

Análisis de Señales ECG y Detección de Arritmias

Proyecto 1

Dina Yared Parra Romero

Cristian Tamayo Arango

Carolina Ruiz Morales

Facultad de Ingeniería

Universidad de Antioquia

Profesor: John Freddy Ochoa / Juliana Moreno Rada

Análisis de señales ECG y detección de arritmias

1. Resumen

Las arritmias cardíacas son alteraciones en la frecuencia o ritmo del corazón que pueden derivar en complicaciones graves si no son detectadas a tiempo. El electrocardiograma (ECG) constituye una herramienta clave en el diagnóstico de estas condiciones. En este proyecto se realiza un análisis estadístico detallado de una base de datos de señales ECG, con el objetivo de identificar patrones de prevalencia de arritmias según edad y género, así como establecer un subconjunto óptimo de datos para futuros modelos de aprendizaje automático. Los resultados permiten comprender mejor la distribución demográfica de las arritmias y sentar bases sólidas para modelos predictivos..

2. Introducción

Las arritmias cardíacas, trastornos que afectan el ritmo del corazón, representan un riesgo significativo para la salud, lo que subraya la importancia de la detección temprana para una intervención oportuna y la prevención de complicaciones graves como la insuficiencia cardíaca y los accidentes cerebrovasculares [1]. El electrocardiograma (ECG) es una herramienta indispensable en este campo, ya que ofrece un medio no invasivo para monitorear la actividad eléctrica del corazón e identificar desviaciones de los ciclos cardíacos normales [2]. Este informe detalla un análisis de los datos de ECG, centrándose en la identificación y caracterización de diversas arritmias mediante el examen de características clave del ECG como la onda P, el complejo QRS y la onda T.

El proyecto utiliza el archivo `Diagnostics.xlsx`, que contiene información clínica vinculada a los registros de ECG, y el archivo `ECGDataDenoised.zip`, que proporciona señales de ECG preprocesadas. A través del análisis estadístico y la aplicación de técnicas de procesamiento de señales, el objetivo es discernir patrones indicativos de arritmias prevalentes, explorar posibles tendencias relacionadas con el género y la edad, y determinar conjuntos de datos óptimos para entrenar algoritmos de detección de arritmias. Además, el informe evalúa la eficacia de diferentes metodologías en la detección de componentes críticos del ECG, sobre todo los picos R, utilizando librerías como `NeuroKit` y `Detectors`. El análisis se extiende a la variabilidad de la frecuencia cardíaca (VFC), una medida que refleja la influencia del sistema nervioso autónomo en la función cardíaca, y la extracción de características relevantes de la VFC utilizando la librería `hrvanalysis`. Al comparar estas características entre diferentes grupos de arritmias, este estudio busca identificar marcadores sólidos para distinguir entre diversas afecciones cardíacas, lo que en última instancia contribuye al avance del análisis automatizado de ECG y la detección de arritmias.

3. Metodología

3.1. Materiales y Métodos

- **Base de datos:** Se utiliza la base de datos Chapman ECG [14], la cual contiene registros de señales ECG de 10,646 pacientes con diversas condiciones cardíacas.

- **Archivo clínico:** Se utiliza el archivo *Diagnostics.xlsx*, el cual proporcionó información como edad, género y diagnóstico de cada paciente.
- **Software de análisis:** Python 3.9, con las librerías:
 1. *NeuroKit2* para la detección de picos R y análisis de señales biológicas.
 2. *py-ecg-detectors* para la comparación de métodos de detección de picos.
 3. *hrvanalysis* para la extracción de características de variabilidad de la frecuencia cardíaca (HRV).
 4. *pandas*, *numpy*, *matplotlib* y *seaborn* para el manejo de datos y la visualización gráfica.
- **Herramientas de análisis:** Se utiliza Jupyter Notebook como entorno de trabajo.

3.2. Procedimiento

3.2.1. Análisis 1: Exploración y selección de datos

Para este análisis, se utilizó un enfoque exploratorio-descriptivo mediante el procesamiento de datos del archivo *Diagnostics.xlsx* en Jupyter Notebook, empleando Python y las bibliotecas *pandas*, *matplotlib* y *seaborn*. Primero, se analizó la prevalencia de arritmias mediante *value_counts()* y gráficos de barras, permitiendo identificar las arritmias más frecuentes. Posteriormente, se estudió su distribución por género a través de tablas cruzadas y gráficos de barras apiladas. También se analizó la relación entre edad y tipo de arritmia utilizando agrupaciones por intervalos de edad (*pandas.cut()*) y gráficos de caja (*boxplot()*).

Con base en estos resultados, se seleccionó un subconjunto de pacientes entre 20 y 80 años para futuros análisis, priorizando representatividad y minimizando sesgos. Finalmente, se compararon los hallazgos con literatura científica (*Scielo*, *Fundación Española del Corazón*, *NHLBI*), confirmando que las tendencias observadas coincidían con estudios previos. Cada etapa fue complementada con visualizaciones y su respectiva interpretación.

3.2.2. Análisis 2: Selección de herramienta de análisis

Para evaluar el mejor modelo de detección de picos R, se seleccionaron dos tipos de arritmias utilizando la información contenida en el archivo *Diagnostics.xlsx* de la base de datos. A través de la función *andas.sample()*^p, se extrajeron aleatoriamente 20 muestras de cada arritmia, obteniendo un total de 40 registros. Cada señal fue procesada inicialmente con la función *nk.ecg-peaks()* de *NeuroKit2*, que permitió identificar los picos R y visualizar la señal asociada. Posteriormente, se empleó la función *Detectors()* de la librería *ecgdetectors*, probando todos los métodos disponibles para evaluar su desempeño de forma visual mediante gráficas. Tras esta comparación, se seleccionó el modelo Engzee por su mejor desempeño en la detección de picos en señales con características complejas. Finalmente, se realizaron comparaciones directas entre los resultados obtenidos con *nk.ecg-peaks()* y el modelo Engzee, utilizando señales más difíciles, para confirmar cuál de los dos métodos ofrecía un rendimiento óptimo.

