

Análisis de señales ECG y detección de arritmias - análisis espectral

Juan Esteban Pineda Lopera*, Marly Alejandra Ortega Andrade†, Elisabeth Gómez Cano‡

*†‡Universidad de Antioquia, Medellín, Colombia

*jesteban.pineda1@udea.edu.co, †alejandra.ortega1@udea.edu.co, ‡elisabeth.gomez@udea.edu.co

I. RESUMEN

Este artículo presenta el análisis y clasificación de señales electrocardiográficas utilizando técnicas de aprendizaje automático y redes neuronales. Se procesaron registros del ECG para identificar cuatro patologías cardíacas: fibrilación auricular (AFIB), bradicardia sinusal (SB), taquicardia sinusal (ST) y ritmo sinusal (SR). Se exploraron distintas arquitecturas de redes neuronales, así como el algoritmo XGBoost, evaluando su rendimiento mediante métricas de clasificación y matrices de confusión. Los resultados muestran una alta precisión en la clasificación, destacando el desempeño de XGBoost y redes profundas para tareas específicas, lo que valida su potencial para aplicaciones clínicas automatizadas en el diagnóstico cardíaco.

II. ABSTRACT

This article presents the analysis and classification of electrocardiographic (ECG) signals using machine learning techniques and neural networks. ECG recordings were processed to identify four cardiac conditions: atrial fibrillation (AFIB), sinus bradycardia (SB), sinus tachycardia (ST), and normal sinus rhythm (SR). Various neural network architectures were explored, as well as the XGBoost algorithm, with performance evaluated through classification metrics and confusion matrices. The results demonstrate high classification accuracy, highlighting the effectiveness of XGBoost and deep networks for specific tasks, thus validating their potential for automated clinical applications in cardiac diagnosis.

III. INTRODUCCIÓN

El análisis automatizado de señales electrocardiográficas (ECG) se ha convertido en una herramienta clave para apoyar el diagnóstico de enfermedades cardíacas, gracias a los avances recientes en inteligencia artificial. Entre las técnicas más utilizadas se encuentran las redes neuronales y los algoritmos de aprendizaje automático, que han mostrado gran potencial en la clasificación precisa de arritmias y otros trastornos del ritmo cardíaco. Particularmente, modelos como las redes multicapa y métodos de ensamble como XGBoost han sido ampliamente explorados en la literatura debido a su capacidad para aprender patrones complejos a partir de grandes volúmenes de datos [1]. Sin embargo, persisten desafíos relacionados con la interpretación clínica de los modelos, el manejo de datos desbalanceados y la necesidad de una validación robusta. En este contexto, el

presente estudio explora distintas estrategias para la clasificación automática de señales de ECG provenientes de pacientes con cuatro condiciones cardíacas: fibrilación auricular (AFIB), ritmo sinusal (SR), bradicardia sinusal (SB) y taquicardia sinusal (ST). Se desarrolló un flujo completo de procesamiento que incluye la limpieza y codificación de los datos, seguido por la aplicación de modelos de redes neuronales con distintas arquitecturas y del clasificador XGBoost. Además, se evalúa el desempeño de estas aproximaciones tanto en escenarios multiclase como binarios, para determinar su efectividad en tareas clínicas concretas.

IV. ANTECEDENTES

IV-A. Enfoque óptimo para la clasificación multietapa de arritmias [2]

Este estudio muestra las dificultades presentes en la clasificación de arritmias utilizando redes neuronales. Para llevar a cabo esta tarea, en trabajos previos se han empleado transformaciones secuenciales de los datos brutos del ECG como características de entrada. Sin embargo, este enfoque requiere grandes volúmenes de datos y un alto costo computacional para entrenar los modelos de forma adecuada.

Además, las características tradicionales como la magnitud de los picos, la duración y las distancias entre ellos no siempre proporcionan suficiente información para clasificar ciertos tipos de arritmias, especialmente aquellas con distorsiones o ausencia de componentes, como ocurre en los casos de fibrilación auricular (AFIB) y flutter auricular (AFL). Por otro lado, métodos como la transformada Wavelet o la transformada rápida de Fourier, aunque útiles para obtener características en el dominio de la frecuencia, tienden a perder información importante del dominio temporal.

Uno de los principales retos en la extracción de características es la alta variabilidad morfológica del ECG entre individuos, dada por factores como la edad, el género y la raza. Por esta razón, se aplicó una normalización de la amplitud utilizando el algoritmo de máximo-mínimo. Además, se incorporaron variables demográficas como lo son la edad y el género y métricas específicas del ECG, como la media y varianza de los intervalos RR.

Este artículo también propone un nuevo método interpretable de extracción de características, evaluado a través de 11 escenarios distintos, con conjuntos de datos que iban desde 11 hasta 39.830 variables. Se definieron varios grupos, de los cuales se detallan a continuación los principales:

Grupo 1: Incluye 11 características clínicas básicas, como la frecuencia ventricular y auricular, duración del complejo QRS, intervalo QT, ejes R y T, conteo de complejos QRS, y el inicio y final de la onda Q.

Grupo 2: Contiene 23 características morfológicas extraídas específicamente de la derivación II del ECG, como la media y varianza de los intervalos RR, y de la altura, anchura y prominencia del complejo QRS, picos no QRS y valles.

Grupo 6: Aplica un enfoque más avanzado basado en la distribución empírica de frecuencias de picos y valles detectados en la derivación II. Estos elementos se clasificaron en tres subconjuntos: QRS, no QRS y valles, y se midieron combinaciones entre ellos para calcular razones morfológicas (altura/tiempo, anchura/tiempo y prominencia/tiempo). Estos datos se organizaron en tablas de distribución con longitud uniforme. En total, el Grupo 6 incluye 900 características morfológicas, 600 relacionadas con diferencias de ubicación, y 1800 correspondientes a relaciones entre combinaciones, sumando 3300 características.

Los demás grupos de características evaluados en el estudio se derivan de los Grupos 1, 2 y 6, a través de combinaciones, expansiones o adaptaciones de estas variables base. Este enfoque permitió capturar de manera más precisa la variabilidad de las ondas P, Q, R, S, T y los segmentos intermedios, factores clave en la identificación de distintos ritmos cardíacos.

