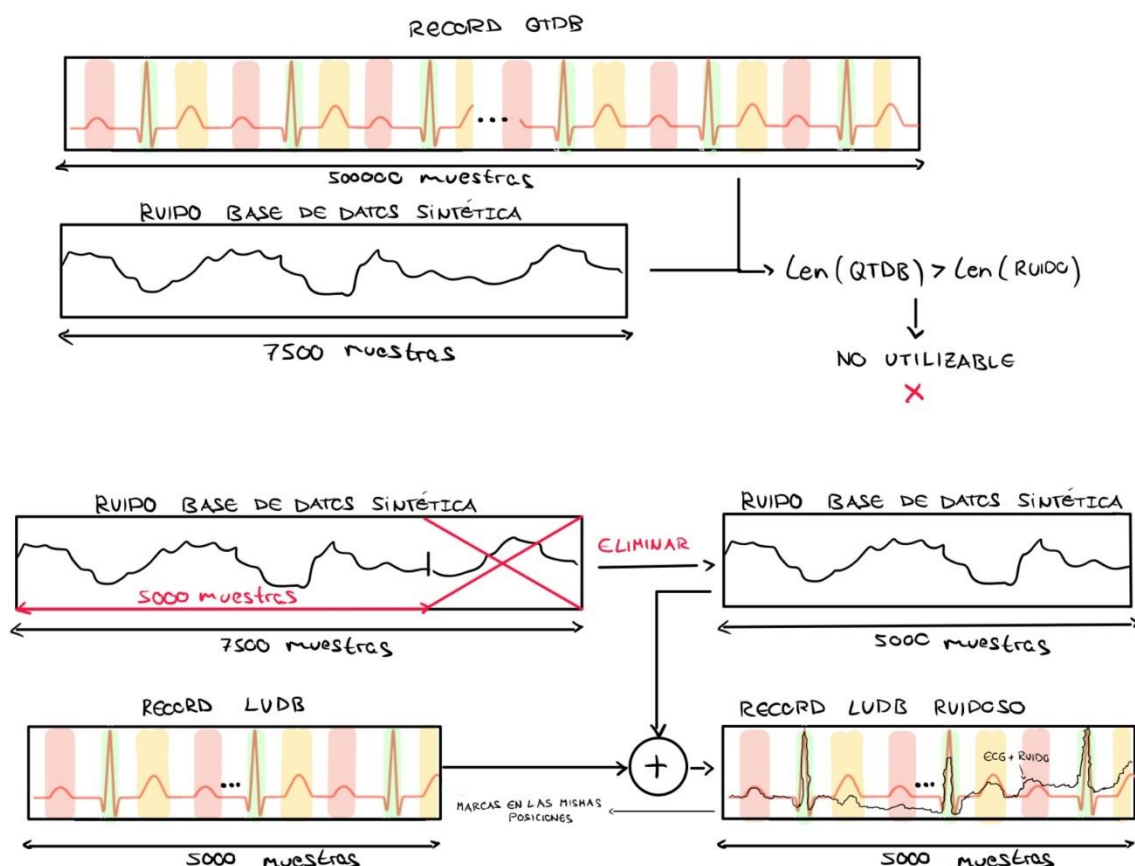


Figura 20: Esquema de adición de ruido sintético.

Para ello se han considerado las realizaciones de ruido simuladas para 300 de los registros sintéticos obtenidos con el simulador. De este modo, para cada registro de la LUDB se generan dos grabaciones adicionales con dos ruidos distintos, seleccionados aleatoriamente entre las nueve opciones disponibles (ver tablas 1 y 2).

A pesar de la adición de ruido a una señal de la base de datos Lobachevsky, las marcas de inicio y finalización de las ondas permanecen inalteradas, ya que la señal subyacente no se modifica. Por consiguiente, las nuevas grabaciones generadas conservarán las mismas marcas que las grabaciones originales correspondientes.

En consecuencia, se han generado 400 grabaciones adicionales de 10 segundos cada una. Las anotaciones de referencia para ellas son las realizadas por cardiólogos sobre las señales sin ruido.

2.3.7 Delineación de datos sintéticos

El generador de ECGs sintéticos no permite obtener los inicios y finales de las ondas generadas, lo cual limita su utilidad como fuente de datos de entrenamiento, ya que carece de las etiquetas correspondientes. Para paliar esta limitación, se ha utilizado utilizar un delineador automático, obteniendo así las anotaciones de referencia para los ECGs sintéticos.

Para ello, se ha empleado un delineador de una sola derivación basado en la transformada wavelet, reconocido por su robustez y alta precisión [24]. Se han delineado las primeras 400 carpetas correspondientes a todas las condiciones cardíacas, excepto las arritmias, debido a la irregularidad de los ritmos cardíacos en dicha condición como se puede apreciar en la figura 21, lo cual podría

dificultar la delineación y afectar negativamente los resultados del entrenamiento de la red. Esta delineación se realiza únicamente en los ECGs que no contienen ruido, con el fin de obtener un resultado más preciso. Posteriormente, las marcas obtenidas se aplican al resto de los archivos con ruido, dado que la señal subyacente sigue siendo la misma.

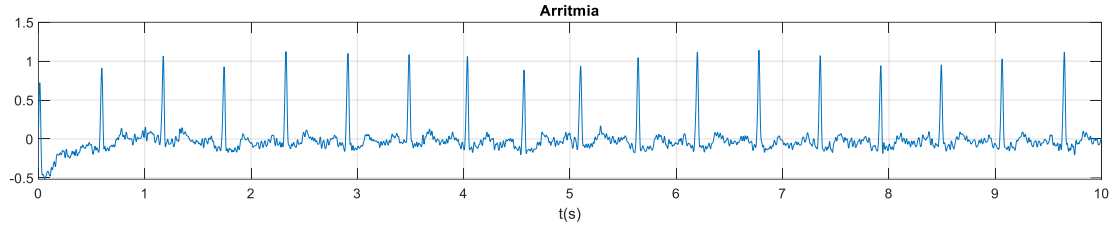


Figura 21: Electrocardiograma de una taquicardia supraventricular obtenido de la base de datos sintética.

Es importante destacar la delineación de la fibrilación auricular, que se ha realizado utilizando el archivo que contiene exclusivamente información de la activación ventricular, específicamente de las ondas QRS y T (ver Anexo I para información detallada del ciclo cardíaco y el electrocardiograma). Esto se debe a que, como se observa en la figura 22, en la fibrilación auricular no debería existir la onda P y, al tratarse de un delineador basado en DSP, el cual no es aplicable a cualquier afección cardíaca, existe la posibilidad de que detecte una onda P aunque esta no esté presente.

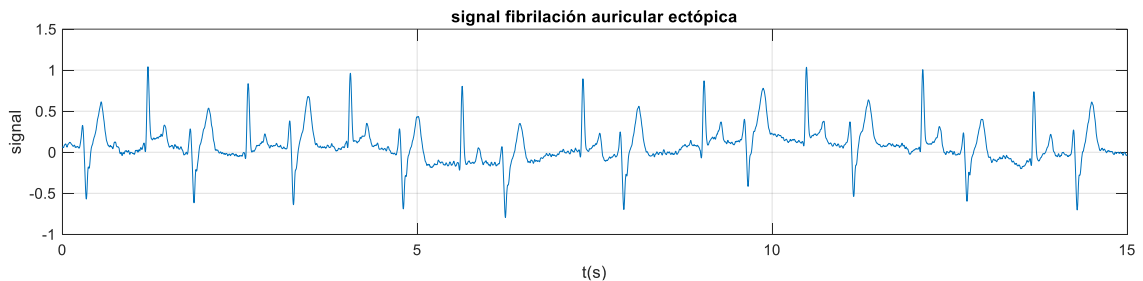


Figura 22: Electrocardiograma de la activación ventricular de paciente con fibrilación auricular ectópica obtenida de la base de datos sintética.

3. ENTRENAMIENTO DE LA RED

3.1 Estrategia de entrenamiento

3.1.1 Entrenamiento y preentrenamiento

En este trabajo, se ha decidido realizar un preentrenamiento con el objetivo de mejorar las métricas de la red y mitigar la escasez de datos de entrenamiento. El preentrenamiento implica entrenar la red utilizando anotaciones automáticas de la base de datos QTDB y registros generados por un generador sintético, y cuyas anotaciones son proporcionadas por un delineador automático. De esta manera, se espera alcanzar un rendimiento comparable al de los delineadores automáticos.

En la segunda fase, la red neuronal se entrena a partir de los pesos obtenidos durante el preentrenamiento. En esta etapa, se utiliza tanto la base de datos QTDB como la LUDB, pero ahora con anotaciones realizadas por cardiólogos expertos. Este proceso permite mejorar las métricas de las anotaciones automáticas al ajustar ligeramente los pesos de la red.

3.1.2 Validación cruzada

Una estrategia muy común usada en el entrenamiento de redes neuronales es la validación cruzada. Esta técnica consiste en la división del dataset en k subconjuntos (folds) del mismo tamaño, donde n de estos subconjuntos se utilizarán para entrenar la red y m como datos de validación [2] [14] [25] [26] [27]. En este trabajo en concreto se ha utilizado un 80% del dataset en la etapa de entrenamiento y validación y un 20% para evaluación. El 20% de ese 80% se ha utilizado para validación y el resto para entrenamiento. La figura 23 muestra la organización de los datos.

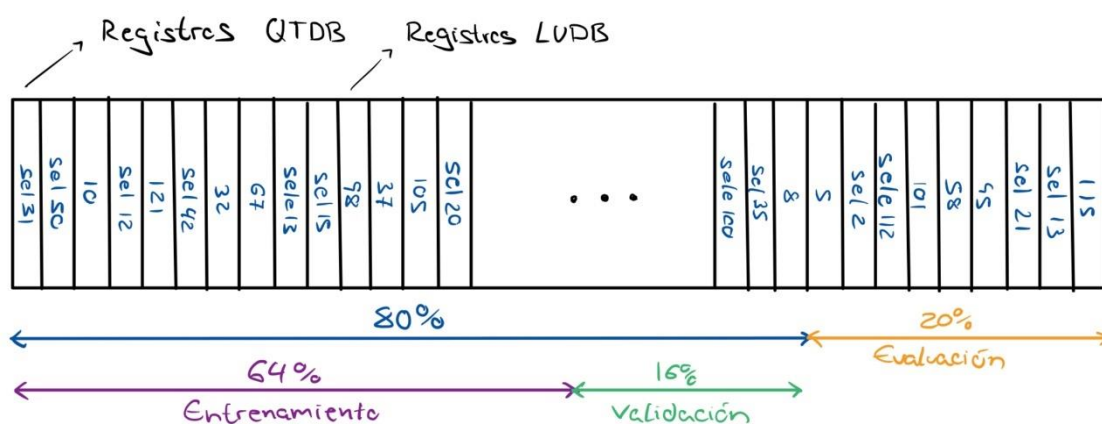


Figura 23: Organización de los datos para validación cruzada

Esta técnica es utilizada para evaluar la capacidad de generalización de un modelo de aprendizaje automático, asegurando que el modelo funcione bien no solo en los datos de entrenamiento sino también en datos no vistos. Esta técnica es particularmente útil para evitar el sobreajuste, donde el modelo se ajusta demasiado a los datos de entrenamiento y falla al generalizar a datos nuevos.

Cabe mencionar que tanto el subconjunto de validación y entrenamiento como el de evaluación se han obtenido de forma aleatoria y asegurando que todos los datos del mismo paciente están en el mismo subconjunto, para que no existan sesgos en los resultados por elección indebida.