3.2.3. Análisis 3: Variabilidad de la frecuencia cardiaca (HRV) y otros índices ECG

Primero, se realizó un filtrado de outliers en el conjunto de datos agrupando las observaciones según el tipo de ritmo cardiaco registrado en la columna 'Rhythm'. Para cada grupo, se calcularon el primer cuartil (Q1) y el tercer cuartil (Q3) de la edad de los pacientes ('PatientAge'), determinando así el rango intercuartílico (IQR). Con base en estos valores, se establecieron límites de aceptación definidos por el intervalo $[Q1 - 1,5 \times IQR, Q3 + 1,5 \times IQR]$. Posteriormente, se filtraron y conservaron únicamente las observaciones cuya edad se encontraba dentro de estos límites, eliminando así los valores atípicos. Finalmente, los datos filtrados de cada grupo se combinaron para formar un nuevo DataFrame libre de outliers. A continuación, se eliminaron los registros corres-

pondientes a edades extremas, excluyendo pacientes menores de 20 años y mayores de 80 años. Posteriormente, se realizó un balanceo cuidadoso de la distribución por género para asegurar la representatividad de la muestra. Este balanceo se llevó a cabo midiendo el tamaño del grupo más pequeño en cada tipo de arritmia y, para el grupo más grande, seleccionando aleatoriamente datos dentro del rango de edad del grupo menor, de modo que ambos quedaran igualados en número. Para la detección de los picos R en los regis-

tros de ECG, se implementó la librería NeuroKit2, utilizando específicamente la función *nk.ecg_process()*. De la amplia cantidad de parámetros generados, se seleccionaron únicamente los valores de SDNN, RMSSD y LF/HF, ya que son considerados los más relevantes para el análisis de variabilidad de la frecuencia cardíaca, además de optimizar el uso de recursos computacionales. Paralelamente, mediante la librería hrvanalysis y su función *get_time_domain_features()*, se extrajeron todas las métricas de variabilidad proporcionadas, dado que el volumen de datos que proporciona la función es considerablemente menor. Finalmente, se construyeron dos bases de datos distintas, cada una conteniendo las métricas de variabilidad obtenidas de ambos métodos de análisis.

3.2.4. Análisis 4: Comparación de características entre grupos

Para la etapa de comparación de características entre grupos, se trabajó sobre las bases de datos construidas en el Análisis 3, que contenían métricas de variabilidad de la frecuencia cardíaca (HRV) extraídas mediante NeuroKit2 y hrvanalysis y se generaron gráficos de distribución (boxplots y violin plots) para representar visualmente las diferencias entre los grupos. Luego se eligieron características representativas como SDNN, RMSSD y LF/HF, basándose en su relevancia clínica documentada en la literatura. A partir de estos parámetros, se realizaron comparaciones estadísticas entre los grupos de arritmias seleccionados. Se aplicaron pruebas de hipótesis, utilizando la prueba t de Student para

muestras independientes o la prueba de Mann-Whitney U en caso de no cumplir supuestos de normalidad (evaluados mediante la prueba de Shapiro-Wilk). Finalmente se calculó el valor p de cada comparación para determinar la significancia estadística de las diferencias observadas. A partir de los resultados, se identificó cuál de las métricas proporcionaba una mejor discriminación entre los grupos, basándose en el nivel de separación visual en los gráficos y en la magnitud del efecto observado en las pruebas estadísticas.

4. Análisis y Resultados

4.1. Análisis 1: Exploración y selección de datos

Para responder a las preguntas planteadas, se cargó el archivo `Diagnostics.xlsx` y se realizó una limpieza básica de los datos. A partir de ello, se estructuró un análisis exploratorio en varios apartados, considerando aspectos de prevalencia global, diferencias por género, distribución por edad, selección de subconjunto útil para modelos de aprendizaje, y comparación con literatura científica.

4.1.1. ¿Qué tipo de arritmias son más prelevantes?

Se utilizó la función `value_counts()` sobre la columna correspondiente al tipo de ritmo cardíaco para conocer la distribución general. Los resultados fueron representados mediante un gráfico de barras Figura 1 que mostró que los ritmos más frecuentes fueron el ritmo sinusal normal (NSR), la fibrilación auricular (AFIB), la taquicardia supraventricular (ST), la bradicardia sinusal (SB) y el ritmo sinusal irregular (SBR). Estos hallazgos son consistentes con lo observado en entornos clínicos reales, donde la fibrilación auricular es la arritmia sostenida más prevalente, especialmente en personas mayores. Es la primera causa de eventos embólicos como: el deterioro cognitivo, los episodios de insuficiencia cardíaca y la disminución de la calidad de vida [3]

En la Figura 1 Se seleccionaron pacientes entre 20 y 80 años. Las arritmias seleccionadas: SB, SR, ST, AFIB, SVT, AF, SA, AT. Además, el total de registros seleccionados fueron 9122

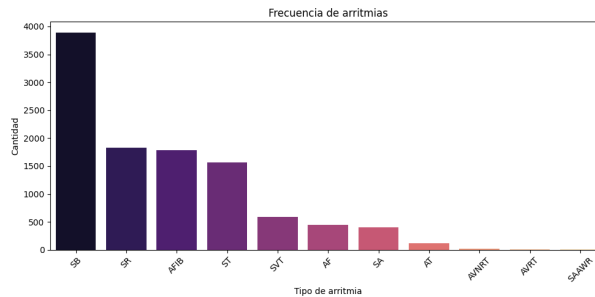


Figura 1: Resultados de la prevalencia de las arritmias en diferentes grupos.