IV-B. Detección confiable de ondas P en señales ECG patológicas [3]

Del complejo PQRS, la onda más comunmente perdida es la P, esto debido a que es la de menor voltaje y no posee características temporales o frecuenciales exclusivas que la diferencie de otras ondas, lo que provoca que frecuentemente no sea detectada correctamente por los equipos, o que ciertas patologías provoquen su encubrimiento, deformación, perdida o solapamiento con otras ondas del complejo. Esto es especialmente limitante debido a que su correcta detección y clasificación es esencial en diversas patologías.

Ante esto, se proponen ciertas reglas de decisión que buscan optimizar el proceso de clasificación y encontrar la onda P en la posición correcta o, alternativamente, no buscarla en absoluto. Esto se logra, primero, al buscar por patologías que afectan la presencia de la onda P a través de otras características, por ejemplo la contracción ventricular prematura (PVC), que puede ser clasificada con el área bajo el complejo QRS, o la fibrilación auricular (AFIB), que se detecta con la entropía de Shannon (SH). En caso de poseer estas patologías, no se busca la onda P en absoluto.

Por otro lado otras patologías pueden afectar la relación de P con el QRS, como el bloque AV de segundo grado, por lo que se alteran los parámetros de búsqueda de acuerdo a la morfología patológica.

Finalmente, se verifica que en efecto la onda detectada sea P a partir de su posición y amplitud, descartando posibles confusiones respecto a la onda T anterior.

En resumen, este proceso considera distintos factores adicionales basados en las características fisiológicas de las

señales cardíacas, tanto patológicas como sanas, para hacer más robusta y confiable la detección y caracterización de la onda P.

V. METODOLOGÍA

Se definió un conjunto de características compuesto por atributos extraídos desde tres enfoques principales: análisis de variabilidad del ritmo cardíaco (HRV), análisis espectral mediante transformada de Fourier, y coeficientes cepstrales MFCC. Esta selección se basó en su capacidad para capturar información relevante del ritmo cardíaco, como se evidenció en proyectos previos.

V-A. Extracción de características

V-A1. *Características HRV en el dominio temporal:* Se procesaron registros ECG denoised, extrayendo la derivación II. Usando el detector de Hamilton, se identificaron los picos R y se calcularon los intervalos NN. A partir de estos, se obtuvieron métricas como:

- mean_nni
- sdnn
- rmssd
- pnni_50
- pnni_20
- Cantidad total de R-peaks por señal

Estas métricas capturan aspectos fundamentales de la variabilidad del ritmo cardíaco.

V-A2. *Características espectrales (FFT):* A cada señal se le aplicó la transformada de Fourier, dividiendo el espectro en cinco bandas de frecuencia:

- F1: 0–20 Hz
- F2: 20–60 Hz
- F3: 60–120 Hz
- F4: 120–180 Hz
- F5: 180–250 Hz

Por cada banda se extrajeron:

- Picos espectrales
- Promedios de amplitud
- Razones entre áreas y magnitudes:
 - area_ratio
 - magnitude_ratio

Estas características permiten identificar patrones de frecuencia relevantes para el análisis del ritmo cardíaco.

V-A3. *Características MFCC:* Se calcularon 13 coeficientes MFCC por señal, a partir de los cuales se extrajeron estadísticas globales como el promedio, la desviación estándar y la varianza. Estas características permiten representar variaciones finas en la forma de onda.

Finalmente, todas las características fueron integradas en un único DataFrame, unidas por el identificador del registro. Se añadieron además variables clínicas complementarias, como edad, género y medidas del ECG clínico, conformando un conjunto final de 49 características.

V-B. Análisis comparativo de arquitecturas neuronales para la clasificación de cuatro patologías cardíacas

Para la clasificación automática de cuatro patologías cardíacas (Fibrilación Auricular, Ritmo Sinusal, Bradicardia Sinusal y Taquicardia Sinusal), se desarrolló un modelo de aprendizaje supervisado basado en redes neuronales artificiales, utilizando las características extraídas anteriormente.

V-B1. Procedimiento:

- **Preprocesamiento de datos:** Se trabajó con un archivo .csv que contenía las características obtenidas por paciente. Se eliminaron columnas irrelevantes y valores nulos. Se codificaron variables categóricas y se aplicó normalización estándar a las características numéricas. La variable objetivo (Estado) se transformó mediante codificación one-hot. Los datos se dividieron en conjuntos de entrenamiento (80 %) y prueba (20 %) de manera estratificada.
- **Visualización exploratoria:** Se aplicó Análisis de Componentes Principales (PCA) para representar la distribución de las clases en un espacio 2D y explorar la separabilidad de las patologías.
- **Modelado:** Se implementaron tres arquitecturas distintas de redes neuronales:
 - Red neuronal multicapa básica (MLP).
 - MLP con capas profundas y normalización por lotes (Batch Normalization).
 - Red de memoria a corto-largo plazo (LSTM) simulando entrada secuencial.

Se usó la función de pérdida `categorical_crossentropy` y el optimizador Adam. Se aplicaron técnicas como Dropout para reducir el sobreajuste. Esta metodología permitió entrenar y validar modelos para la clasificación automática de señales ECG en función de la patología cardíaca correspondiente.

V-C. Análisis comparativo de arquitecturas neuronales para la clasificación de dos patologías cardíacas

Se desarrolló un sistema de clasificación binaria, enfocado exclusivamente en distinguir entre dos tipos de ritmos cardíacos: fibrilación auricular (AFIB) y ritmo sinusal (SR). El flujo metodológico seguido fue el siguiente:

V-C1. Procedimiento: Se extrajeron únicamente los registros correspondientes a las clases AFIB y SR del conjunto de datos original. Posteriormente, se codificaron las etiquetas utilizando LabelEncoder para facilitar el entrenamiento supervisado.

Los datos fueron divididos en conjuntos de entrenamiento y prueba, usando un 80 % para entrenamiento y 20 % para prueba, manteniendo el balance entre clases mediante estratificación (`stratify=y_dos`).