3.1.3 Inicialización de pesos

La inicialización de los pesos se realiza siempre con la misma semilla, de tal forma que se garantiza que los resultados obtenidos en diferentes ejecuciones del mismo experimento sean idénticos, permitiendo verificar y validar resultados de manera consistente (reproducibilidad) [2].

3.2 Introducción de datos en la red

3.2.1 Dimensiones

Es importante definir las dimensiones de entrada a la red. En este trabajo la entrada (electrocardiograma) se divide en ventanas temporales por ser secuencias relativamente largas. De esta forma, la red captura mejor las relaciones locales que existen en los datos de entrada y permite simplificar el código debido a la obtención de secuencias de longitud fija.

Se ha elegido un tamaño de ventana de 1024 muestras ya que, considerando que la frecuencia cardiaca promedio sean 60 pulsaciones por segundo, se obtienen aproximadamente entre uno y tres latidos por muestra, siendo la red capaz de aprender relaciones temporales entre latidos (por abarcar más de un latido) y patrones dentro de cada latido

Se realiza una agrupación en lotes de datos de 64 ventanas (batch) con la finalidad de procesar múltiples ventanas simultáneamente aprovechando el cómputo paralelo de las GPUs utilizadas y la optimización de la memoria (En el anexo II se desarrolla la elección del tamaño del batch).

Como muestra la figura 24, los datos de entrada serán tensores de dimensiones (64, 1024,1) correspondientes a (batch_size, timesteps, input_dim). De igual forma, el tensor de salida tendrá las dimensiones (64, 1024, 4), las mismas a excepción del número de clases, que serán 4 (P, QRS, T, NULL).

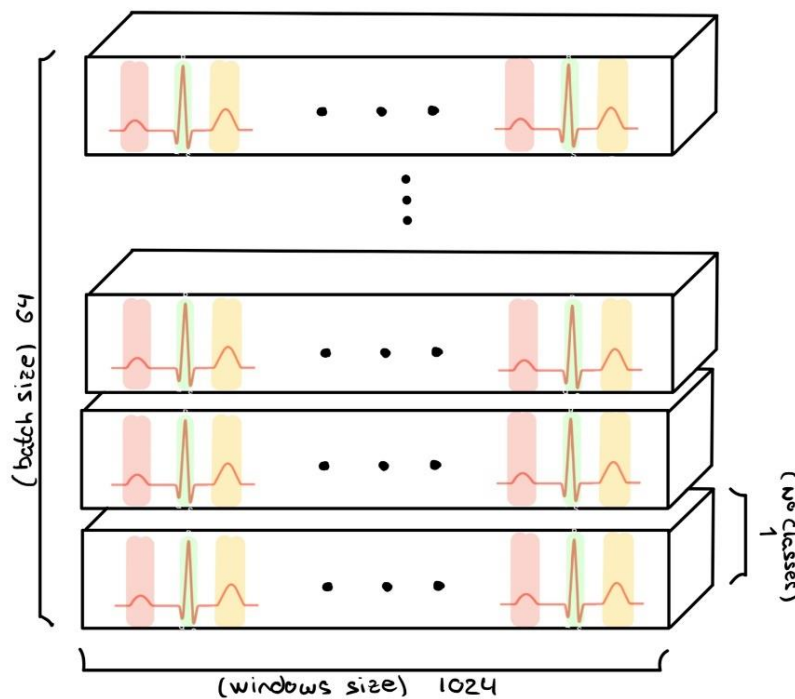


Figura 24: Representación del tensor de entrada a la red U-NET

3.2.2 Solapamiento

El solapamiento de ventanas es una estrategia fundamental para capturar la dinámica temporal en el análisis de señales, ya que permite que cada punto de la señal contribuya a múltiples ventanas, mejorando así la representación de la evolución temporal de la señal [14]. Se obtiene por tanto una reconstrucción de la señal mediante las distintas ventanas.

Esta técnica es especialmente relevante en el procesamiento de electrocardiogramas (ECG), donde la señal presenta cambios abruptos. Sin solapamiento, los inicios y finales de las ventanas podrían seccionar una onda, como se ilustra en la Figura 25, lo que impediría una captura precisa de la señal durante el entrenamiento del modelo y aumentaría significativamente la varianza de las estimaciones.

En el presente estudio, se ha aplicado un solapamiento del 87.5%, lo cual implica que cada muestra de la señal ECG se procesa hasta ocho veces por la red neuronal. La Figura 25 muestra claramente este proceso.

Cabe mencionar que la reconstrucción de la señal es necesaria únicamente en la validación y evaluación de la señal, donde se requiere la señal completa.

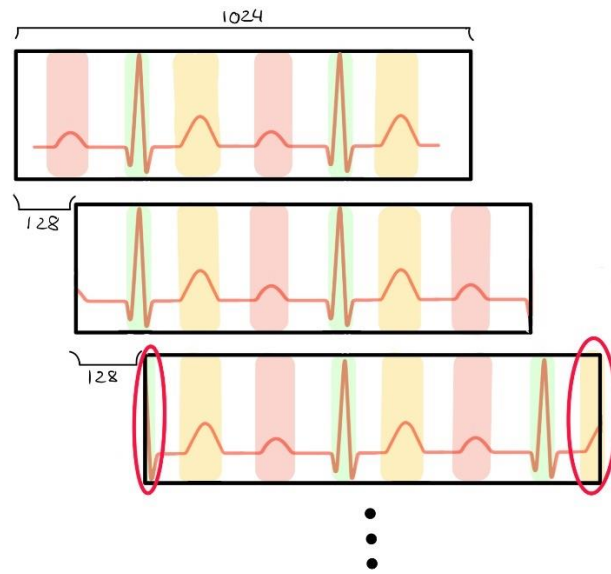


Figura 25: Representación del solapamiento de las ventanas a la entrada de la red U-NET

3.2.3 Máscara (función de coste)

Las bases de datos empleadas en este estudio no están completamente delineadas por cardiólogos. En el caso de la base de datos LUDB, ni el inicio ni el final de cada registro están anotados, ya que estas secciones pueden contener errores indeseados. Por otro lado, en la base de datos QTDB, los registros están delineados solo en ciertas secciones debido a su larga duración; al ser la señal cardíaca de naturaleza periódica, una delineación exhaustiva de un ECG de un mismo paciente no aportaría información adicional.

Las secciones que carecen de anotaciones, pero no de actividad cardíaca, requieren una función de coste modificada. Si estas secciones se consideran de forma incorrecta, el comportamiento de la red neuronal se deterioraría, ya que se podrían interpretar erróneamente como zonas isoelectricas (nulas) cuando no lo son. Para evitar este problema, se implementa una función de coste personalizada.

Primero se calcula la cross-entropía (CE), que mide la diferencia entre la distribución de probabilidad real (las etiquetas verdaderas) y la distribución estimada por un modelo. En el formato de este trabajo (one-hot), las etiquetas son vectores en los que solo una clase tiene probabilidad 1, y las demás tienen probabilidad 0. Se calcula usando la función de entropía cruzada de pytorch, explicada en el Anexo II.

Posteriormente, se crea una máscara que indique las áreas con anotaciones válidas. Para construir esta máscara, se concatenan las máscaras binarias de todas las ondas de cada registro, considerando las secciones con actividad cardíaca sin anotar como un intervalo de 500 muestras entre anotaciones. Este intervalo es suficiente para inferir la presencia de un latido entre las marcas, aunque no esté anotado. Además, se añaden 20 muestras adicionales al inicio y al final de estos intervalos para proporcionar un margen respecto al final de la onda. La máscara resultante contiene valores de 1 en las áreas con anotaciones y 0 en las demás.

Una vez obtenida la máscara binaria de cada registro, se multiplica por el resultado de la función de pérdida de entropía cruzada (cross-entropy).

Finalmente, se realiza la suma de todo el vector resultante y se divide por la suma de la máscara binaria aplicada. Este cálculo genera un valor único que se utiliza para actualizar los pesos de la red neuronal. De esta manera, las áreas donde la máscara binaria es 0 no influirán en la actualización de los pesos.

El proceso del cálculo de la función de pérdida se representa en la figura 26.

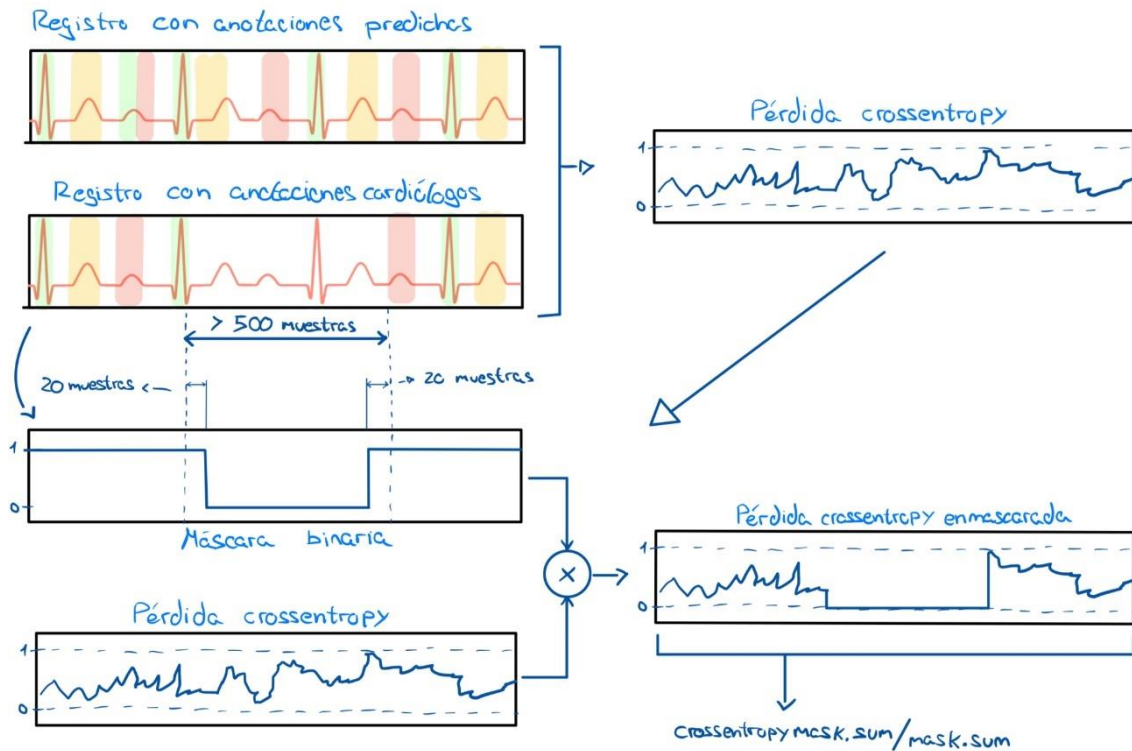


Figura 26: Esquema del cálculo de la función de pérdida personalizada.

3.3 Métricas

En el contexto de la evaluación de la delineación de un electrocardiograma (ECG) mediante redes neuronales, es fundamental emplear una variedad de métricas que proporcionen una visión exhaustiva del rendimiento del modelo. Entre las más relevantes se encuentran la precisión, el recall, la F1-score y la matriz de confusión, así como métricas específicas para el análisis temporal y de localización, como el onset, offset y la desviación estándar de los bordes de las ondas.

La precisión o valor predictivo positivo se define como la proporción de verdaderos positivos sobre el total de predicciones positivas realizadas por el modelo, es decir, identifica la proporción de ondas correctas detectadas con respecto a todas las detectadas [29].

$$Sensibilidad = \frac{TP}{TP+FP} \quad (3)$$

Por otro lado, el recall o sensibilidad es una métrica que mide la frecuencia con la que un modelo de aprendizaje automático identifica correctamente los verdaderos positivos entre todas las ondas positivas reales del conjunto de datos [29].