4.1.2. Prevalencia de arritmias por género.

Mediante la función `groupby()` se agruparon los datos por tipo de arritmia y género, reorganizándolos con `unstack()` para facilitar su graficación. Se utilizó un gráfico de barras apiladas Figura 2, con una codificación de color diferenciada (*azul para hombres y rosado para mujeres*), lo que permitió identificar que ciertas arritmias, como la fibrilación auricular y la taquicardia supraventricular, tienen mayor prevalencia en el género masculino. Otras, como la bradicardia sinusal, presentaron una distribución más equilibrada. Este análisis respalda lo encontrado, que reportan una mayor frecuencia de fibrilación auricular en hombres, asociada a factores estructurales y hormonales [4].

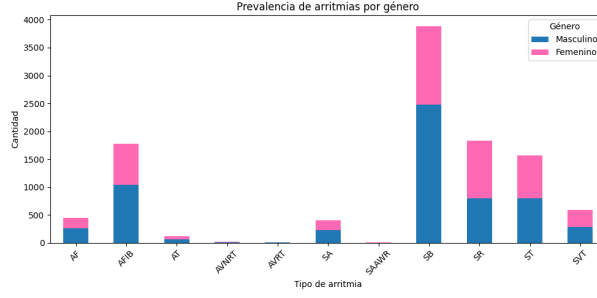


Figura 2: Resultados de las arritmias más prelevantes por género.

4.1.3. Distribución de arritmias según edad del paciente.

Para este análisis, se seleccionaron las cinco arritmias más frecuentes y se utilizó un gráfico combinado de caja de bigotes (boxplot) con puntos individuales (swarmplot) Figura 3. Este tipo de visualización permite observar tanto la distribución estadística (mediana, cuartiles, valores extremos) como la densidad real de los datos. Se observó que los pacientes con fibrilación auricular tienden a tener mayor edad, mientras que ritmos como NSR y ST son más comunes en adultos jóvenes. La visualización con swarmplot resulta útil para confirmar que la tendencia central reflejada en la caja de bigotes representa adecuadamente la distribución de la muestra [5].

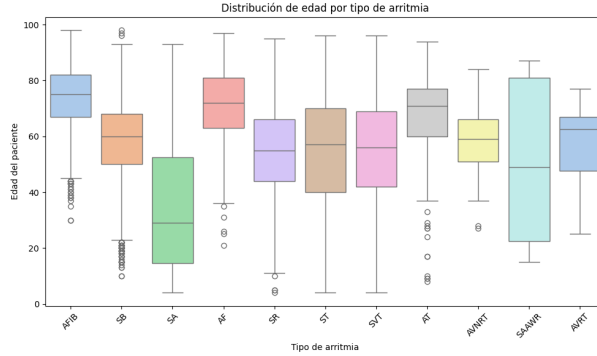


Figura 3: Resultados de la variabilidad de edad para cada tipo de arritmia.

4.1.4. Selección de subconjunto de datos para aprendizaje automático.

Para la creación de modelos de aprendizaje automático orientados a la clasificación de ritmos cardíacos, es fundamental contar con un subconjunto de datos que contenga suficientes muestras por clase y que sea representativo del grupo poblacional afectado por arritmias relevantes. En este sentido, se exploraron diferentes variables para establecer criterios de selección, principalmente edad y género. A partir del análisis descriptivo de la base de datos, se observó que las arritmias más prevalentes como la fibrilación auricular (AFIB), el ritmo sinusal irregular (SBR) y el ritmo de marcapasos ventricular (VP) tienden a presentarse en individuos de mayor edad. Esta tendencia fue confirmada mediante gráficos de caja (boxplots), donde se evidencia que la mediana de edad para estas arritmias se ubica por encima de los 60 años. Con base en estos hallazgos, se decidió limitar el subconjunto de entrenamiento a pacientes con edades entre 20 y 80 años, rango en el cual se concentra la mayor cantidad de casos clínicamente relevantes..

Aunque el análisis por género mostró diferencias en la distribución de ciertas arritmias por ejemplo, una mayor frecuencia de AFIB en hombres, se optó por no segmentar los datos por esta variable en la fase de selección del subconjunto. Esta decisión se basa en dos consideraciones fundamentales. En primer lugar, al reducir el conjunto de entrenamiento filtrando también por género, se corre el riesgo de disminuir el poder estadístico del modelo al limitar la cantidad de muestras por clase. En segundo lugar, diversos estudios y artículos recientes indican que, si bien existen diferencias de presentación y sintomatología entre hombres y mujeres, la incidencia global de arritmias puede ser similar entre ambos sexos. Por ejemplo, la Fundación Española del Corazón señala que “las arritmias pueden afectar tanto a hombres como a mujeres” [3]. De forma complementaria, la Sociedad Española de Cardiología también destaca que, aunque hay diferencias fisiopatológicas y de respuesta al tratamiento, no siempre existe un sesgo claro en la prevalencia general [6]

Por lo tanto, el criterio de edad se utilizó como principal variable de segmentación para asegurar una representación clínica válida y una distribución suficiente de casos para entrenar modelos robustos. Se mantiene el género como una variable disponible en el conjunto de datos para análisis posteriores, como la detección de sesgos en modelos o la personalización de predicciones, pero no se utilizó en esta fase para evitar comprometer la generalización del modelo por exceso de segmentación Figura 4.