Se entrenaron tres modelos de redes neuronales artificiales con arquitectura secuencial, empleando la biblioteca Keras (TensorFlow). Todos los modelos utilizaron la función de activación ReLU en las capas ocultas y sigmoide en la capa

de salida, dado que se trata de una clasificación binaria. La función de pérdida usada fue binary crossentropy, y el optimizador fue Adam. Las arquitecturas fueron las siguientes:

- **Modelo 1 (Simple):** Una sola capa oculta con 32 neuronas.
- **Modelo 2 (Intermedio):** Dos capas ocultas, con 64 y 32 neuronas respectivamente.
- **Modelo 3 (Complejo):** Tres capas, incluyendo una capa de Dropout del 30 % para mitigar el sobreajuste. Las capas densas tienen 128 y 64 neuronas.

Todos los modelos fueron entrenados durante 20 épocas con un tamaño de batch de 32 muestras, y sin mostrar salida durante el proceso (`verbose=0`).

Se realizaron predicciones sobre el conjunto de prueba con cada modelo, y se generaron matrices de confusión para evaluar su desempeño clasificatorio. Estas matrices muestran la cantidad de aciertos y errores en la identificación de las clases AFIB y SR.

V-D. Aplicación del algoritmo XGBoost

También se implementó el algoritmo XGBoost (Extreme Gradient Boosting) para la clasificación multiclase de señales cardíacas en cuatro categorías: AFIB (Fibrilación Auricular), SB (Bradicardia Sinusal), SR (Ritmo Sinusal) y ST (Taquicardia Sinusal).

Se utilizó el mismo conjunto de datos previamente preprocesado y codificado (usando LabelEncoder), que incluye múltiples características extraídas de registros de señales cardíacas. Los datos se dividieron en un 80 % para entrenamiento y un 20 % para prueba, manteniendo la proporción de clases mediante estratificación.

- `use_label_encoder=False`: se utiliza para evitar advertencias asociadas a versiones modernas de la API de XGBoost.
- `eval_metric='mlogloss'`: seleccionada por ser adecuada para tareas de clasificación multiclase, permitiendo una mejor optimización del modelo.
- `random_state=42`: se establece para asegurar la reproducibilidad de los resultados en diferentes ejecuciones.

El modelo se entrenó utilizando el conjunto de entrenamiento (`X_train`, `y_train`) y posteriormente se realizaron predicciones sobre el conjunto de prueba (`X_test`).

Para evaluar el desempeño del modelo, se generaron los siguientes indicadores:

- Un reporte de clasificación que incluye métricas clave como precisión, *recall* y *f1-score* para cada clase, proporcionando una evaluación detallada del desempeño del modelo.
- Una matriz de confusión que resume visualmente los aciertos y errores de clasificación para cada clase, facilitando la identificación de patrones de confusión entre etiquetas.

Ambas métricas permitieron comparar objetivamente el desempeño de XGBoost frente a los modelos de redes neu-

ronales previamente desarrollados, especialmente en términos de balance entre clases y precisión global.

V-E. Aplicación del algoritmo K-Means

Se utilizó K-Means con el objetivo de analizar cómo se agrupan las señales ECG según sus características, sin hacer uso de las etiquetas reales. Para ello, se trabajó con las características extraídas, luego, se modificó la columna Gender, asignando el valor 1 para "FEMALE" y 0 para "MALE", utilizando la función map(). Luego, se eliminaron las columnas Registro y Estado, ya que no eran útiles para el clustering no supervisado.

Luego de preparar los datos, se aplicó un proceso de normalización usando la función StandardScaler() de la librería Scikit-learn. Esto fue necesario porque el algoritmo K-Means trabaja midiendo distancias, y si las variables tienen escalas diferentes, los resultados pueden verse afectados. Una vez los datos estaban en la misma escala, se aplicó el algoritmo K-Means pidiendo que formara cuatro grupos ($n_{clusters}=4$), ya que originalmente se tenían cuatro tipos de arritmia. Además, se usó el parámetro random state=42 para que el agrupamiento fuera siempre igual cada vez que se ejecutara el código, es decir, para que los resultados fueran repetibles. También se incluyó el parámetro n init='auto', que permite que el algoritmo pruebe varias formas de iniciar los grupos antes de quedarse con la mejor, ayudando así a que los grupos finales sean más estables y representativos.

Para facilitar la visualización de los resultados del agrupamiento, se aplicó el Análisis de Componentes Principales (PCA), una técnica de reducción de dimensionalidad que permitió representar los datos en un plano de dos dimensiones sin perder demasiada información relevante. De esta forma, fue posible graficar los puntos agrupados por K-Means y observar la distribución de los clústeres en el espacio. Luego, con el fin de comparar los grupos generados por el algoritmo con las clases reales del conjunto de datos, se utilizó el algoritmo de asignación linear-sum-assignment, que permite encontrar la mejor correspondencia posible entre clústeres y etiquetas conocidas. A partir de esta correspondencia, se generó un reporte de clasificación y una matriz de confusión, lo cual permitió comparar de manera estructurada las etiquetas reales con las asignadas por K-Means.

VI. RESULTADOS

Las características HRV, como sdnm, rmssd y pnni_50, mostraron variaciones entre diferentes ritmos, reflejando diferencias en la variabilidad cardíaca. El análisis espectral evidenció patrones específicos en distintas bandas de frecuencia, destacando las zonas F1 y F2. Por su parte, las características MFCC aportaron descriptores sobre la forma de la señal. Finalmente, se añadieron variables clínicas relevantes como edad, frecuencia cardíaca y duración del QRS.

Este conjunto integrado permitió abordar con mayor solidez las tareas de clasificación en los siguientes puntos del proyecto.

Se evaluaron tres modelos de clasificación de señales ECG: una red neuronal densa simple (Modelo 1), una red densa con

Batch Normalization (Modelo 2) y una red LSTM (Modelo 3). A continuación se presentan los resultados obtenidos en las predicciones sobre el conjunto de prueba, destacando la precisión global (accuracy) y el comportamiento frente a cada clase según las matrices de confusión.

- **Modelo 1 – Red Densa Simple:** El Modelo 1 alcanzó una accuracy de 95.92 %. La matriz de confusión (Figura 1) muestra que:

- La clase SB fue correctamente identificada en 773 ocasiones, con muy pocos errores.
- Se cometieron errores principalmente entre AFIB y SR, y entre ST y otras clases. Por ejemplo:
 - 14 muestras reales de ST fueron clasificadas como AFIB.
 - 17 muestras de SR fueron confundidas como SB.