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

El coeficiente DICE representa la medida de concordancia entre las máscaras binarias generadas por la red neuronal para cada onda del electrocardiograma y las correspondientes máscaras provenientes de las anotaciones realizadas por cardiólogos o sistemas de delineación automática. En esencia, evalúa la similitud entre ambas representaciones binarias para cada tipo de onda cardíaca [30].

$$DICE = 2 * \frac{|A \cap B|}{|A+B|} \quad (5)$$

Este coeficiente ha inspirado funciones de coste para entrenar redes de segmentación [14].

Por último, se evaluarán los errores temporales, tanto el inicio (onset) como el final (offset) de los bordes de las ondas. Se cuantifica la desviación media de los errores en los puntos de inicio y finalización de la onda generada por la red neuronal respecto a las etiquetas de referencia. Además, se realiza el cálculo de la desviación estándar tanto del onset como del offset para analizar la variabilidad en estos puntos.

$$\mu = \frac{\sum_{j=1}^p \sum_{i=1}^l |x_{i,j} - m_{i,j}|}{p} \quad (6)$$

$$\sigma = \sqrt{\frac{1}{p} * \sum_{j=1}^p (\sum_{i=1}^l |x_{i,j} - m_{i,j}| - \mu)^2} \quad (7)$$

Parámetros:

μ = Media de las sumas de los errores de cada registro.

σ = Varianza de las sumas de los errores de cada registro.

$x_{i,j}$ = Marca i del registro p obtenida por la red entrenada.

$m_{i,j}$ = Anotación i del registro p considerada la solución correcta.

Cabe mencionar que el cálculo de estas métricas se basa en la detección o no de las ondas, sin embargo, en la prueba de robustez de la red U-NET este cálculo se realiza muestra a muestra ya que las métricas muestran una mayor variabilidad, cualidad que se busca para medir la robustez.

3.4 Pruebas

3.4.1 Robustez U-NET

Una de las pruebas más importantes y frecuentemente olvidadas es conocer la variabilidad de la red, es decir, conocer la varianza que tienen los resultados y así saber si el resultado es consistente o simplemente producto de la casualidad. Para lograr esto, se fijan todos los parámetros de la red, así como los datos de entrada, y se ejecuta varias veces el mismo proceso.

En este caso, se ha elegido una red U-NET con 4 niveles y una profundidad de 5 capas por nivel, una red intermedia a las futuras variaciones de esta.

Se han realizado 4 ejecuciones de 70 epochs cada una con todos los datos de entrenamiento (aumento de datos incluido), guardando las métricas de precisión, recall y pérdida de los datos de validación y entrenamiento (estas métricas se explican en la sección 3.1) en cada una de las epochs de las 4 ejecuciones.

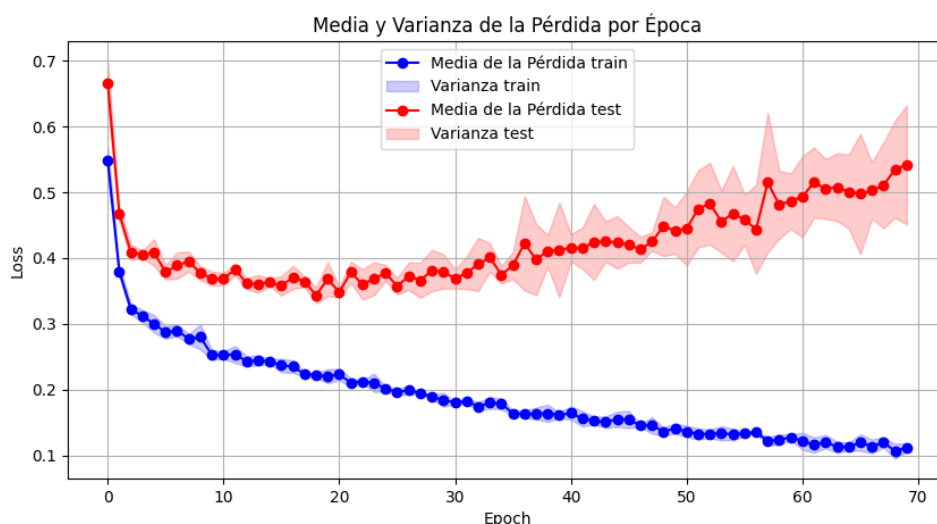


Figura 27: Media y varianza de función de pérdida (cross entropy) durante el entrenamiento de la red U-NET de 4 niveles y 4 capas con datos de entrenamiento.

En la Figura 27, se observa un aumento de la varianza en la función de pérdidas en el conjunto de prueba conforme incrementa el número de epochs. A partir de la epoch 25, este incremento se vuelve significativo, lo cual se considera un indicativo de que la red ha comenzado a sobreajustar el conjunto de datos. Este fenómeno de sobreajuste se evidencia adicionalmente en la diferencia entre las pérdidas de las funciones de prueba y entrenamiento. La red neuronal empieza a memorizar los datos de entrenamiento, lo que impide su capacidad de generalización. Por lo tanto, los resultados en el conjunto de prueba se vuelven cada vez más erráticos y menos representativos de la capacidad de la red para manejar datos no vistos previamente.

Si se quiere realmente conocer la robustez de la red neuronal y asegurar que una ejecución sea suficiente para obtener un resultado representativo del aprendizaje de la red se tienen que observar las métricas de precisión y recall.

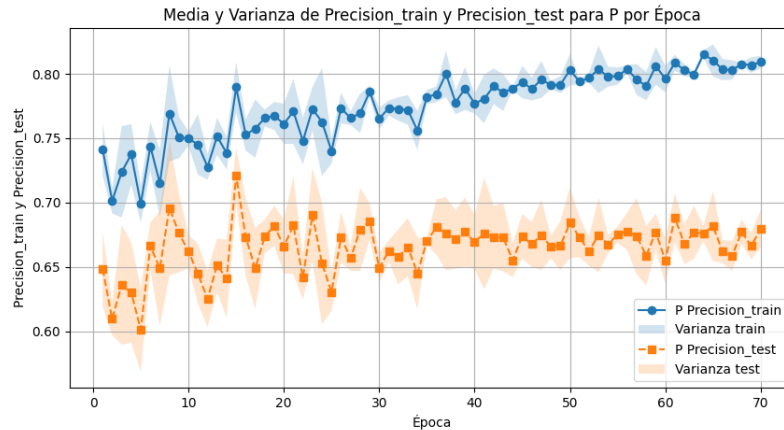


Figura 28: Media y varianza de la precisión de la onda P durante el entrenamiento de la red U-NET de 4 niveles y 4 capas con datos de entrenamiento.

En la Figura 28, se ha optado por representar la precisión de la onda P, dado que esta métrica presenta la mayor varianza entre todas las consideradas (el conjunto completo de gráficas se encuentra en el Anexo IV). No obstante, la varianza observada en el conjunto de prueba es notablemente baja y la media no exhibe grandes oscilaciones. En conclusión, los resultados indican que, con los datos de entrenamiento disponibles, la red neuronal demuestra una notable robustez, no siendo necesario realizar múltiples ejecuciones del mismo proceso para obtener resultados concluyentes.

De igual forma que con la función de pérdida, todas las métricas presentan una desviación de los datos de entrenamiento con respecto a los de validación conforme aumenta el número de epochs debido al sobreajuste. En esta prueba el sobreajuste es indiferente debido a que únicamente se mide la robustez de la red, sin embargo, en las pruebas restantes será necesario delimitar a partir de qué número de epochs sucede.

Debido a la diferencia de tamaños y procedencias de los datos de entrenamiento y preentrenamiento se ha considerado realizar adicionalmente esta prueba para los datos de entrenamiento con la misma red. En este caso el solape es del 25% con la intención de que la proporción de los datos de entrenamiento y preentrenamiento sea razonable ya que el tamaño de la base de datos de preentrenamiento es muy superior al de entrenamiento.

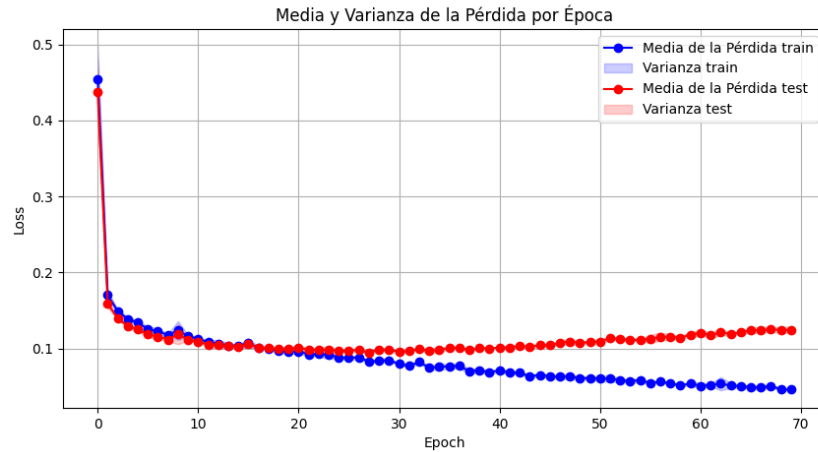


Figura 29: Media y varianza de la función de pérdida durante el entrenamiento de la red U-NET de 4 niveles y 4 capas con datos de preentrenamiento.

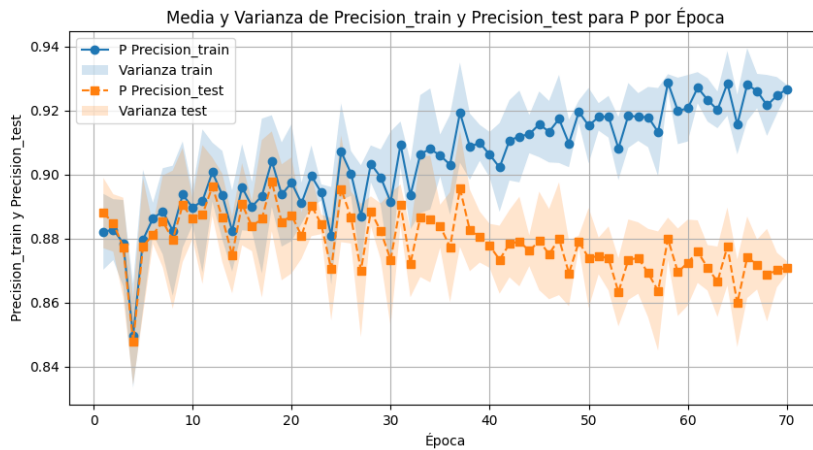


Figura 30: Media y varianza de la precisión de la onda P durante el entrenamiento de la red U-NET de 4 niveles y 4 capas con datos de preentrenamiento.

Las figuras 29 y 30 muestran que los datos de preentrenamiento también presentan una varianza baja, concluyendo así que el modelo es robusto a la hora de entrenar con cualquier base de datos (el conjunto completo de gráficas se encuentra en el anexo IV).

Cabe mencionar que el sobreajuste se produce mucho más tarde comparado con los datos de entrenamiento debido a que, como ya se ha mencionado, el tamaño de la base de datos de preentrenamiento es mayor. Esto también provoca que la varianza de datos sea menor.

3.4.2 Profundidad y niveles U-NET

Unos de los hiperparámetros que se debe definir a la hora de realizar el entrenamiento son la profundidad de la red y el número de niveles que esta tiene.

No siempre es idóneo escoger la red que tenga un mayor valor debido a que, al tener mayor número de parámetros, la red tiene mayor capacidad para memorizar el conjunto de

entrenamiento, sobreadaptándose a los mismos. Adicionalmente, si el conjunto de datos no es lo suficientemente grande para soportar el tamaño de la red, esta no aprenderá de forma efectiva. Finalmente, el coste computacional de las redes con gran cantidad de parámetros hace que no sean útiles para tareas a tiempo real o con baja latencia.