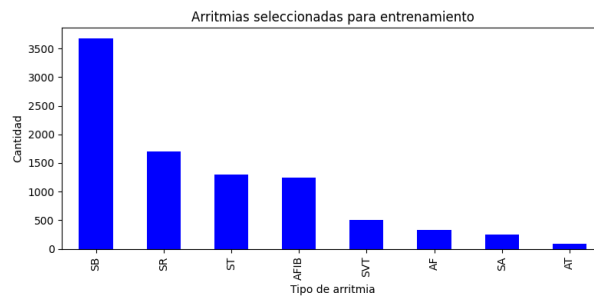


Figura 4: Resultados de las arritmias seleccionadas para entrenamiento.

4.1.5. Comparación con literatura científica.

Los resultados obtenidos en el análisis exploratorio muestran que la fibrilación auricular (AFIB) es la arritmia no fisiológica más prevalente en la base de datos, con una alta concentración en pacientes mayores de 60 años, predominantemente de sexo masculino. Este hallazgo coincide con múltiples estudios y revisiones clínicas que destacan la edad avanzada y el género masculino como factores de riesgo relevantes para el desarrollo de esta arritmia. En particular, la revisión publicada en la revista cubana de Cardiología y Cirugía Cardiovascular subraya que la fibrilación auricular representa una de las arritmias más frecuentes en la práctica clínica, con una incidencia creciente con la edad, especialmente en hombres mayores de 65 años [3]. Este patrón se observa también en nuestro análisis, donde los boxplots revelan que la mediana de edad para AFIB se encuentra por encima de los 70 años, lo cual sugiere una clara asociación entre envejecimiento y riesgo arrítmico.

Asimismo, la Fundación Española del Corazón señala diferencias notables entre sexos en la presentación de arritmias. Se menciona que las arritmias como la fibrilación auricular son más prevalentes en hombres, mientras que las mujeres presentan con mayor

frecuencia taquicardias supraventriculares (ST) y mayor sensibilidad a desencadenantes emocionales [4]. Esto se alinea con los hallazgos del gráfico de barras apiladas, donde se observa una clara predominancia masculina en el caso de AFIB, y una distribución más equilibrada para otros tipos de ritmos. Por otra parte, el Instituto Nacional del Corazón,

los Pulmones y la Sangre (NHLBI, por sus siglas en inglés) del NIH también confirma la influencia de la edad como uno de los principales desencadenantes de arritmias, con mayor riesgo en adultos mayores debido al deterioro natural del tejido cardíaco y a la acumulación de enfermedades estructurales previas [5]. Este factor se refleja claramente en los datos, donde se identifican umbrales superiores de edad en arritmias como AFIB y SB, en comparación con ritmos fisiológicos como NSR, que se distribuyen con mayor uniformidad en pacientes jóvenes y adultos medios. Estos puntos de coincidencia entre los datos

empíricos y la evidencia científica fortalecen la validez del conjunto de datos utilizado y permiten inferir que su uso en modelos de predicción de arritmias mediante inteligencia artificial estaría clínicamente justificado. Además, la coherencia observada respalda el enfoque metodológico empleado en este estudio, tanto en la selección de variables como en el tratamiento estadístico de las mismas.

4.1.6. Propuestas de análisis adicionales.

Finalmente, proponemos realizar análisis más profundos, como la relación entre frecuencia cardíaca y tipo de arritmia, la evolución temporal de pacientes si se cuenta con registros longitudinales, y el uso de algoritmos de clustering no supervisado para explorar agrupamientos ocultos en los datos. Estos enfoques podrían complementar los modelos supervisados y aportar una perspectiva más integral del fenómeno de las arritmias.

4.2. Análisis 2: Selección herramienta de análisis.

Se realizó una evaluación comparativa entre dos librerías especializadas en procesamiento de señales ECG: NeuroKit2 y py-ecg-detectors. El análisis se llevó a cabo utilizando un conjunto de 40 registros de arritmias cardíacas de diversa complejidad, con el objetivo de determinar cuál de las dos herramientas ofrecía un mejor desempeño en la identificación precisa de los picos R bajo condiciones variables.

Tras realizar la comparación de los algoritmos disponibles en la librería py-ecg-detectors, se observó que EngzeeMod presentó el mejor desempeño dentro de dicho conjunto. Este algoritmo demostró una mayor precisión en la identificación de picos R reales en comparación con otros métodos como evaluados (como Hamilton o Christov) como se puede evidenciar en la Figura 5, mostrando una menor tasa de falsos positivos, la capacidad de detectar de manera mas constante y una mejor capacidad para manejar segmentos de señal afectados por variaciones morfológicas o ruido moderado.

Sin embargo, al realizar una comparación directa entre el algoritmo de mejor rendimiento de py-ecg-detectors (Engzee) y la librería NeuroKit2, se observó que esta última demostró un desempeño superior en términos de precisión y robustez. NeuroKit2 mostró una mayor consistencia en los resultados, particularmente al procesar señales ECG con características complejas (arritmias, variaciones morfológicas y presencia de artefactos),