Estos resultados muestran una buena precisión global, aunque con limitaciones al distinguir ritmos con patrones similares.

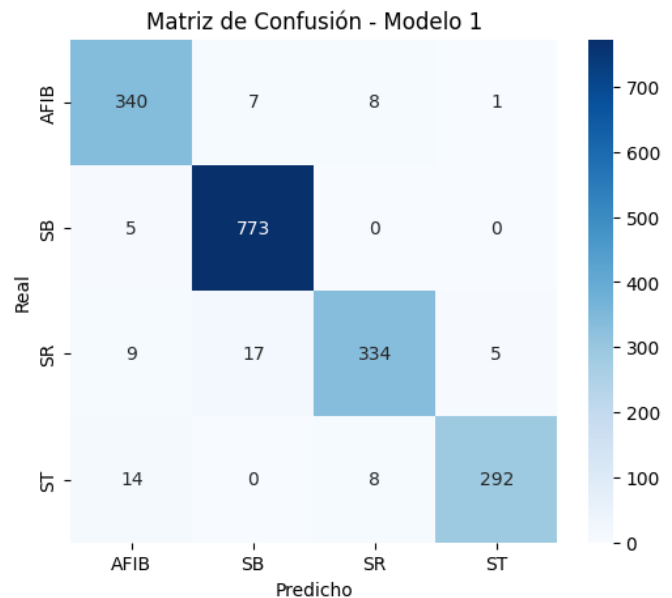


Figura 1: Matriz de confusión del Modelo 1.

Cuadro I: Reporte de clasificación multiclase con XGBoost

Clase	Precisión	Recall	F1-score	Soporte
AFIB	0.92	0.96	0.94	356
SB	0.97	0.99	0.98	778
SR	0.95	0.92	0.93	365
ST	0.98	0.93	0.95	314
Exactitud (accuracy)		0.96 (1813 muestras)		
Promedio macro	0.96	0.95	0.95	1813
Promedio ponderado	0.96	0.96	0.96	1813

El modelo muestra buen rendimiento general, pero presenta confusión entre AFIB-SR y errores notables al clasificar ST.

- **Modelo 2 – Red Densa con Batch Normalization:** El Modelo 2 obtuvo la mayor precisión entre los tres modelos, con una *accuracy* de 96.75 %. En su matriz de confusión (Figura 2) se observan mejoras en comparación al Modelo 1:

- La clase ST fue mejor clasificada:
 - Solo 11 errores como AFIB.
 - 0 errores como SB o SR.
- Disminuyeron los errores entre SR y SB:
 - Solo 15 casos de confusión.
- Se mantuvo una clasificación casi perfecta de SB:
 - 773 aciertos.

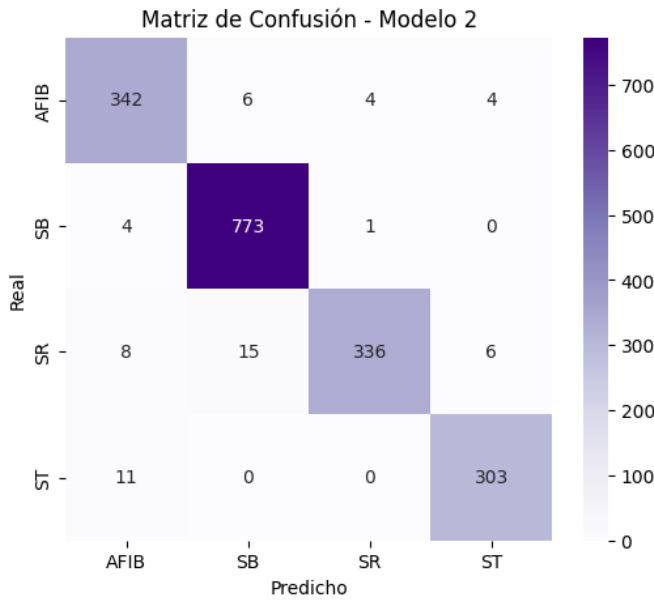


Figura 2: Matriz de confusión del Modelo 2.

Cuadro II: Reporte de clasificación multiclase con XGBoost

Clase	Precisión	Recall	F1-score	Soporte
AFIB	0.92	0.96	0.94	356
SB	0.97	0.99	0.98	778
SR	0.95	0.92	0.93	365
ST	0.98	0.93	0.95	314
Exactitud (accuracy)		0.96 (1813 muestras)		
Promedio macro	0.96	0.95	0.95	1813
Promedio ponderado	0.96	0.96	0.96	1813

Se observa una reducción importante de errores, especialmente en la clase ST. Este modelo logra una mejor diferenciación entre ritmos similares

- **Modelo 3 – Red LSTM:** El Modelo 3 alcanzó una *accuracy* de 96.64 %, muy cercana a la del Modelo 2. Sus resultados reflejan un comportamiento más balanceado:
 - Se clasificaron correctamente 339 instancias de SR, superando ligeramente al Modelo 2.

- SB se mantuvo con 771 aciertos, similar a los otros modelos.
- ST fue clasificada correctamente en 297 casos, con solo 12 errores como AFIB.

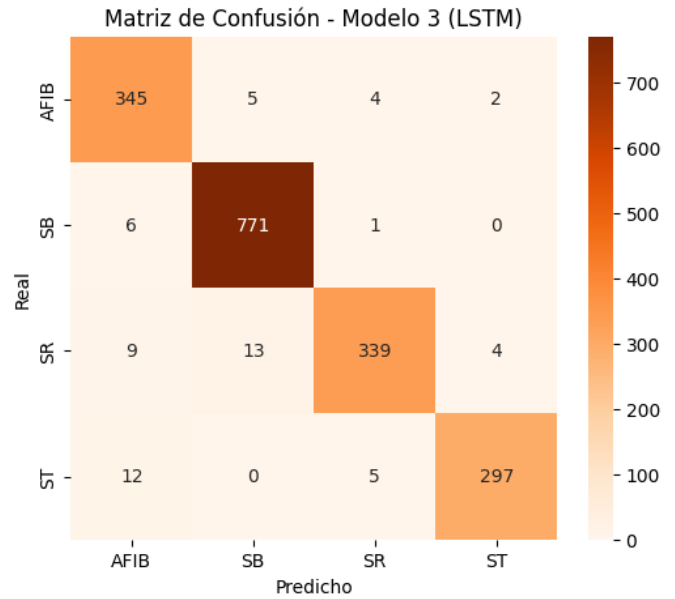


Figura 3: Matriz de confusión del Modelo 3.