Para poder escoger estos parámetros se han realizado nueve entrenamientos distintos con nueve configuraciones diferentes con los datos de entrenamiento, guardando las métricas de precisión y recall de cada epoch, así como la pérdida. Todas las pruebas se han realizado con un ratio de aprendizaje de 0.0003 y 15 epochs.

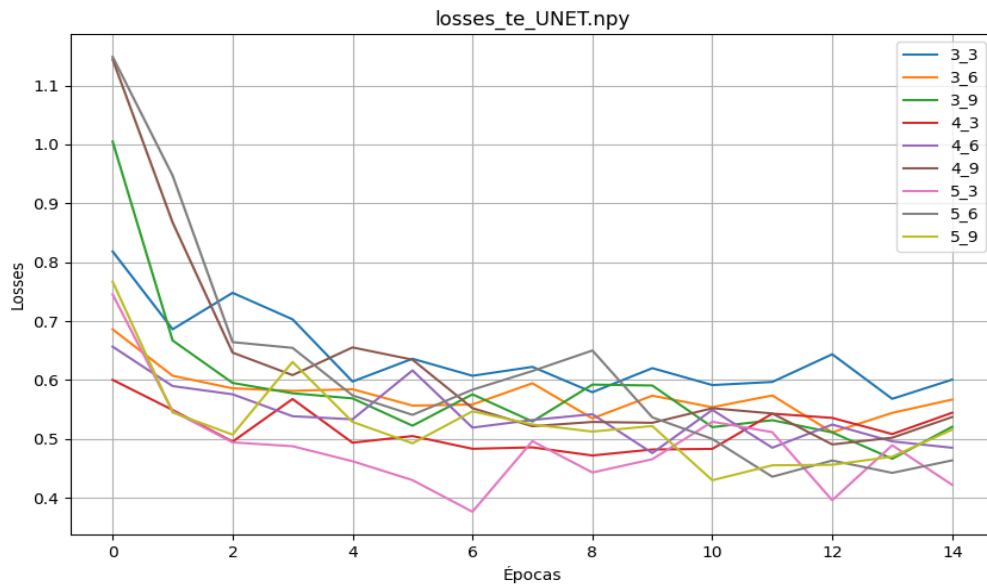


Figura 31: Función de pérdida durante el entrenamiento con las bases de datos QT y Lobachevsky obtenida con los datos de entrenamiento para cada configuración.

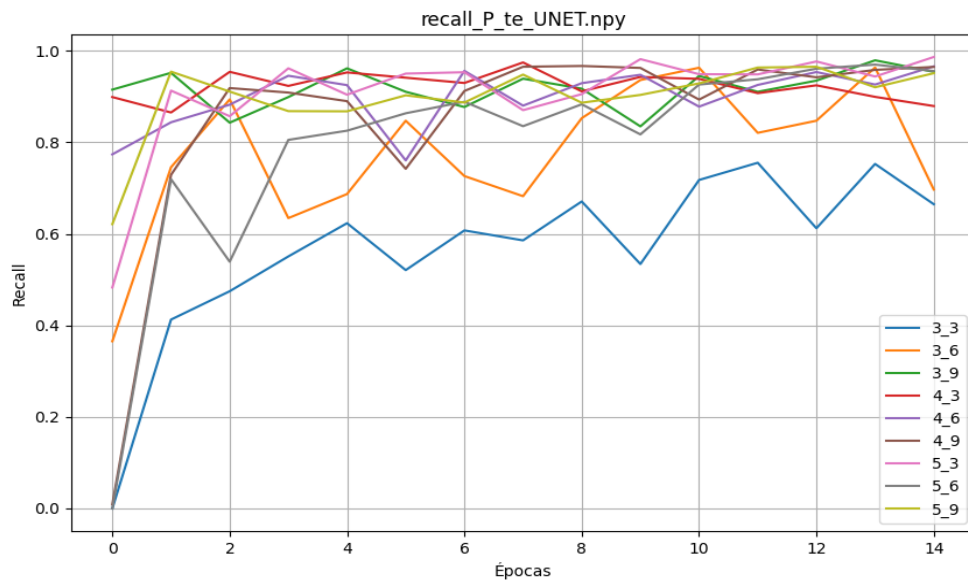


Figura 32: Recall de la onda P durante el entrenamiento con las bases de datos QT y Lobachevsky obtenida con los datos de entrenamiento para cada configuración.

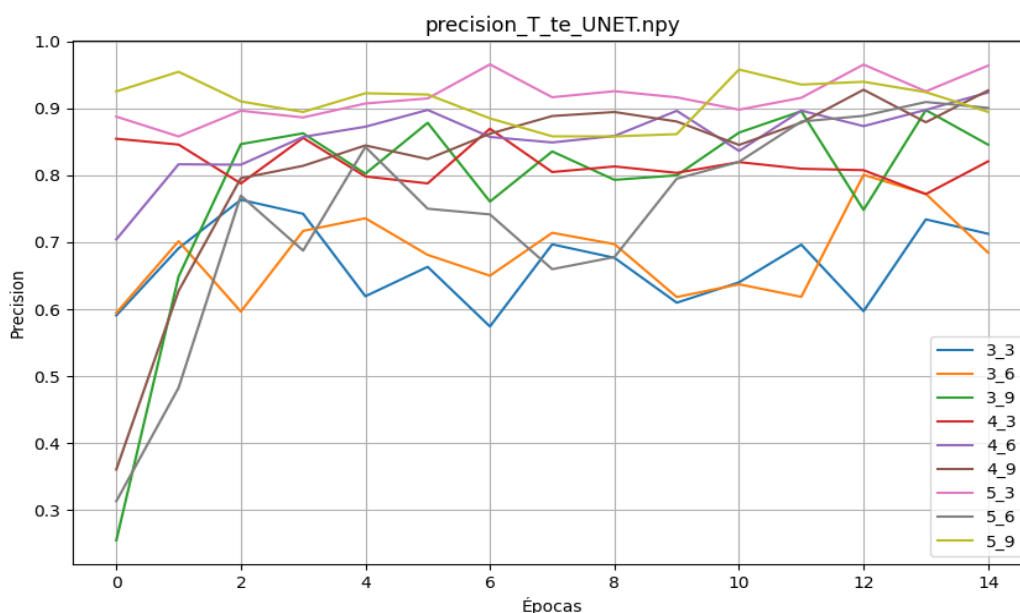


Figura 33: Precisión de la onda T durante el entrenamiento con las bases de datos QT y Lobachevsky obtenida con los datos de entrenamiento para cada configuración.

En la Figura 31 se evidencia de manera clara que la red neuronal con 5 niveles y 3 capas presenta una menor pérdida, lo cual indica un aprendizaje superior. Este resultado se confirma mediante las Figuras 32 y 33, en las cuales, en promedio, la mencionada red se posiciona por encima de la mayoría de las configuraciones evaluadas (el conjunto completo de gráficas se encuentra en el Anexo IV).

Adicionalmente, la red menos profunda, compuesta por 3 niveles y 3 capas, demuestra un desempeño de aprendizaje inferior, debido a que su simplicidad resulta inadecuada para la tarea de delineación en cuestión.

Se han evaluado dichas arquitecturas mediante la base de datos Lobachevsky obteniendo los siguientes resultados:

Prec	3_3	3_6	3_9	4_3	4_6	4_9	5_3	5_6	5_9
P	0.415	0.819	0.867	0.899	0.893	0.886	0.892	0.924	0.901
QRS	1	0.998	0.999	0.999	1	0.999	0.998	0.999	0.998
T	0.642	0.698	0.879	0.868	0.930	0.938	0.958	0.925	0.940

Tabla 3: Precisión obtenida con los datos de evaluación para las diferentes configuraciones de la red U-NET

Rec	3 3	3 6	3 9	4 3	4 6	4 9	5 3	5 6	5 9
P	0.668	0.754	0.961	0.878	0.949	0.939	0.964	0.922	0.936
QRS	0.952	0.960	0.981	0.959	0.977	0.976	0.983	0.973	0.991
T	0.852	0.948	0.885	0.878	0.946	0.995	0.987	0.953	0.994

Tabla 4: Recall obtenido con los datos de evaluación para las diferentes configuraciones de la red U-NET

F1	3 3	3 6	3 9	4 3	4 6	4 9	5 3	5 6	5 9
P	0.512	0.785	0.911	0.888	0.920	0.912	0.927	0.923	0.918
QRS	0.975	0.979	0.990	0.979	0.988	0.987	0.990	0.986	0.994
T	0.732	0.804	0.882	0.873	0.938	0.966	0.972	0.939	0.966

Tabla 5: Coeficiente F1 obtenido con los datos de evaluación para las diferentes configuraciones de la red U-NET

On	3 3	3 6	3 9	4 3	4 6	4 9	5 3	5 6	5 9
P	-10.2± 39.4	-18.8± 27.4	-3.9± 22.1	-9.6± 22.6	-2.5± 19.5	-1.9± 21.3	3.6± 18.5	3.4± 19.5	2.9± 24.7
QRS	-2.1± 8.2	0.2± 8.9	0.8± 8.9	-2.4± 6.6	0.0± 8.8	-2.4± 8.9	-1.7± 10.1	1.2± 10.4	-1.9± 9.4
T	7.5± 66.2	15.4± 71.6	11.2± 64.5	12.4± 76.4	10.1± 54.5	19.7± 66.4	18.0± 40.8	17.8± 41.1	27.8± 48.0

Tabla 6: Media y varianza del error del onset de las ondas en milisegundos, obtenidas con los datos de evaluación para las diferentes configuraciones de la red U-NET

Off	3 3	3 6	3 9	4 3	4 6	4 9	5 3	5 6	5 9
P	22.3± 32.8	13.9± 19.7	8.5± 19.9	13.9± 17.7	10.9± 18.0	11.9± 19.2	6.1± 15.9	13.1± 19.9	13.2± 22.4
QRS	2.0± 12.3	-2.6± 13.1	-0.2± 19.4	-0.5± 12.6	-1.4± 16.1	0.7± 16.5	-0.8± 12.5	-2.0± 17.4	3.3± 18.6
T	37.4± 47.3	6.4± 67.0	13.5± 54.0	14.7± 63.4	7.8± 50.5	17.2± 54.8	15.6± 36.1	19.9± 37.3	18.6± 42.4

Tabla 7: Media y varianza del error del offset de las ondas en milisegundos, obtenidas con los datos de evaluación para las diferentes configuraciones de la red U-NET

La evaluación de las arquitecturas mostrada en las tablas 3, 4, 5, 6 y 7 confirma la decisión previamente tomada, demostrando que, en promedio, la red de 5 niveles y 3 capas obtiene los resultados más favorables.

3.4.3 Función de coste personalizada

Debido a utilización de una función de coste modificada, es necesario saber si esta aporta una mejora con respecto a la función de coste crossentropy que provee PyTorch. Para ello, realizamos dos entrenamientos de la red U-Net de 5 niveles y 3 capas sin modificar ningún parámetro, pero cambiando la función de coste.

	Precisión	Recall	F1	Error onset (media±desv. est.) (ms)	Error offset (media±desv. est.) (ms)
P	0.859	0.940	0.897	4.4±18.6	9.76±19.1
QRS	0.986	0.992	0.989	-0.41±11.1	4.1±14.3
T	0.915	0.974	0.943	6.5±39.1	14.7±36.4

Tabla 8: Métricas obtenidas con los datos de entrenamiento para la red U-NET con 5 niveles y 3 capas con función de coste crossentropy.