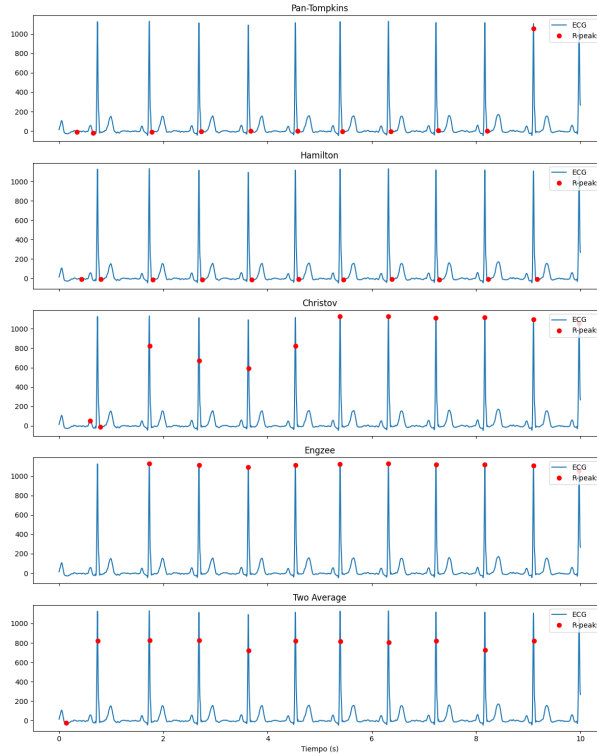


Figura 5: Comparacion de los diferentes modelos de py-ecg-detectors.

como se evidencia en la Figura 6. Dejando como la mejor opcion la libreria NeuroKit2, para procesar los picos R, del grupo de datos seleccionado.

4.3. Análisis 3: Variabilidad de la frecuencia cardíaca (HRV) y otros índices ECG.

NeuroKit2 demostró funcionalidades para la detección de picos R, sino tambien para el análisis de la variabilidad de la frecuencia cardíaca (HRV). La librería ofrece un conjunto completo de herramientas para calcular métricas de HRV tanto en el dominio temporal (RMSSD, SDNN, pNN50) como en el dominio frecuencial (potencia en bandas LF, HF). Su función integrada `hrv_time()` permite obtener estas mediciones de forma inmediata a partir de los picos R detectados. Además, NeuroKit2 incluye algoritmos avanzados para el análisis no lineal de la HRV a través de funciones como `hrv_nonlinear()`, que calcula parámetros como la entropía muestral (SampEn) y los exponentes de escalamiento DFA, particularmente útiles para evaluar la complejidad y regulación autonómica del ritmo cardíaco. Estas capacidades se complementan con el procesamiento robusto de la señal ECG, donde funciones como `ecg_delineate()` permiten identificar todas las componentes del complejo QRS, mientras que `ecg_quality()` y `signal_fixpeaks()` aseguran la fiabilidad de los resultados incluso en presencia de artefactos. Al comparar con otras alternativas como

`hrvanalysis`, se observó que aunque esta última ofrece herramientas especializadas para el análisis de la variabilidad cardíaca en los dominios frecuencial y no lineal, NeuroKit2 presenta una solución más integral. Cabe mencionar que el análisis se vio limitado por la duración de los registros disponibles (10 segundos), lo que restringió el cálculo de

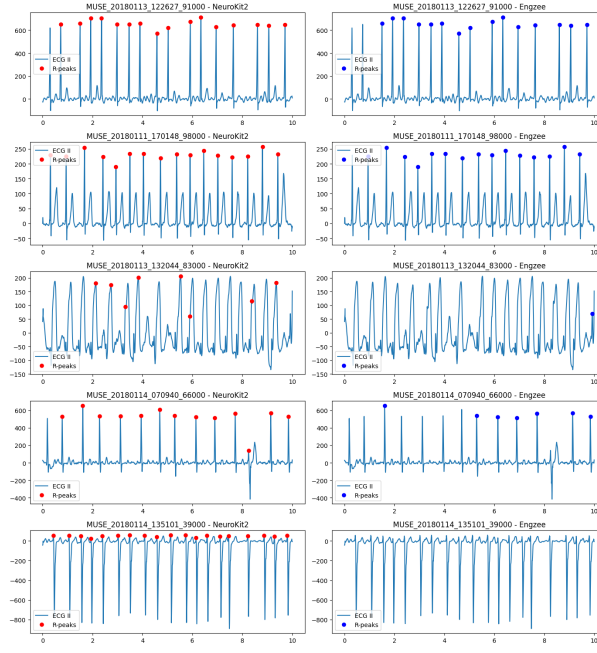


Figura 6: Comparación entre py-ecg-detectors y NeuroKit2.

ciertos parámetros de HRV que requieren series temporales más extensas para su correcta estimación. Esta restricción llevó a enfocar el estudio principalmente en las métricas del dominio temporal, donde NeuroKit2 demostró un rendimiento óptimo incluso con señales de corta duración.

4.4. Análisis 4: Comparación de características entre grupos

4.4.1. Comparación de las características de HRV como SDNN, RMSSD y LF/HF

El análisis de la variabilidad de la frecuencia cardíaca (HRV) permite evaluar la dinámica del sistema nervioso autónomo sobre el corazón, proporcionando información relevante tanto en estados fisiológicos como patológicos. Comparar características específicas de HRV como SDNN, RMSSD y LF/HF resulta fundamental, ya que cada una ofrece una perspectiva diferente del comportamiento cardíaco.

1. SDNN (Desviación estandar de los intervalos entre latidos RR normales):

SDNN mide la variabilidad global de la frecuencia cardiaca. Captura todas las fluctuaciones del ritmo cardiaco, tanto como cambios lentos (como dormir), así como cambios rápidos (como estar estresado). Un SDNN muy bajo puede indicar estrés ó enfermedad [9]. SDNN representa la variabilidad total de los intervalos entre latidos, lo que nos permite mirar la cómo actúa el sistemas simpático y parasimpático a lo largo del tiempo de registro. Su análisis permite identificar alteraciones globales en el control autonómico del corazón [7]. La Figura 7 presenta la comparación de

los valores de SDNN entre los diferentes grupos de ritmos cardíacos analizados: AF,

AFIB, AT, SA, SB, SR, ST y SVT. El análisis se realizó mediante diagramas de caja (boxplots), los cuales permiten visualizar la mediana (línea verde), los cuartiles, la dispersión de los datos y los valores atípicos en cada grupo.