Cuadro III: Reporte de clasificación multiclase con XGBoost

Clase	Precisión	Recall	F1-score	Soporte
AFIB	0.92	0.96	0.94	356
SB	0.97	0.99	0.98	778
SR	0.95	0.92	0.93	365
ST	0.98	0.93	0.95	314
Exactitud (accuracy)		0.96 (1813 muestras)		
Promedio macro	0.96	0.95	0.95	1813
Promedio ponderado	0.96	0.96	0.96	1813

El modelo LSTM logra un desempeño uniforme en todas las clases, aprovechando la naturaleza secuencial de los datos ECG.

VI-A. Comparación General

En la Figura 4 se presenta la comparación directa de la *accuracy* entre los tres modelos:

- **Modelo 2 (Denso + BN):** 96.75 %
- **Modelo 3 (LSTM):** 96.64 %
- **Modelo 1 (Denso Simple):** 95.92 %

Aunque las diferencias en *accuracy* son pequeñas, se observan mejoras consistentes en las predicciones clase a clase, especialmente con el Modelo 2.

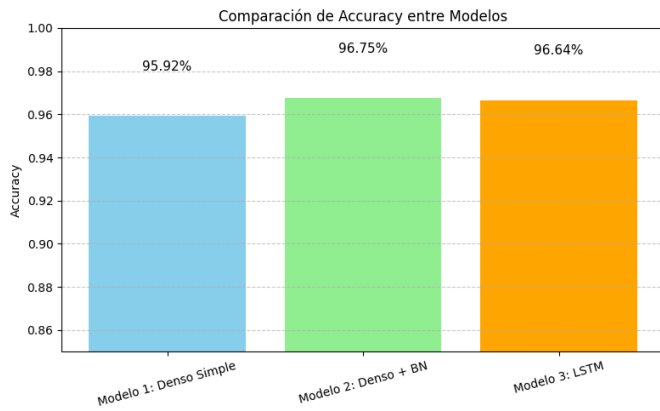


Figura 4: Comparación de accuracy entre modelos.

El Modelo 2 supera ligeramente al resto en precisión general. El Modelo 3 es competitivo, y el Modelo 1 queda rezagado en comparación.

VI-B. Resultados parte 5

En cuanto a la clasificación binaria AFIB vs SR usando tres modelos: simple, intermedio y complejo.

- **Modelo 1:** El Modelo 1 alcanzó una clasificación correcta de 342 señales de AFIB y 350 de SR, mostrando 14 falsos positivos de AFIB como SR y 15 falsos negativos. Este modelo demostró un buen desempeño general, aunque con limitaciones al diferenciar ritmos con patrones similares.
- **Modelo 2:** El Modelo 2 logró reducir los errores al clasificar señales de SR, presentando solo 6 muestras mal clasificadas. Sin embargo, mostró mayor confusión en la clasificación de señales de AFIB, con 102 errores, lo que sugiere que el modelo favoreció la predicción de SR y podría haber incurrido en sobreajuste hacia esa clase.
- **Modelo 3:** El Modelo 3 evidenció un mejor equilibrio entre sensibilidad y especificidad. Clasificó correctamente 346 señales de AFIB, con solo 10 errores, y cometió 19 errores en la clasificación de señales SR, mejorando así respecto al Modelo 1. La inclusión de la capa *Dropout* parece haber contribuido a reducir el sobreajuste y estabilizar el rendimiento general del modelo.

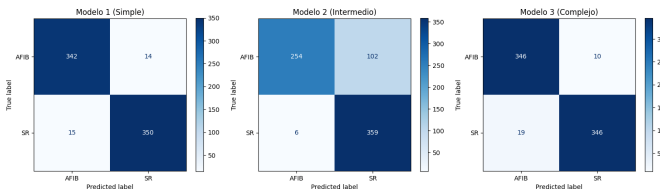


Figura 5

Aunque los tres modelos obtuvieron buenos resultados en la clasificación binaria de señales ECG, el Modelo 3 demostró ser el más robusto y equilibrado frente a ambas clases. Su capacidad para mantener el rendimiento sin caer en sobreajuste lo

posiciona como una arquitectura efectiva para tareas similares en el dominio biomédico.

Para la clasificación multiclase utilizando el algoritmo XGBoost, entrenado sobre las cuatro clases de patologías cardíacas: AFIB, SB, SR y ST. La matriz de confusión obtenida (ver Figura 5) y los valores de precisión, recall y F1-score reflejan un excelente rendimiento general, alcanzando una precisión total del 98 % sobre el conjunto de prueba

- **AFIB:** Se clasificaron correctamente 338 señales, con 18 errores repartidos entre SB, SR y ST. Esto representa un *recall* del 95 % y una precisión también del 95 %, indicando un equilibrio sólido entre detección y especificidad.
- **SB (ritmo sinusal lento):** Mostró el mejor rendimiento del modelo, con 775 clasificaciones correctas sobre 778 casos. La precisión fue del 99 % y el *recall* del 100 %, evidenciando que esta clase fue claramente reconocida por el modelo.
- **SR (ritmo sinusal normal):** Se identificaron correctamente 357 de 365 muestras, con solo 8 errores, principalmente confundidas con AFIB. La precisión y el *recall* fueron ambos del 98 %, reflejando una muy alta fiabilidad.
- **ST (ritmo sinusal taquicárdico):** Se obtuvo un total de 305 clasificaciones correctas sobre 314 muestras. La mayoría de los errores correspondieron a confusiones con AFIB, con un pequeño margen. La precisión fue del 99 % y el *recall* del 97 %.