	Precisión	Recall	F1	Error onset (media±desv. est.) (ms)	Error offset (media±desv. est.) (ms)
P	0.892	0.964	0.927	3.6± 18.5	6.1± 15.9
QRS	0.998	0.983	0.990	-1.7± 10.1	-0.8± 12.5
T	0.958	0.987	0.972	18.0± 40.8	15.6± 36.1

Tabla 9: Métricas obtenidas con los datos de entrenamiento para la red U-NET con 5 niveles y 3 capas con función de coste personalizada.

El uso de la función de coste personalizada muestra una mejora sistemática con respecto a usar la función crossentropy, como se muestra en las tablas 8 y 9.

4. EVALUACIÓN

4.1 Método de evaluación con ventanas

La red U-NET definida procesa los datos de entrada en ventanas de 1024 muestras de longitud con solapamiento (stride = 128). Este procedimiento requiere el establecimiento de un método capaz de serializar los resultados de una grabación completa.

Para abordar esta cuestión, se adoptó un enfoque en el que todas las máscaras binarias con el onset correspondiente son sumadas. En este método, la salida correspondiente a cada muestra de la señal resultante es la suma de las máscaras binarias de hasta 8 ventanas solapadas. Posteriormente, se calcula el promedio de estas ventanas para cada muestra, obteniendo así un valor entre 0 y 1 (que puede interpretarse como la probabilidad de estar dentro de la máscara correspondiente a la onda), que luego es redondeado para asegurar que los valores sean binarios. Este método se esquematiza en la figura 34

Este método resulta efectivo para suavizar los datos de entrada y mitigar el efecto de los bordes de las ventanas, asegurando una evaluación más precisa y confiable de las grabaciones completas.

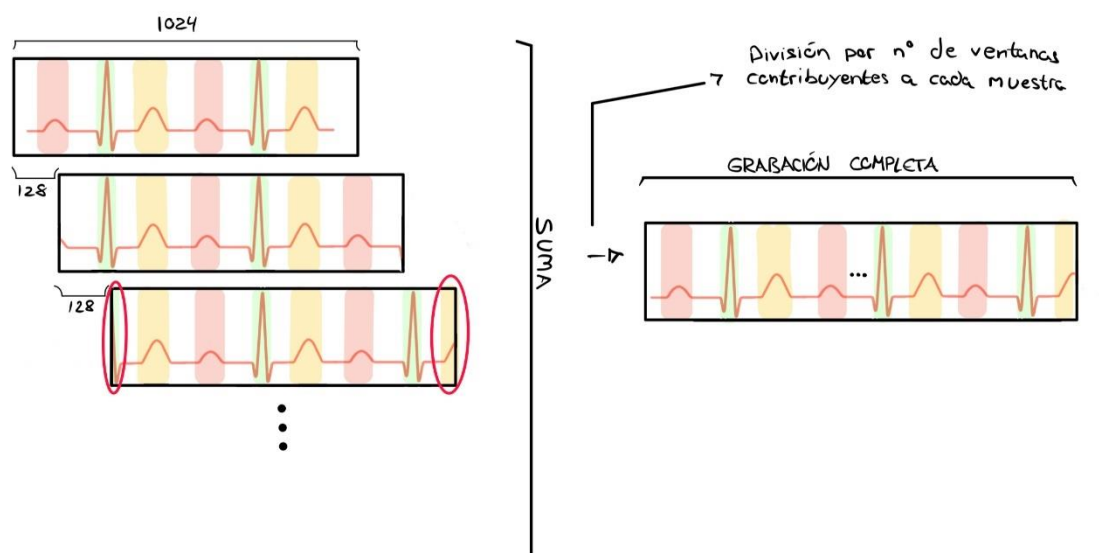


Figura 34: Proceso de serialización de los datos en evaluación.

4.2 Resultados

4.2.1 Red U-NET

Se inicia el proceso de entrenamiento utilizando datos sintéticos generados (preentrenamiento) con 30 epochs, stride = 1024, tamaño de ventana = 1024, $lr = 3 \cdot 10^{-4}$ y dropout = 0.25. La delineación de estos datos sintéticos se llevó a cabo mediante un delineador automático, lo cual implica que la capacidad máxima de la red para alcanzar métricas elevadas estará limitada por el desempeño de dicho delineador en esta base de datos.

A pesar de que estas marcas no han sido realizadas por cardiólogos, proporcionan a la red un aprendizaje básico que se complementará posteriormente con las bases de datos Lobachevsky y QT.

Dado el considerable tamaño de los datos sintéticos, se llevó a cabo el preentrenamiento con un 12 % de los datos sintéticos aproximadamente, elegido para que exista variedad de ritmos, ruidos, afecciones, entre otros; y el total de las marcas del delineador automático de la base de datos QT.

El entrenamiento con datos reales se realizó posteriormente con la combinación de datos de la base de datos Lobachevsky y QT con 15 epochs y stride = 128, obteniendo los resultados mostrados en la tabla 10.

	Precisión	Recall	F1	Error onset (media±desv. est.) (ms)	Error offset (media±desv. est.) (ms)
P	0.919	0.945	0.932	0.7 ± 18.2	14.7 ± 19.0
QRS	0.999	0.996	0.997	-1.2 ± 7.7	-0.1 ± 15.3
T	0.967	0.964	0.965	8.3 ± 36.5	14.2 ± 35.0

Tabla 10: Métricas obtenidas con los datos de preentrenamiento y entrenamiento para la red U-NET con 5 niveles y 3 capas.

En la tabla 10 se presentan las métricas representativas del desempeño de la red neuronal analizada. No obstante, en el presente estudio se ha desarrollado una función de postprocesado multiderivacional capaz de mejorar la delineación considerando las distintas derivaciones de las grabaciones analizadas. La explicación detallada del funcionamiento de dicha función se expone en el Anexo V.

	Precisión	Recall	F1	Error onset (media±desv. est.) (ms)	Error offset (media±desv. est.) (ms)
P	0.936	0.952	0.944	0.6±18.6	14.5±18.6
QRS	1	1	1	-1.1±7.5	-0.1±14.8
T	0.999	0.958	0.978	7.0±12.3	12.3±32.1

Tabla 11: Métricas obtenidas con los datos preentrenamiento y entrenamiento para la red U-NET con 5 niveles y 3 capas aplicando postprocesado multiderivacional.

Se observa una mejora notable en la mayoría de las métricas, destacándose la detección del 100% de los complejos QRS, así como una mayor dificultad en la detección de la onda P.

4.2.2 Adaptaciones y Mejoras de la Red U-NET

La U-NET que se ha descrito y usado hasta este punto del trabajo no añade ninguna modificación con respecto a la original [13]. Por ello se han realizado varios entrenamientos con modificaciones de la arquitectura, únicamente con datos de entrenamiento, suficiente para comparar modelos.

- 1) Conexiones residuales. Se ha modificado la red U-NET añadiendo conexiones residuales en los distintos niveles. Estas permiten una mejor propagación de los gradientes, sobre todo en redes neuronales profundas (una explicación más extensa se encuentra en el Anexo IV).
- 2) Capas recurrentes LSTM. Otra de las modificaciones ha sido la inclusión de varias capas LSTM. Las redes LSTM han demostrado ser muy efectivas en la tarea de delineación ECG [1] [16]. Por ello se han introducido dos capas LSTM bidireccionales entre el encoder y el decoder, lo que permite capturar relaciones temporales de la señal, muy útiles en la tarea de delineación (la descripción de las capas LST se encuentra en el Anexo IV).
- 3) Capas de atención. Finalmente, se han introducidos mecanismos de atención, los cuales ayudan al modelo a enfocarse en las partes más relevantes del ECG (ver Anexo IV).

Estas pruebas se han realizado con 15 epochs y un learning rate de 0.0003, únicamente con los datos de entrenamiento, cuyos resultados se representan en las tablas 12, 13, 14, 15 y 16.

Prec	U-NET	U-NET RESIDUAL	U-NET + LSTM	U-NET + ATENCIÓN
P	0.892	0.900	0.928	0.942
QRS	0.998	0.960	0.999	0.993
T	0.958	0.997	0.998	0.992

Tabla 12: Precisión obtenida con los datos de evaluación para diferentes modificaciones de la red U-NET.

Rec	U-NET	U-NET RESIDUAL	U-NET + LSTM	U-NET + ATENCIÓN
P	0.964	0.981	0.975	0.956
QRS	0.983	0.998	0.990	0.998
T	0.987	0.990	0.995	0.993

Tabla 13: Recall obtenido con los datos de evaluación para diferentes modificaciones de la red U-NET.

F1	U-NET	U-NET RESIDUAL	U-NET + LSTM	U-NET + ATENCIÓN
P	0.927	0.939	0.951	0.949
QRS	0.990	0.979	0.994	0.995
T	0.972	0.994	0.996	0.992

Tabla 14: Coeficiente F1 obtenido con los datos de evaluación para diferentes modificaciones de la red U-NET.

On	U-NET	U-NET RESIDUAL	U-NET + LSTM	U-NET + ATENCIÓN
P	3.6± 18.6	3.1±21.3	0.8 ±19.9	-3.3±20.4
QRS	-1.7±10.1	3.3±20.2	-0.8 ±9.1	-3.4± 7.7
T	18.0±40.9	5.1 ±38.7	14.9± 35.8	9.9±37.8

Tabla 15: Media y varianza del error del onset de las ondas en milisegundos, obtenidas con datos de evaluación para diferentes modificaciones de la red U-NET.

Off	U-NET	U-NET RESIDUAL	U-NET + LSTM	U-NET + ATENCIÓN
P	6.1 ± 15.9	7.2±17.6	7.9±16.9	14.2±17.4
QRS	-0.8±12.5	-7.7±28.4	-0.3 ±12.8	-1.2± 11.5
T	15.6±36.1	11.3± 32.8	11.2 ±33.7	18.6±36.1

Tablas 16: Media y varianza del error del offset de las ondas en milisegundos, obtenidas con datos de evaluación para diferentes modificaciones de la red U-NET.

En términos generales, las arquitecturas de red U-NET modificadas exhiben un rendimiento superior en la tarea de delineación en comparación con la red U-NET básica. Aunque no es sistemática, la incorporación de un bloque LSTM en la red U-NET constituye la mejora más significativa.

Si se aplica el bloque de postprocesado multiderivacional a la arquitectura U-NET con bloque LSTM se obtienen los resultados mostrados en la tabla 17.

	Precisión	Recall	F1	Error onset (media±desv. est.) (ms)	Error offset (media±desv. est.) (ms)
P	0.915	0.979	0.946	1.5±16.4	7.7±16.2
QRS	1	1	1	-0.7±7.7	-0.1±12.9
T	1	0.993	0.996	13.9±31.1	10.2±31.5

Tablas 17: Métricas obtenidas aplicando postprocesado para red U-NET con bloque bidireccional LSTM

4.2.3 Comparación de resultados

Los mejores resultados se han obtenido con el modelo U-NET con bloque bidireccional LSTM. Estos resultados se comparan con algunos ejemplos de la literatura actual en las tablas 18, 19, 20 y 21, para conocer la relevancia de este trabajo.