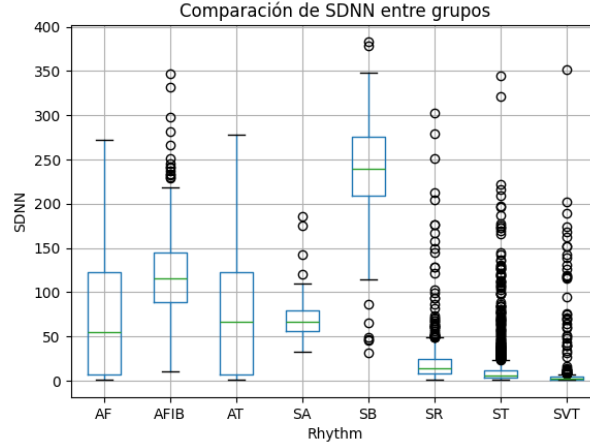


Figura 7: Comparación de los valores de SDNN entre los diferentes grupos.

Se observa que el grupo SB (bradicardia sinusal) presenta los valores más altos de SDNN, indicando una mayor variabilidad global de la frecuencia cardíaca, consistente con el ritmo lento característico de la bradicardia. En contraste, los grupos ST (taquicardia sinusal) y SVT (taquicardia supraventricular) muestran los valores más bajos de SDNN, mostrando una reducida variabilidad de la frecuencia cardíaca asociada a ritmos rápidos y respalda que taquicardias (ST, SVT) bajan RMSSD por mayor estrés simpático [9]. El grupo SR (ritmo sinusal normal) presenta valores de

SDNN intermedios, aunque con una dispersión y también algunos valores atípicos, lo cual es esperable en una población sana con variabilidad fisiológica natural. Por otro lado, los grupos AF y AFIB también muestran una amplia dispersión, mostrando heterogeneidad en las arritmias.

2. RMSSD (Raíz cuadrada de la media de las diferencias sucesivas entre intervalos RR):

RMSSD mide la variabilidad a corto plazo de los latidos cardiacos, no mira cambios lentos o a corto plazo como SDNN, sino que mira cómo cambia el latido de un latido a otro (como la respiración). Un RMSSD muy bajo puede indicar estrés, fatiga ó enfermedad [9]. RMSSD cuantifica la variabilidad de corto plazo entre intervalos consecutivos, siendo un marcador sensible de la actividad parasimpática. Compararlo permite detectar variaciones rápidas asociadas principalmente a mecanismos de relajación y recuperación fisiológica [7].

La Figura 8 muestra la comparación de los valores de RMSSD entre los diferentes grupos de ritmos cardiacos (AF, AFIB, AT, SA, SB, SR, ST, SVT) mediante diagramas de caja (boxplots). En cada boxplot se visualiza la mediana (línea verde), el rango intercuartil (caja), la dispersión general de los datos (bigotes) y los valores atípicos (círculos).

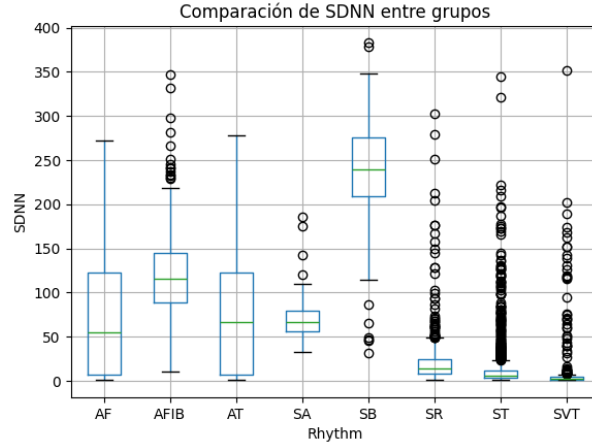


Figura 8: Comparación de los valores de RMSSD entre los diferentes grupos.

Se observa que el grupo SB (bradicardia sinusal) presenta nuevamente los valores de RMSSD más elevados entre todos los grupos, indicando una alta variabilidad de corto plazo entre intervalos RR, característica típica de ritmos lentos modulados principalmente por el sistema parasimpático [9].

En contraste, los grupos ST (taquicardia sinusal) y SVT (taquicardia supraventricular) muestran valores de RMSSD bajos, reflejando una escasa variabilidad de corto plazo propia de ritmos rápidos, en los cuales predomina la modulación simpática [9].

3. LFHF_NeuroKit (Cociente entre la energía en la banda de baja frecuencia (LF) y la alta frecuencia (HF) en el espectro de frecuencias de la HRV):

LF/HF Indica el balance autónomo entre el S. Simpático y el Parasimpático. Un LF/HF alto, indica predominio del simpático, y un LF/HF bajo indica predominio del parasimpático. Además, estima el balance entre las modulaciones simpática y parasimpática, ofreciendo una medida indirecta del equilibrio autonómico. La comparación de este índice permite identificar predominio de estrés o relajación en distintas condiciones [7].