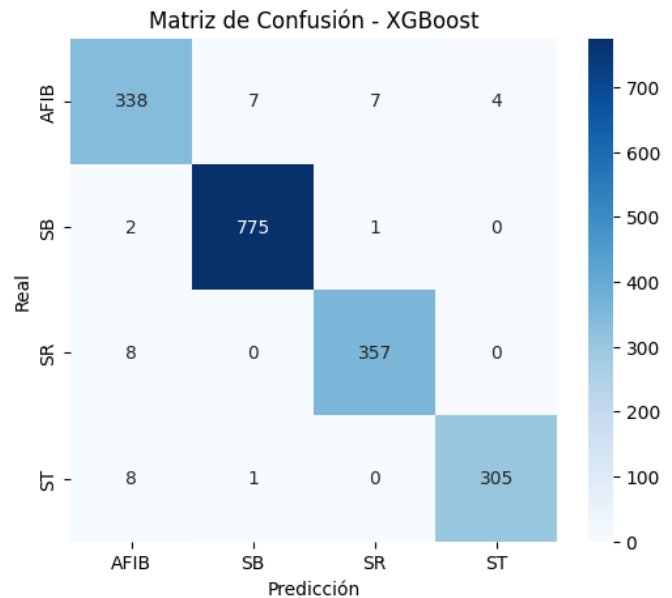


Figura 6: Matriz de confusión obtenida con el modelo XG-Boost.

Muestra un alto grado de precisión en todas las clases, con mínimos errores de clasificación. El mejor desempeño se observó en la clase SB, mientras que AFIB y ST mostraron

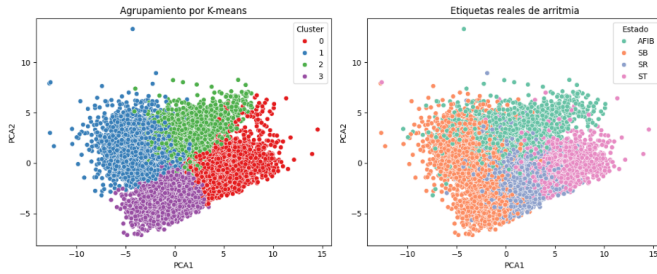


Figura 7: Agrupamiento con K-Means (izquierda) comparado con las etiquetas reales de arritmia (derecha) en el espacio reducido con PCA.

ligeras confusiones entre sí, aunque con resultados generales altamente satisfactorios.

VI-C. Resultados del algoritmo K-Means:

En la Figura 7 se presentan dos visualizaciones en el espacio bidimensional obtenido mediante Análisis de Componentes Principales (PCA). En la imagen de la izquierda se muestra el resultado del agrupamiento realizado con el algoritmo K-Means, donde se observaron cuatro clústeres diferenciados. Cada punto representa una observación, coloreada según el clúster asignado por el modelo. En la imagen de la derecha se muestran las mismas observaciones, pero coloreadas según las etiquetas reales de arritmia (AFIB, SB, SR, ST), lo que permite una comparación visual entre los grupos generados por el algoritmo y las clases verdaderas.

Además, se generó un reporte de clasificación tras alinear los clústeres obtenidos mediante K-Means con las etiquetas reales, utilizando el algoritmo de asignación:

Cuadro IV: Reporte de clasificación tras mapear clústeres a clases con K-Means

Clase	Precisión	Recall	F1-score	Soporte
AFIB	0.79	0.70	0.74	1779
SB	0.76	0.66	0.71	3889
SR	0.39	0.47	0.43	1826
ST	0.70	0.86	0.77	1568
Accuracy	0.66 (9062 muestras)			
Promedio macro	0.66	0.67	0.66	9062
Promedio ponderado	0.68	0.66	0.67	9062

También se generó una matriz de confusión que compara las predicciones del algoritmo K-Means con las clases reales, una vez que se asignaron los clústeres a etiquetas usando el algoritmo de asignación. En esta, se observa que la clase SB fue la mejor identificada, con 2.565 instancias correctamente agrupadas. La clase ST también presentó buenos resultados, con 1.346 casos bien clasificados. En cambio, la clase SR tuvo una dispersión considerable, con solo 864 casos correctamente agrupados, mientras que una gran cantidad fueron confundidos con ST y SB. La clase AFIB tuvo 1.249 aciertos, pero también un número significativo de errores, especialmente hacia SB. Estos resultados reflejan que, si bien el algoritmo

logró identificar algunas estructuras en los datos, la separación entre ciertas clases fue limitada, especialmente en aquellas con patrones más similares entre sí.

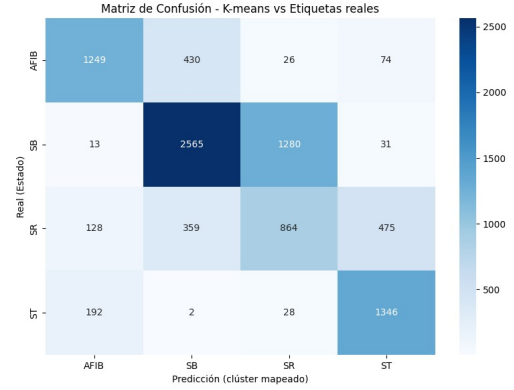


Figura 8: Matriz de confusión entre las predicciones de K-Means y las clases reales, tras la asignación óptima de clústeres.

VII. DISCUSIÓN

VII-A. Modelos supervisados multiclase:

Los resultados obtenidos con los modelos de clasificación supervisada evidencian un alto nivel de desempeño general en la identificación automática de las cuatro arritmias cardíacas seleccionadas (AFIB, SB, SR y ST). En particular, los tres modelos basados en redes neuronales artificiales mostraron una precisión superior al 95 %, lo que indica que las características extraídas del ECG contienen información relevante y discriminativa para esta tarea [4]. A pesar de tener arquitecturas diferentes, los modelos presentaron patrones de error similares, principalmente entre clases con morfologías cercanas.

El Modelo 1, una red densa simple, obtuvo una precisión global del 95.92 %. Aunque eficaz en términos generales, este modelo mostró confusiones significativas entre clases como AFIB y SR, lo que sugiere una limitada capacidad para capturar relaciones no lineales complejas. Este tipo de error es común en arquitecturas densas poco profundas, que suelen tener un rendimiento adecuado pero inferior frente a datos con alta variabilidad fisiológica [5].