Precisión	Este trabajo	Martínez JP (Procesado de señal) [24]	Jimenez-Perez G (U-NET) [14]	Iana Sereda (Convolucional) [31]
P	91.59	91.03	90.12	90.48
QRS	100	99.86	99.14	98.27
T	100	97.79	98.25	96.23

Tabla 18: Comparación de la precisión de la mejor prueba en este trabajo con las de otros trabajos

Recall	Este trabajo	Martínez JP (Procesado de señal) [24]	Jimenez-Perez G (U-NET) [14]	Iana Sereda (Convolucional) [31]
P	97.95	98.87	98.73	97.36
QRS	100	99.80	99.94	99.86
T	99.36	99.77	99.88	93.51

Tablas 19: Comparación del recall de la mejor prueba en este trabajo con las de otros trabajos

Onset	Este trabajo	Martínez JP (Procesado de señal) [24]	Jimenez-Perez G (U-NET) [14]	Iana Sereda (Convolucional) [31]
P	1.4±16.4	2± 14.8	1.5±22.9	3.4±18.4
QRS	-0.7±7.7	4.6±7.7	-0.1±8.4	1.7±10
T	13.8±31.1	N/A	21.5±66.3	9.2±28.2

Tablas 20: Comparación de la media y varianza del error del onset de las ondas en milisegundos de la mejor prueba en este trabajo con las de otros trabajos

Offset	Este trabajo	Martínez JP (Procesado de señal) [24]	Jimenez-Perez G (U-NET) [14]	Iana Sereda (Convolutacional) [31]
P	7.7±16.2	1.9± 12.8	0.3 ±16	-4.1±19.4
QRS	-0.1 ±12.9	0.8± 8.7	3.6±12.5	-3.4±12.3
T	10.2±31.5	-1.6 ± 18.1	4.5±31.1	-6±25

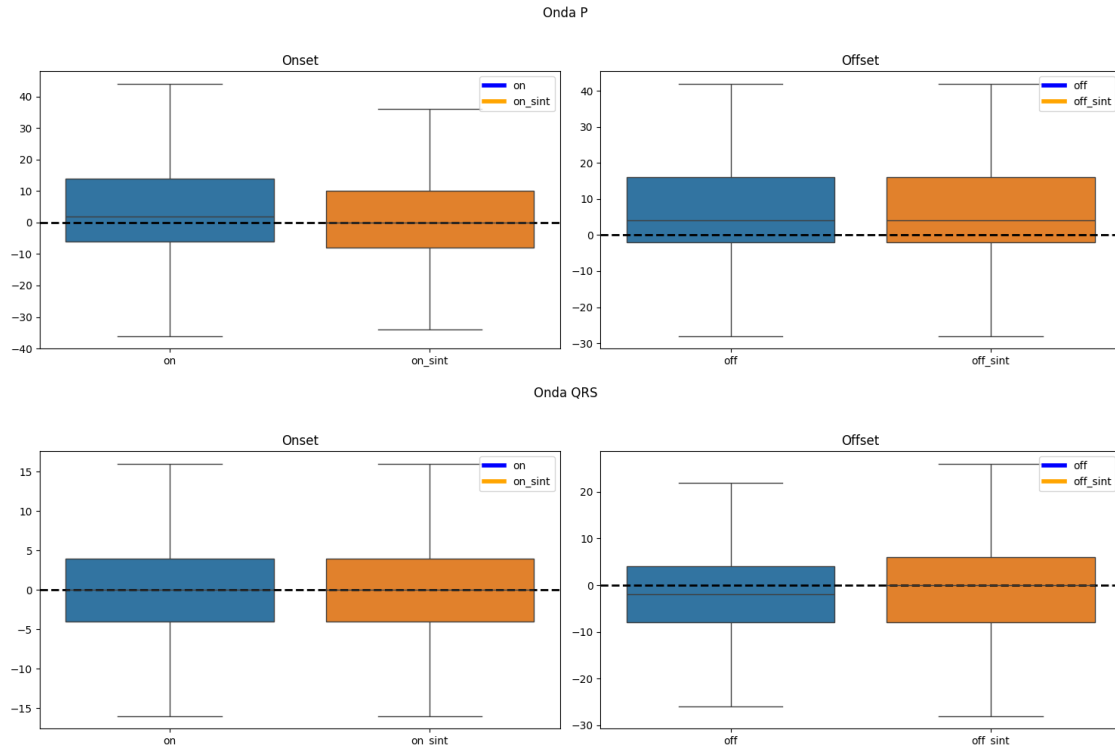
Tablas 21: Comparación de la media y varianza del error del offset de las ondas en milisegundos de la mejor prueba en este trabajo con las de otros trabajos

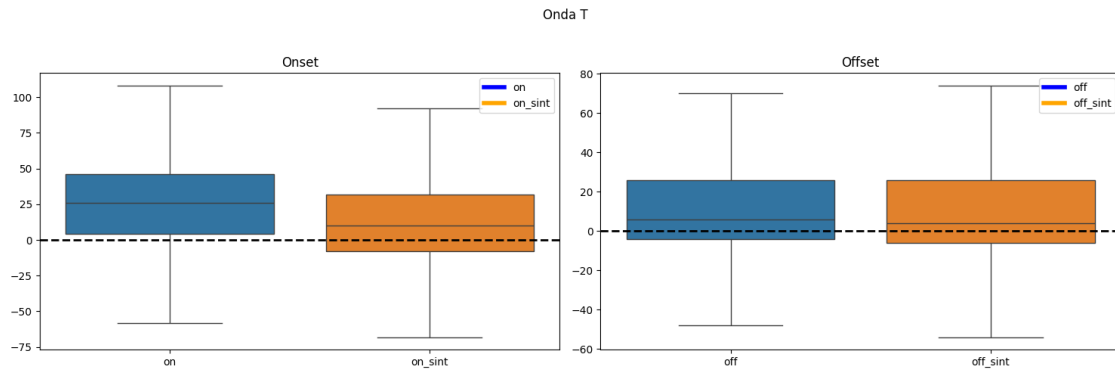
5. CONCLUSIONES Y PROYECCIONES

5.1 Conclusiones

En este trabajo se ha estudiado y evaluado las prestaciones de la red U-NET para la delineación de electrocardiogramas. El principal objetivo del trabajo era crear un delineador que fuese capaz de delinear una gran variedad de afecciones cardiacas, frecuencias cardiacas y ruido sin necesidad de la intervención humana.

Uno de los principales problemas en la implementación de algoritmos de aprendizaje profundo en el entorno clínico es la dependencia de estos con el tamaño de entrenamiento de los datos, lo cual puede ser costoso de adquirir. Por ello en este trabajo se ha utilizado el aumento de datos mediante adicción de ruido y generación de electrocardiogramas sintéticos, paliando parcialmente dicho problema.





Figuras 35, 36, 37: Gráficas de caja de las métricas onset y offset obtenidas con los datos de evaluación para las diferentes ondas.

Tal como se evidencia en las figuras 35, 36 y 37, se aprecia una mejora notable al implementar un preentrenamiento con datos sintéticos acompañados de ruido. En particular, se observa que la mediana se ubica habitualmente más próxima al valor cero. Asimismo, tanto el primer cuartil como el tercer cuartil se sitúan más cercanos a la mediana, indicando una reducción en la variabilidad. Por último, se constata una disminución significativa en la extensión de los bigotes, observable en casos específicos como el inicio del complejo QRS y el final de la onda P.

A pesar de lograr obtener datos sintéticos y añadir ruido, la red sigue estando entrenada por bases de datos muy limitadas, las cuales no tienen toda la variedad de ejemplos de afecciones cardiacas deseados. Aun así, consigue obtener una buena generalización, consiguiendo así lograr uno de los principales objetivos de este trabajo, obteniendo buenos resultados en una mayor variedad de rangos que los algoritmos DSP como [24].

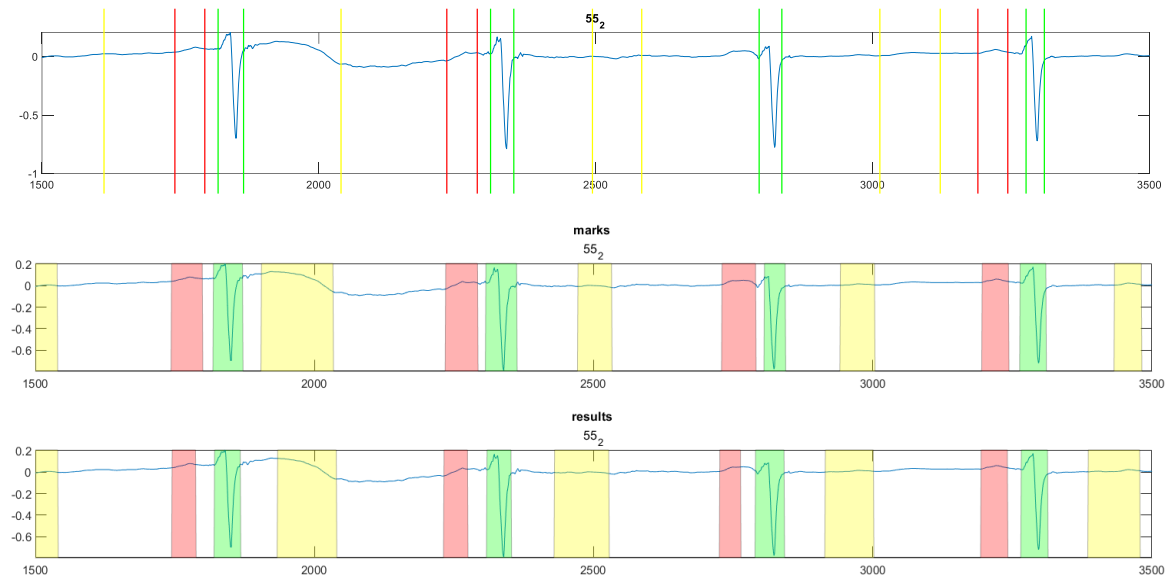


Figura 38: Comparación de la grabación 55 delineada con [24] (arriba) y con la red neuronal de este trabajo (medio), junto con las anotaciones de referencia (abajo).

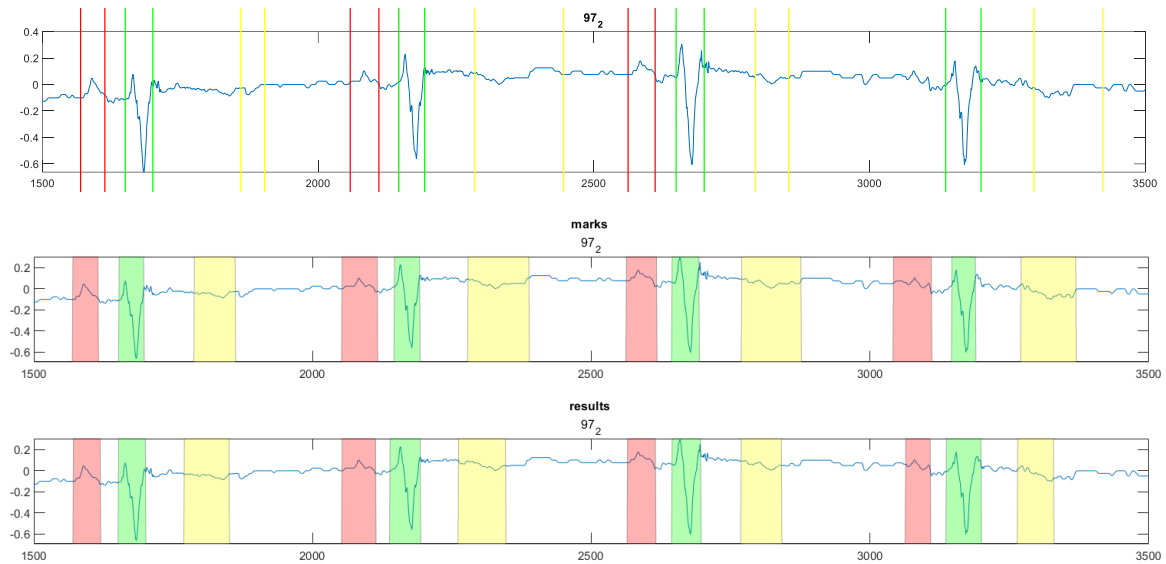


Figura 39: Comparación de la grabación 97 delineada con [24] (arriba) y con la red neuronal de este trabajo (medio), junto con la solución (abajo).