Sin embargo, debido a que la mayoría de los datos de "LFHF_NeuroKit" están vacíos (NaN), no podemos realizar una comparación entre ellos, por lo que el gráfico sale en blanco. La ausencia de datos en la variable LFHF_NeuroKit puede deberse a que el cálculo del índice LF/HF requiere la obtención de un espectro de potencia fiable, lo cual generalmente exige señales de al menos 1 minuto de duración (*Las nuestras fueron señales de aproximadamente 10 segundos*) para poder capturar adecuadamente las modulaciones de baja frecuencia (LF) y alta frecuencia (HF) [7].

4.5. Pruebas de Hipótesis

Se decidió comparar solamente algunos de los grupos para realizar las pruebas estadísticas, uno de ellos correspondientes a ritmo sinusal normal (SR) y fibrilación auricular (AFIB), debido a que, el ritmo sinusal (SR) representa el patrón eléctrico cardíaco normal, caracterizado por intervalos regulares entre latidos [11]. Por otro lado, AFIB

es una de las arritmias más relevantes clínicamente, asociada a riesgos aumentados de accidente cerebrovascular e insuficiencia cardíaca. Lo que nos permitirá encontrar diferencias en la variabilidad de la frecuencia cardíaca (HRV) como SDNN y RMSSD [12]. Además, se realizaron análisis adicionales entre otros grupos clínicamente relevantes. Se

seleccionó la comparación entre SR y SB porque ambos corresponden a ritmos de origen sinusal, pero con diferencias fisiológicas ya que el ritmo sinusal normal presenta una frecuencia cardíaca dentro de rangos típicos, la bradicardia sinusal (SB) se caracteriza por una frecuencia cardíaca anormalmente baja, lo cual podría influir en la variabilidad de la frecuencia cardíaca (HRV) [13]. Por otro lado, se decidió comparar también ST

y SVT debido a que ambos representan ritmos rápidos, aunque de diferente origen. ST corresponde a una taquicardia sinusal, generalmente benigna y adaptativa, mientras que SVT (taquicardia supraventricular) suele involucrar mecanismos patológicos de reentrada a nivel auricular o nodal [7]. Esta comparación busca evaluar si las métricas de HRV son sensibles para distinguir entre taquicardias de diferente naturaleza. Para realizar las

pruebas estadísticas, primero filtramos los datos de SR y AFIB, luego probamos la normalidad con Shapiro-Wilk para cada grupo. Seguidamente, según el resultado, si es normal, usamos t-test, y si no es normal, usamos Mann-Whitney U test (no paramétrica).

4.5.1. Comparación de SDNN y RMSSD entre los grupos de SR y AFIB

Se realizó una prueba de normalidad mediante el test de Shapiro-Wilk para los grupos SR y AFIB, tanto en la variable SDNN como en RMSSD. Los resultados indicaron que en todos los casos los datos no siguen una distribución normal (p -valor ≤ 0.05), por lo cual se optó por utilizar la prueba no paramétrica de Mann-Whitney U para comparar las medianas entre los grupos. En la comparación de SDNN entre SR y AFIB, se obtuvo un

p -valor menor a 0.05, indicando diferencias estadísticamente significativas entre ambos grupos. De manera similar, en la comparación de RMSSD se obtuvo un p -valor menor a 0.05, evidenciando también diferencias estadísticamente significativas. Lo que indica que tanto la variabilidad global (SDNN) como la variabilidad a corto plazo (RMSSD) [9] difieren significativamente entre individuos con ritmo sinusal normal y aquellos con fibrilación auricular, lo que refleja las alteraciones patológicas en la dinámica cardíaca asociadas a esta arritmia [11].

4.5.2. Comparación de la característica SDNN y RMSSD entre los grupos SR vs SB

Al realizar la prueba de normalidad (Shapiro-Wilk) para los grupos SR y SB tanto en SDNN como en RMSSD, vimos que los datos no siguen una distribución normal, pues tienen un p -valor ≤ 0.05 en ambos casos, mostrando en ambos casos que tienen diferencias estadísticamente significativas entre SR y SB, por lo cual se aplicó la prueba no paramétrica de Mann-Whitney U. Por otro lado, al observar la Figura 4 y 5, vemos que los valores de mediana de SDNN y RMSSD fueron considerablemente mayores en el grupo de bradicardia sinusal (SB) en comparación con el ritmo sinusal normal (SR), lo que nos indica

que tiene una mayor variabilidad de la frecuencia cardíaca en pacientes con bradicardia.

4.5.3. Comparación de la característica SDNN y RMSSD entre los grupos ST y SVT

Para los grupos ST y SVT, también se evaluó la normalidad mediante la prueba de Shapiro-Wilk, confirmando la ausencia de distribución normal en las variables analizadas. En consecuencia, se utilizó nuevamente la prueba de Mann-Whitney U para la comparación ya que en ambos casos se encontró que el p-valor era menor a 0.05. Aunque ST y SVT son ritmos cardíacos rápidos, se encuentra que hay diferencias en la variabilidad de la frecuencia cardíaca.

4.6. ¿Qué características logra mejor discriminación entre grupos?

Ambas variables (SDNN y RMSSD) mostraron diferencias estadísticamente significativas, pues ambas tienen un p-valor bastante pequeño, mucho menor que 0.05 en todas las comparaciones realizadas (SR vs AFIB, SR vs SB, ST vs SVT). Sin embargo, al analizar las distribuciones mediante los gráficos de caja y bigotes Figura 4 y 5, se observó que SDNN presentó una separación más marcada entre los grupos en comparación con RMSSD, lo que indica que esta característica logra mejor discriminación entre grupos.