El Modelo 2, que incorpora normalización por lotes, alcanzó la mayor precisión entre los tres modelos neuronales, con un 96.75 %. Esta técnica ha demostrado ser eficaz en la mejora de la estabilidad y la velocidad de entrenamiento de redes profundas, además de reducir el riesgo de sobreajuste [6]. En este modelo se observó una reducción de errores en la clasificación de ST y una menor confusión entre SR y SB, en comparación con el Modelo 1, lo que refleja una mejor diferenciación de ritmos cardíacos fisiológicamente cercanos.

El Modelo 3, basado en una red LSTM, logró una precisión del 96.64 %, apenas por debajo del Modelo 2. Las redes LSTM están diseñadas para capturar relaciones temporales y secuenciales dentro de los datos, lo cual es especialmente relevante para señales ECG, cuya morfología depende del

tiempo [7]. Este modelo presentó un rendimiento equilibrado entre las clases, y fue particularmente eficiente en la clasificación de ritmos normales y taquicardias, lo que sugiere que la representación temporal de los intervalos RR y otros patrones dinámicos fue aprovechada por la arquitectura.

Finalmente, el modelo XGBoost alcanzó la mayor precisión global, 98 %, en la clasificación multiclase. Este algoritmo de ensamble basado en árboles de decisión ha demostrado ser altamente competitivo en tareas estructuradas, con buena tolerancia al ruido y capacidad para manejar conjuntos de datos de alta dimensionalidad [8]. En este caso, la clase SB obtuvo un desempeño casi perfecto (99 % de precisión y 100 % de recall), mientras que las demás clases también superaron el 95 % en todas las métricas. Estos resultados coinciden con estudios previos donde XGBoost ha superado a redes neuronales en tareas similares cuando se trabaja con características preprocesadas [9].

VII-B. Clasificación binaria AFIB vs SR:

Además del análisis multiclase, se realizó una clasificación binaria centrada en diferenciar fibrilación auricular (AFIB) del ritmo sinusal (SR), debido que estas tenían la mayor cantidad de muestras en los datos. Para esta tarea se evaluaron las mismas arquitecturas de red neuronal utilizadas previamente, ahora adaptadas a una salida binaria.

El Modelo 1, logró resultados aceptables, pero mostró una tendencia a favorecer la clase con mayor número de casos (SR), evidenciando un sesgo en las predicciones. Este comportamiento es común cuando no se aplica un balance adecuado de clases o técnicas de regularización, especialmente en redes simples sin capas adicionales [10]

El Modelo 2, alcanzó una precisión alta en la clase SR pero a costa de una sensibilidad reducida hacia AFIB, lo cual puede ser clínicamente problemático, ya que el costo de no detectar casos de fibrilación auricular puede ser elevado. A pesar de sus buenos valores de accuracy, este modelo presentó un desequilibrio en la sensibilidad, lo que indica que no generaliza de manera óptima frente a clases más pequeñas.

Por otro lado, el Modelo 3, fue el que logró el desempeño más equilibrado entre ambas clases. Este modelo aprovechó mejor las secuencias temporales implícitas en las señales ECG, mejorando la capacidad de detección de eventos arrítmicos más sutiles, como los intervalos R-R irregulares típicos de AFIB, gracias a su estructura recurrente y a la aplicación de técnicas de regularización como Dropout, la red LSTM logró una mayor sensibilidad en la clase AFIB sin sacrificar la precisión en SR [11].

VII-C. Modelo no supervisado (K-Means):

Además de los modelos supervisados, se aplicó el algoritmo no supervisado K-Means con el objetivo de explorar la estructura interna de los datos y determinar si era posible identificar agrupamientos coherentes sin utilizar las etiquetas reales de arritmia. Este tipo de enfoque es útil en contextos donde no se dispone de etiquetas o se busca validar la separabilidad

inherente de las clases a partir de las características extraídas [12]

El modelo K-Means alcanzó una precisión global del 66 % al comparar los clústeres con las clases reales. Las clases SB y ST fueron las mejor representadas, mientras que SR y AFIB presentaron mayores confusiones. Estos resultados son consistentes con estudios previos que han identificado limitaciones de los métodos de clustering tradicionales para diferenciar patrones sutiles en señales fisiológicas, especialmente cuando la separación entre clases no es lineal o los datos presentan ruido [13]. La naturaleza de K-Means, que asume clústeres esféricos y separación basada en distancias euclidianas, puede limitar su capacidad para capturar relaciones complejas entre las variables, especialmente en señales fisiológicas donde los patrones pueden ser más abstractos o secuenciales [14].

A pesar de estas limitaciones, el resultado obtenido demuestra que existe cierta estructura en los datos que permite una agrupación parcial útil. Además, la visualización con PCA mostró que algunos grupos tienden a concentrarse en regiones separadas del espacio reducido, lo cual refuerza la validez de las características extraídas para representar la información cardíaca. Esta observación sugiere que, aunque K-Means no es competitivo con los modelos supervisados en cuanto a precisión, puede ser útil en fases exploratorias, como la detección de patrones, la identificación de subgrupos, o como base para métodos híbridos.

VII-D. Equipos comerciales que traen incorporados algoritmos de ayuda diagnóstica a partir de señales EKG

Actualmente, existen varios dispositivos comerciales que incorporan algoritmos avanzados de ayuda diagnóstica para el análisis automático de señales EKG, lo que ha permitido mejorar la precisión clínica y reducir los tiempos de interpretación. Estos sistemas se basan en muchos de los principios estudiados en este curso, como el análisis de bioseñales, transformaciones espectrales, extracción de características, clasificación y procesamiento digital mediante paquetes computacionales como NumPy, SciPy, Pandas y Matplotlib.

Uno de estos equipos es el Philips PageWriter TC70, que utiliza el algoritmo DXL de 16 derivaciones. Este algoritmo realiza análisis automáticos del ECG para detectar ritmos anormales como fibrilación auricular, flutter auricular y taquicardias, aplicando una lógica similar a los flujos de procesamiento vistos en el curso: primero realiza filtrado digital y extracción de características temporales y espectrales, luego analiza patrones de forma de onda y toma decisiones basadas en reglas o modelos entrenados previamente [15].