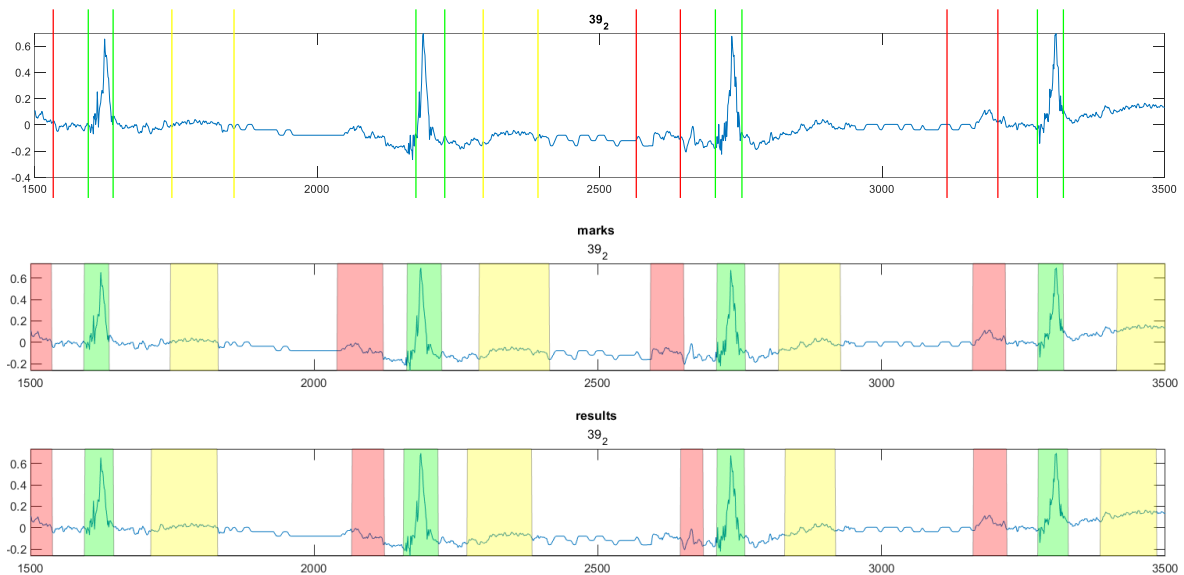


Figura 40: Comparación de la grabación 39 delineada con [24] (arriba) y con la red neuronal de este trabajo (medio), junto con la solución (abajo).

Las figuras 38, 39 y 40 presentan alguno de los ejemplos donde la red de este trabajo proporciona un mejor funcionamiento que en [24]. Se observa en la figura 38 que, al ser un algoritmo DSP, no es capaz de detectar la onda T debido a cambios en esta, que pueden ser indicativos de condiciones cardíacas isquémicas.

En la figura 40 aparecen ondas irregulares, pudiendo ser producidas por un mal contacto en el electrodo, por la presencia de interferencias o por una alteración en la conducción eléctrica del corazón. La onda P y onda T no están delimitadas correctamente por el algoritmo DSP, ondas clave para detectar patologías y arritmias como la fibrilación auricular. Asimismo, en la figura 39 no se realiza una delineación correcta de la onda T por el algoritmo DSP.

La mejor ejecución de este estudio presenta una mejora en comparación con [24] en términos de precisión de la onda P en un 0.62%, del complejo QRS en un 0.14% y de la onda T en un 2.21%. En cuanto al recall del complejo QRS, se observa una mejora del 0.21%. Además, la media del error de inicio de la onda P mejora en un 30%, la del complejo QRS en un 84.8%, y la media del error de fin del complejo QRS en un 93.75%. Las demás métricas son similares en ambos estudios, con la excepción de la media y la varianza del error de offset de la onda T, donde el algoritmo basado en la transformada Wavelet muestra un desempeño significativamente superior.

A pesar del buen funcionamiento desempeñado en la mayoría de ejemplos, existen casos concretos los cuales la red no ha sido capaz de delinear correctamente. Se muestran algunos ejemplos de los errores existentes con la base de datos Lobachevsky.

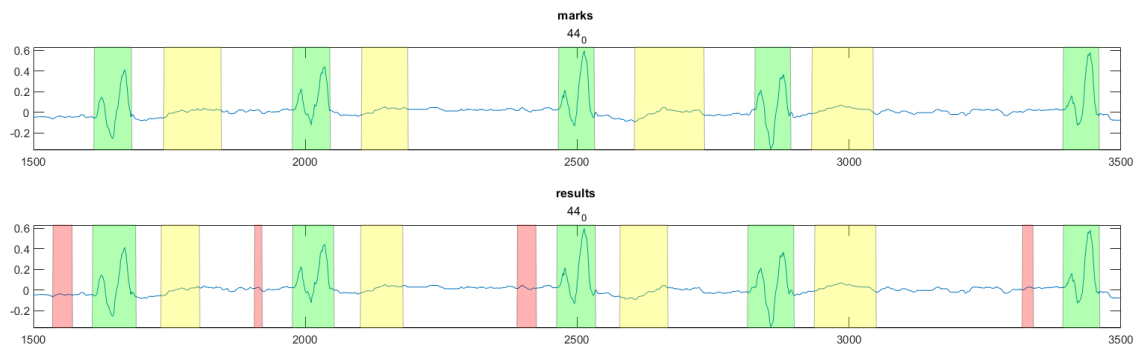


Figura 41: Primera derivación de la grabación 44 presentada con las marcas hechas por cardiólogos (arriba) y las obtenidas con la red neuronal de este trabajo.

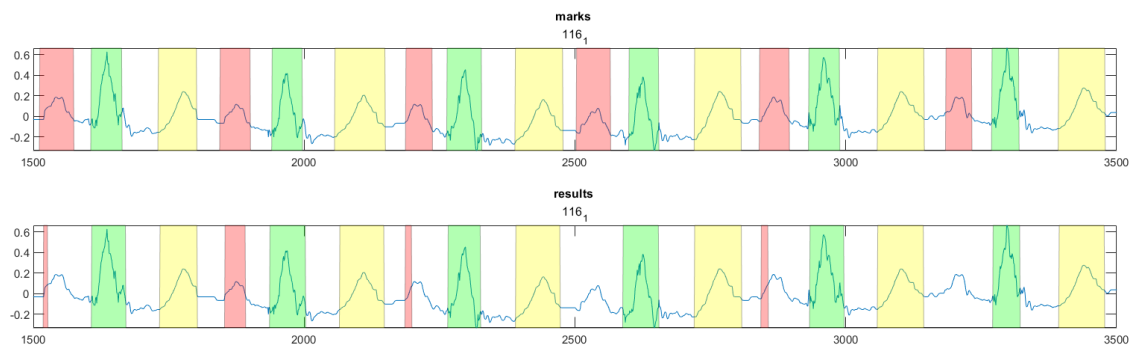


Figura 42: Segunda derivación de la grabación 116 presentada con las marcas hechas por cardiólogos (arriba) y las obtenidas con la red neuronal de este trabajo.

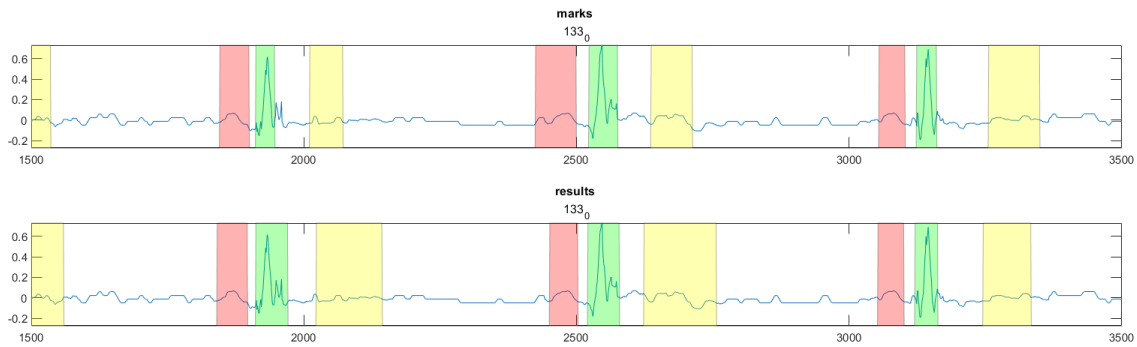


Figura 43: Primera derivación de la grabación 133 presentada con las marcas hechas por cardiólogos (arriba) y las obtenidas con la red neuronal de este trabajo.

En la figura 41 se presenta un caso en el que se identifica la onda P, a pesar de su ausencia en las marcas. No queda del todo claro si este ejemplo está debidamente delineado, dado que parece observarse una cierta elevación asociada a la onda P detectada por la red neuronal, lo cual podría comprometer la validez de las soluciones. No obstante, se asume que es correcto. Asimismo, a pesar de ser detectada, la onda P en la figura 42 no está adecuadamente delineada, lo que confirma que esta onda presenta la mayor deficiencia en su delineación, como ya se había evidenciado en las métricas previamente analizadas.

Adicionalmente, en la figura 43 se observa que el inicio (onset) y el final (offset) de la onda T presentan una variabilidad significativa. En algunas grabaciones, se tiende a sobreestimar, es decir, a delinear la onda más allá de su extensión real, mientras que, en otros casos, aunque menos frecuentes, se subestima su anchura.

Existen otros ejemplos en los que la delineación no resulta precisa; sin embargo, tales discrepancias se presentan únicamente en una o muy pocas derivaciones, lo que no afecta significativamente el resultado final si se trata de una situación clínica, debido a que se presenta como una combinación de todas las derivaciones. Para abordar esto, se puede emplear la delineación multilead con reglas, detallado en el Anexo V, que permite obtener una solución única considerando todas las derivaciones, sin que un error en una sola afecte el resultado final.

En general, las métricas demuestran un comportamiento comparable a la mayoría de los trabajos de la literatura hasta el conocimiento actual, observando en las tablas 18, 19, 20 y 21 mejora con respecto a otros trabajos en la media en el onset de la onda P y desviación estándar en el complejo QRS así como en la precisión de todas las ondas y el recall del complejo QRS, demostrando que la red estudiada es una red robusta, útil para la delineación.

El incremento en el rendimiento observado en ciertas métricas puede atribuirse a la innovadora combinación entre la red U-NET y un bloque LSTM bidireccional, lo cual abarca tanto la dimensión temporal como la espacial, una aproximación no documentada en la literatura existente hasta el conocimiento actual; así como la aplicación de una máscara personalizada. Además, la integración de datos sintéticos generados por un sistema validado por cardiólogos y la incorporación de ruido a las grabaciones preexistentes contribuyen a una mayor generalización en los electrocardiogramas posiblemente analizados en entornos hospitalarios, considerando interferencias comunes como marcapasos, actividad muscular y artefactos debidos al movimiento de los electrodos, entre otros.

Finalmente, es importante resaltar que este trabajo establece una base valiosa para otros investigadores que continúen esta línea de investigación, permitiendo un ahorro significativo de tiempo en la extracción, lectura y adaptación de datos, así como una función de coste personalizada para las bases de datos analizadas que aporta un incremento de prestaciones durante el entrenamiento. El sistema desarrollado cuenta con parámetros de fácil modificación, facilitando así la experimentación con una variedad de redes, hiperparámetros, o incluso su aplicación en otros tipos de aprendizaje que no necesariamente se enfoquen en la delineación.