4.6.1. Algoritmo para calcular el valor de SDNN a partir de una señal de ECG:

1. Detección de picos R: Se identifican los latidos del corazón localizando los máximos de los complejos QRS en la señal de ECG. Debido a que la HVR (Variabilidad de la frecuencia cardíaca), depende de los latidos del corazón, por lo cual es útil detectar la frecuencia en la que se detectan esos picos R
2. Cálculo de intervalos RR: Se mide el tiempo entre cada par de picos consecutivos R.
3. Obtención de la serie de intervalos RR: Construir una lista de todos los intervalos RR obtenidos en el registro. Esta lista representaría la variación latido a latido.
4. Cálculo de la media de intervalos: Para esto, se suman todos los intervalos RR y se dividen entre el número total de intervalos, obteniendo el valor promedio.
5. Cálculo de la desviación estándar de los intervalos RR: Para cada intervalo, se calcula su diferencia respecto a la media, se eleva al cuadrado cada diferencia, se obtiene el promedio de estos cuadrados, y finalmente se extrae la raíz cuadrada del promedio. Esto nos brindará la medida de dispersión de la serie RR, o sea, la SDNN.
6. Reporte del SDNN: El valor final de SDNN se reporta en milisegundos (ms) y refleja la variabilidad global del ritmo cardíaco durante el registro.

5. Conclusiones

- El análisis de los datos permitió identificar patrones de prevalencia de arritmias asociados a edad y género, destacando la fibrilación auricular en hombres mayores. Los resultados fueron consistentes con la literatura, fortaleciendo la validez del conjunto de datos para su uso en modelos de predicción futuros. Este trabajo establece una base sólida para el desarrollo posterior de sistemas de predicción de arritmias, con potencial de expansión hacia técnicas más avanzadas de análisis de series temporales y clustering no supervisado.
- NeuroKit2 fue la herramienta que mejor detectó los latidos en las señales de ECG, funcionando de manera más precisa y confiable que las otras opciones que probamos.
- Entre las características que se analizaron, el SDNN fue la que mejor permitió diferenciar los distintos grupos de arritmias. Esto se vio reflejado en los gráficos, donde las diferencias entre los grupos eran más claras.

Referencias

- [1] Mayo Clinic, “Arritmia cardíaca - Síntomas y causas”, Mayo Clinic, [En línea]. Disponible en: <https://www.mayoclinic.org/es/diseases-conditions/heart-arrhythmia/symptoms-causes/syc-20350668>. [Accedido: 26-abr-2025].
- [2] Mayo Clinic, “Electrocardiograma (ECG o EKG)”, Mayo Clinic, [En línea]. Disponible en: <https://www.mayoclinic.org/es/tests-procedures/ekg/about/pac-20384983>. [Accedido: 26-abr-2025].
- [3] A. González y D. Gutiérrez, “Arritmias cardíacas. Conceptos actuales”, Revista Cubana de Medicina General Integral, vol. 34, no. 4, 2018. [En línea]. Disponible en: http://scielo.sld.cu/scielo.php?pid=S1560-43812018000400014&script=sci_arttext. [Accedido: 26-abr-2025].
- [4] Fundación Española del Corazón, “Arritmias: así nos avisan de que algo falla”, Fundación del Corazón, [En línea]. Disponible en: <https://fundaciondelcorazon.com/blog-impulso-vital/3575-arritmias-asi-nos-avisar-de-que-algo-falla.html>. [Accedido: 26-abr-2025].
- [5] National Heart, Lung, and Blood Institute, “¿Qué causa las arritmias?”, NHLBI, [En línea]. Disponible en: <https://www.nhlbi.nih.gov/es/salud/arritmias/causas>. [Accedido: 26-abr-2025].
- [6] O. Bernal y C. Moro, “Arritmias cardiacas en la mujer”, Revista Española de Cardiología, vol. 59, no. 6, pp. 609–618, jun. 2006, doi: 10.1157/13089748.
- [7] J. Tsuji, R. Venditti, M. Manders, B. Evans, M. Larson, and D. Levy, “Determinants of heart rate variability,” *Circulation*, vol. 93, no. 5, pp. 1043–1050, Mar. 1996. [Online]. Available: <https://doi.org/10.1161/01.CIR.93.5.1043>
- [8] A. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” *Frontiers in Public Health*, vol. 5, p. 258, Sep. 2017. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00258/full>

- [9] F. Shaffer, R. McCraty, and C. L. Zerr, “A healthy heart is not a metronome: An integrative review of the heart’s anatomy and heart rate variability,” *Frontiers in Psychology*, vol. 5, Art. no. 1040, 2014. [Online]. Available: <https://doi.org/10.3389/fpsyg.2014.01040>
- [10] G. D. Clifford and L. Tarassenko, “Quantifying errors in spectral estimates of HRV due to beat replacement and resampling,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 4, pp. 630–638, Apr. 2005. [Online]. Available: <https://doi.org/10.1109/TBME.2005.844028>
- [11] Cleveland Clinic, “Sinus Rhythm,” Cleveland Clinic, Mar. 21, 2022. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/21556-sinus-rhythm>
- [12] Mayo Clinic, “Atrial fibrillation – Symptoms and causes,” Mayo Clinic, Mar. 8, 2024. [Online]. Available: <https://www.mayoclinic.org/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>
- [13] Cleveland Clinic, “Bradycardia: Symptoms, Causes & Treatment,” Cleveland Clinic, 2025. [Online]. Available: <https://my.clevelandclinic.org/health/diseases/17841-bradycardia> Cleveland Clinic
- [14] J. Zheng, C. Rakovski, S. Danioko, J. Zhang, H. Yao, and H. Guo, “A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients,” *figshare, Collection*, Nov. 29, 2019. [Online]. Available: <https://doi.org/10.6084/m9.figshare.c.4560497>
- [15] “Interquartile range,” *Wikipedia*, Apr. 26, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Interquartile_range