Otro ejemplo destacado es el GE MAC VU360, que integra el algoritmo Marquette™ 12SL, ampliamente utilizado para el diagnóstico de una variedad de arritmias y condiciones cardíacas. Este algoritmo realiza representaciones de señales en el dominio del tiempo y la frecuencia, aplicando principios similares a la transformada de Fourier para detectar alteraciones morfológicas. Además, se apoya en la correlación cruzada y medidas estadísticas para identificar desviaciones en la actividad eléctrica cardíaca [16].

Finalmente, el KardiaMobile 6L de AliveCor, un dispositivo portátil con seis derivaciones, emplea algoritmos de inteligencia artificial que permiten identificar en segundos eventos como fibrilación auricular, bradicardia y taquicardia sinusal. Este dispositivo refleja cómo el procesamiento digital avanzado —incluyendo transformadas rápidas de Fourier, análisis de componentes espectrales y técnicas de clasificación supervisada como las vistas en Python con scikit-learn o TensorFlow— puede implementarse eficientemente en hardware de consumo para monitoreo ambulatorio [17].

Estos ejemplos muestran cómo los fundamentos teóricos y computacionales abordados en el curso, como el análisis de señales en el dominio del tiempo y la frecuencia, la extracción de características, y la clasificación mediante modelos, son aplicables directamente en soluciones reales utilizadas en entornos clínicos modernos.

VIII. CONCLUSIONES

La combinación de características espectrales, temporales y clínicas extraídas de señales ECG permitió construir representaciones informativas y discriminativas, las cuales fueron efectivas para entrenar modelos de clasificación automática de arritmias. Este enfoque multicaracterístico contribuyó al alto desempeño alcanzado por los modelos supervisados, particularmente en la distinción entre ritmos cardíacos fisiológicamente diferentes.

Los modelos supervisados, tanto redes neuronales profundas como XGBoost, alcanzaron altos niveles de precisión en la clasificación multiclase de arritmias, con desempeños superiores al 95 %. Las redes neuronales con normalización por lotes y las arquitecturas LSTM demostraron ser eficaces para capturar patrones complejos y temporales en las señales, mientras que XGBoost superó ligeramente en precisión global (98 %), confirmando su eficiencia sobre conjuntos de características estructuradas. Estos resultados destacan la importancia de elegir modelos en función del tipo de datos y el objetivo clínico.

Este trabajo confirma que el análisis automatizado de señales ECG, apoyado en técnicas de aprendizaje automático y extracción adecuada de características, puede contribuir al desarrollo de herramientas de apoyo diagnóstico confiables y reproducibles, con potencial para ser integradas en entornos clínicos y sistemas portátiles de monitoreo cardíaco.

REFERENCIAS

- [1] D. Patiño, J. Medina, R. Silva, A. Guijarro, and J. Rodríguez, "Predicción de arritmias e infartos agudos de miocardio usando aprendizaje automático," *Ingenius: Revista de Ciencia y Tecnología*, vol. 29, pp. 79–89, 2023. [Online]. Available: <https://doi.org/10.17163/ings.n29.2023.07>
- [2] J. Zheng, H. Chu, D. Struppa, J. Zhang, M. Yacoub, H. El-Askary, A. Chang, L. Ehwerhemuepha, I. Abudayyeh, A. Barrett, G. Fu, H. Yao, D. Li, H. Guo, and C. Rakovski, "Optimal multi-stage arrhythmia classification approach," *Scientific Reports*, vol. 10, no. 2898, pp. 1–12, Feb. 2020. [Online]. Available: <https://doi.org/10.1038/s41598-020-59821-7>
- [3] L. Saclova, A. Nemcova, R. Smisek, L. Smital, M. Vitek, and M. Ronzhina, "Reliable p wave detection in pathological ecg signals," *Scientific Reports*, vol. 12, no. 1, p. 6589, Apr. 2022. [Online]. Available: <https://doi.org/10.1038/s41598-022-10656-4>
- [4] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network," *Information Sciences*, vol. 405, pp. 81–90, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025517306539>
- [5] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, p. 258, 2017. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29034226/>
- [6] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: <https://arxiv.org/abs/1502.03167>
- [7] S. Saadatnejad, M. Oveisi, and M. Hashemi, "Lstm-based ecg classification for continuous monitoring on personal wearable devices," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 515–523, 2020. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/30990452/>
- [8] XGBoost Contributors, "Xgboost documentation: Model tutorial," XGBoost Read the Docs, 2025, version 3.0.2; accessed July 14, 2025. [Online]. Available: <https://xgboost.readthedocs.io/en/stable/tutorials/model.html>
- [9] L. Abualigah, S. A. Alomari, M. H. Almomani, R. A. Zitar, K. Saleem, H. Migdady, V. Snasel, A. Smerat, and A. E. Ezugwu, "Artificial intelligence-driven translational medicine: a machine learning framework for predicting disease outcomes and optimizing patient-centric care," *Journal of Translational Medicine*, vol. 23, no. 1, p. 302, 2025. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11892274/>
- [10] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016. [Online]. Available: <https://link.springer.com/article/10.1007/s13748-016-0094-0>
- [11] A. M. Ruiz, "Experimentos con redes neuronales recurrentes lstm para la predicción del nivel de glucosa de pacientes con diabetes," *Revista Ontare*, vol. 11, no. 1, 2023. [Online]. Available: <https://journal.universidadean.edu.co/index.php/Revistao/article/view/3688>
- [12] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167865509002323>
- [13] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data," *PLOS ONE*, vol. 8, no. 6, p. e66341, 2013. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0066341>
- [14] Neptune.ai, "K-means clustering explained," Neptune.ai Blog, May 2025, accessed July 14, 2025. [Online]. Available: <https://neptune.ai/blog/k-means-clustering>
- [15] Philips Healthcare. (2025) Cardiografos pagewriter tc70. Accedido el 15 de julio de 2025. [Online]. Available: <https://www.philips.es/healthcare/product/HC860315/cardiografos-pagewriter-tc70-cardigrafo>
- [16] GE HealthCare. (2025) Mac vu360 resting ecg. Accedido el 15 de julio de 2025. [Online]. Available: <https://www.gehealthcare.com/products/diagnostic-ecg/resting-ecg/mac-vu360>
- [17] AliveCor. (2025) Kardiabmobile 6l. Accedido el 15 de julio de 2025. [Online]. Available: <https://alivecor.com/products/kardiabmobile6l>