El código implementado se encuentra disponible en la siguiente dirección: [Rub3bn/Electrocardiogram-delineator-based-on-deep-neural-networks-U-NET-pytorch-\(github.com\)](https://github.com/Rub3bn/Electrocardiogram-delineator-based-on-deep-neural-networks-U-NET-pytorch)

5.2 Líneas futuras

Durante la realización de este trabajo se han considerado diversas tareas que, finalmente, no han sido implementadas debido a la naturaleza del Trabajo de Fin de Grado (TFG) y a las limitaciones temporales inherentes a su desarrollo. Entre estas tareas, se incluye la modificación de ciertos hiperparámetros, la cual no ha sido realizada debido a la extensa cantidad de combinaciones posibles que dichos parámetros presentan. Por ejemplo, la modificación del valor de dropout, el stride o el tamaño de la ventana fueron seleccionados basándose en proyectos similares, sin realizar un ajuste exhaustivo.

De igual forma, se podrían probar arquitecturas diferentes a la U-NET, como son los transformers, modelos de espacio de estado estructurado entre otras, las cuales están demostrando un gran desempeño en multitud de tareas [28] [36].

En segundo lugar, se evaluó la posibilidad de implementar una red multiderivacional que permitiera la entrada de las 12 derivaciones comúnmente registradas en un electrocardiograma, generando como salida una única derivación. Este enfoque aprovecharía la información de todas las derivaciones para mejorar los resultados, tal como se ha demostrado en el postprocesado realizado en este estudio. Además, este enfoque proporcionaría una única solución, lo cual es ideal en un entorno clínico.

Adicionalmente, cabe señalar que la variedad de morfologías empleadas en la generación de electrocardiogramas sintéticos no es muy amplia y fueron delineadas mediante un algoritmo DSP [24], lo que limita los resultados a las métricas proporcionadas por dicho delineado, que no puede considerarse un gold standar. Una opción alternativa sería modificar el generador para que ofreciese a su salida los inicios y finales de las ondas utilizadas para generar los electrocardiogramas, permitiendo una mayor precisión que la obtenida con el delineador aplicado.

Finalmente, es importante destacar que la finalidad de este tipo de trabajos es su aplicación en situaciones clínicas reales. Por lo tanto, uno de los aspectos cruciales a considerar en investigaciones futuras es la delineación de electrocardiogramas en tiempo real. Esto se podría lograr optimizando los modelos de redes neuronales empleados, minimizando su complejidad sin comprometer sus prestaciones.

Bibliografía

- [1] H. Abrishami *et al*, "Supervised ECG Interval Segmentation Using LSTM Neural Network," *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)*, pp. 71-77, 2018. Available: <https://www.proquest.com/conference-papers-proceedings/supervised-ecg-interval-segmentation-using-lstm/docview/2139457108/se-2>.
- [2] J. Camps, B. Rodriguez, and A. Mincholé, "Deep learning-based QRS multilead delineator in electrocardiogram signals," in *CinC*, vol. 45. IEEE, 2018, pp. 1–4.
- [3] P. Sodmann, M. Vollmer, N. Nath, and L. Kaderali, "A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms," *Physiol. Meas.*, vol. 39, no. 10, p. 104005, 2018.
- [4] G. H. Tison, J. Zhang, F. N. Delling, and R. C. Deo, "Automated and interpretable patient ECG profiles for disease detection, tracking, and discovery," *Circ. Cardiovasc. Qual.*, vol. 12, no. 9, p. e005289, 2019.
- [5] Becker DE. Fundamentals of electrocardiography interpretation. *Anesth Prog.* 2006 Summer;53(2):53-63; quiz 64. doi: 10.2344/0003-3006(2006)53[53:FOEI]2.0.CO;2. PMID: 16863387; PMCID: PMC1614214.
- [6] J. Park, J. An, J. Kim, S. Jung, Y. Gil, Y. Jang, K. Lee, and I. Oh, "Study on the use of standard 12-lead ECG data for rhythm-type ECG classification problems," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106521, 2021. doi: 10.1016/j.cmpb.2021.106521.
- [7] Meek S, Morris F. ABC of clinical electrocardiography. Introduction. I-Leads, rate, rhythm, and cardiac axis. *BMJ.* 2002 Feb 16;324(7334):415-8. doi: 10.1136/bmj.324.7334.415. PMID: 11850377; PMCID: PMC1122339.
- [8] J. Hampton, *The ECG Made Easy, 9th Edition*, Elsevier, 2019. [Libro electrónico]. Disponible en: <https://pre-med.jumedicine.com/wp-content/uploads/sites/7/2020/10/The-ECG-Made-Easy-9th-Edition.pdf>. [Accessed: Jun. 1, 2024]
- [9] J. Pérez y M. López, "Capítulo III: Análisis de sistemas," *Proyecto Fin de Carrera*, Universidad de Sevilla, Sevilla, España, PFC 11818, 2023. [En línea]. Disponible: <https://biblus.us.es/bibing/proyectos/abreproy/11818/fichero/Capitulo+III.pdf> [Accessed: Jun. 5, 2024].
- [10] S. Naeem, A. Ali, S. Anam, and M. Ahmed, "An unsupervised machine learning algorithms: Comprehensive review," *IJCDS Journal*, vol. 13, pp. 911-921, 2023. doi: 10.12785/ijcds/130172.
- [11] "Reinforcement Learning," *MathWorks*. [Online]. Available: <https://es.mathworks.com/discovery/reinforcement-learning.html>. [Accessed: Jun. 6, 2024].
- [12] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L. and Haworth, A. (2021), A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol*, 65: 545-563. <https://doi.org/10.1111/1754-9485.13261>

- [13] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
- [14] Jimenez-Perez G, Alcaine A, Camara O. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Sci Rep*. 2021 Jan 13;11(1):863. doi: 10.1038/s41598-020-79512-7. PMID: 33441632; PMCID: PMC7806759.
- [15] Ji Jimenez-Perez G, Acosta J, Alcaine A and Camara O (2024) Generalising electrocardiogram detection and delineation: training convolutional neural networks with synthetic data augmentation. *Front. Cardiovasc. Med*. **11**: 1341786. doi: 10.3389/fcvm.2024.1341786
- [16] Darmawahyuni A, Nurmaini S, Rachmatullah MN, Avi PP, Teguh SBP, Sapitri AI, Tutuko B, Firdaus F. Improved delineation model of a standard 12-lead electrocardiogram based on a deep learning algorithm. *BMC Med Inform Decis Mak*. 2023 Jul 28;23(1):139. doi: 10.1186/s12911-023-02233-0. PMID: 37507698; PMCID: PMC10375607.
- [17] Che, C., Zhang, P., Zhu, M. et al. Constrained transformer network for ECG signal processing and arrhythmia classification. *BMC Med Inform Decis Mak* 21, 184 (2021). <https://doi.org/10.1186/s12911-021-01546-2>
- [18] Laguna P, Mark RG, Goldberger AL, Moody GB. A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. *Computers in Cardiology* 24:673-676 (1997).
- [19] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.
- [20] Kalyakulina, A., Yusipov, I., Moskalenko, V., Nikolskiy, A., Kosonogov, K., Zolotykh, N., & Ivanchenko, M. (2021). Lobachevsky University Electrocardiography Database (version 1.0.1). *PhysioNet*. <https://doi.org/10.13026/eegm-h675>.
- [21] Kalyakulina, A.I., Yusipov, I.I., Moskalenko, V.A., Nikolskiy, A.V., Kosonogov, K.A., Osipov, G.V., Zolotykh, N.Yu., Ivanchenko, M.V.: LUDB: A New Open-Access Validation Tool for Electrocardiogram Delineation Algorithms, *IEEE Access*, vol. 8, pp. 186181-186190, 2020, doi: 10.1109/ACCESS.2020.3029211
- [22] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* [Online]. 101 (23), pp. e215–e220.
- [23] L. Bachi et al., "ECG Modeling for Simulation of Arrhythmias in Time-Varying Conditions," in *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 12, pp. 3449-3460, Dec. 2023, doi: 10.1109/TBME.2023.3288701.
- [24] Martínez JP, Almeida R, Olmos S, Rocha AP, Laguna P. A wavelet-based ECG delineator: evaluation on standard databases. *IEEE Trans Biomed Eng*. 2004 Apr;51(4):570-81. doi: 10.1109/TBME.2003.821031. PMID: 15072211.

- [25] A. Peimankar and S. Puthusserypady, "DENS-ECG: A deep learning approach for ECG signal delineation," *Expert Systems with Applications*, vol. 165, p. 113911, 2021. [Online]. Available: <https://doi.org/10.1016/j.eswa.2020.113911>.
- [26] Tison GH, Zhang J, Delling FN, Deo RC. Automated and Interpretable Patient ECG Profiles for Disease Detection, Tracking, and Discovery. *Circ Cardiovasc Qual Outcomes*. 2019 Sep;12(9):e005289. doi: 10.1161/CIRCOUTCOMES.118.005289. Epub 2019 Sep 5. PMID: 31525078; PMCID: PMC6951431.
- [27] Sodmann P, Vollmer M, Nath N, Kaderali L. A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms. *Physiol Meas*. 2018 Oct 24;39(10):104005. doi: 10.1088/1361-6579/aae304. PMID: 30235165.
- [28] T. Mehari and N. Strodthoff, "Advancing the State-of-the-Art for ECG Analysis through Structured State Space Models," *arXiv preprint arXiv:2211.07579*, 2022. doi: 10.48550/arXiv.2211.07579.
- [29] P. Fränti and R. Mariescu-Istodor, "Soft precision and recall," *Pattern Recognition Letters*, vol. 167, pp. 115-121, Mar. 2023. doi: [10.1016/j.patrec.2023.02.005](https://doi.org/10.1016/j.patrec.2023.02.005).
- [30] R. R. Shamir, Y. Duchin, J. Kim, G. Sapiro, and N. Y. Harel, "Continuous Dice Coefficient: a Method for Evaluating Probabilistic Segmentations," *bioRxiv*, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:90993940>.
- [31] Iana Sereda, Sergey Alekseev, Aleksandra Koneva, Roman Kataev, and Grigory Osipov. ECG segmentation by neural networks: errors and correction. In 2019 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2019.
- [32] J. M. Bote, J. Recas, F. Rincón, D. Atienza and R. Hermida, "A Modular Low-Complexity ECG Delineation Algorithm for Real-Time Embedded Systems," in *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 2, pp. 429-441, March 2018, doi: 10.1109/JBHI.2017.2671443.
- [33] Jimenez-Perez G, Alcaine A, Camara O. Delineation of the electrocardiogram with a mixed-quality-annotations dataset using convolutional neural networks. *Sci Rep*. 2021 Jan 13;11(1):863. doi: 10.1038/s41598-020-79512-7. PMID: 33441632; PMCID: PMC7806759.
- [34] Sociedad Española de Cardiología, "Coincidiendo con la pandemia, la mortalidad cardiovascular vuelve a crecer," *Sociedad Española de Cardiología*, 25 de noviembre de 2021. [Online]. Available: <https://secardiologia.es/comunicacion/notas-de-prensa/notas-de-prensa-sec/13104-coincidiendo-con-la-pandemia-la-mortalidad-cardiovascular-vuelve-a-crecer>. [Accedido: 21-ago-2024].
- [35] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998. doi: 10.1109/5.726791.
- [36] Che, C., Zhang, P., Zhu, M. *et al.* Constrained transformer network for ECG signal processing and arrhythmia classification. *BMC Med Inform Decis Mak* **21**, 184 (2021). <https://doi.org/10.1186/s12911-021-01546-2>
- [37] D. M. Bers, "The Cardiac Myocyte as a Model of Cardiac Electrophysiology and Calcium Signaling," *Journal of Molecular and Cellular Cardiology*, vol. 42, no. 6, pp. 921-933, 2007. <https://www.journalwebsite.com/article/12345>. [Accessed: 2-Sep-2024].