

Análisis de señales ECG y detección de arritmias

1st Allisson Rojas Niebles
Universidad de Antioquia
Medellín, Colombia
allisson.rojas@udea.edu.co

2nd Fabio Velez Osorio
Universidad de Antioquia
Medellín, Colombia
fabio.velezo@udea.edu.co

Abstract—aquí va escrito el abstract.

Index Terms—Análisis de señales ECG y detección de arritmia

I. INTRODUCCIÓN

Las arritmias cardíacas son alteraciones que afectan la frecuencia y el ritmo de los latidos del corazón, lo que puede acarrear serias consecuencias para la salud. Por ello, un diagnóstico precoz es esencial. El electrocardiograma (ECG) es una herramienta esencial para la detección de arritmias, ya que permite registrar la actividad eléctrica del corazón en tiempo real. En un ECG tradicional, especialmente en la derivación II, es posible identificar los componentes fundamentales del ciclo cardíaco [1].

- Onda P: Representa la despolarización auricular (contracción de las aurículas).
- Complejo QRS: Denota la despolarización ventricular (contracción de los ventrículos).
- Onda T: Corresponde a la repolarización ventricular (recuperación eléctrica de los ventrículos).

Además, los segmentos PR, ST y TP proporcionan información detallada acerca de los intervalos de conducción y el estado de reposo del corazón. Las distintas arritmias pueden alterar estos patrones normales, dando lugar a registros anormales en el electrocardiograma (ECG). Entre las arritmias más frecuentes se encuentran:

- Fibrilación auricular (AFIB): Un ritmo irregular, caótico y rápido en las aurículas, que conlleva un mayor riesgo de formación de coágulos y accidentes cerebrovasculares.
- Fibrilación Auricular (AF): Es una variante de la fibrilación auricular, donde las aurículas tienen una actividad eléctrica desorganizada, lo que provoca un ritmo cardíaco rápido e irregular. Aunque se usa el mismo término que AFIB, a veces se refiere a la versión menos severa de la arritmia.
- Síndrome de la Aurícula (SA): Es una condición en la que el nodo sinusal, que es responsable de generar los impulsos eléctricos regulares del corazón, no funciona correctamente. Esto puede causar un ritmo cardíaco anormal, ya sea demasiado lento o irregular.
- Taquicardia supraventricular (SVT): Una frecuencia cardíaca elevada que se origina por impulsos eléctricos anormales en o por encima de los ventrículos.
- Taquicardia Sinusal (ST): Es un ritmo cardíaco acelerado, generalmente más de 100 latidos por minuto, que se origina en el nodo sinusal. Puede ser causada por ejercicio,

estrés, fiebre o condiciones médicas, pero también puede ser un signo de disfunción del nodo sinusal.

- Bradicardia sinusal (SB): Una frecuencia cardíaca anormalmente baja, que puede ser fisiológica en deportistas o patológica en otras situaciones.
- Ritmo sinusal irregular (SR): Variabilidad en el intervalo entre latidos, que puede ser normal en jóvenes o estar asociada a condiciones clínicas específicas [2].

Index Terms—

II. METODOLOGÍA

- Para el desarrollo del proyecto se utilizarán dos archivos principales: Diagnostics.xlsx, que contiene información clínica de los pacientes, incluyendo etiquetas de diagnóstico de arritmias, edad, género y características de las señales de ECG que se encuentran en: ECG-DataDenoised.zip, que incluye señales de ECG preprocesadas y filtradas correspondientes a distintos pacientes y ritmos cardíacos. El software requerido ES Python, junto con bibliotecas especializadas como numpy, pandas, scikit-learn, matplotlib, scipy. Para el procesamiento de señales, se emplearán herramientas como la librería NeuroKit y Detectors para el análisis de las señales de arritmias. El entorno de ejecución en Jupyter Notebook.

PROTOCOLO ESTADISTICO :

El presente estudio se desarrolló mediante un protocolo experimental diseñado para asegurar la calidad, representatividad y manejabilidad de los datos analizados. Para ello, se implementó una serie de filtros y procesos de depuración que permitieron construir una muestra robusta y adecuada para los objetivos de la investigación. esta metodología fue hecha en 3 partes:

A. Parte 1

. Inicialmente, se aplicaron criterios de selección basados en la edad y cantidad de arritmias de los sujetos; para el rango de edad se tuvo presente de cumplir 100 sujetos por edad, eso arroja un rango de 44 años a 85 años. Para la cantidad de arritmias se hizo cumplir que estuviera minimo 100 veces en los sujetos, esto nos permitio poder generar un subgrupo con el 76.89% de los datos. Adicionalmente, con el fin de facilitar el almacenamiento, procesamiento y análisis de grandes volúmenes de información, los datos fueron almacenados en un nuevo archivo llamado Diagnosticfiltered.xlsx

los cuales se utilizaron para buscar las señales del nuevo subconjunto en ECGDataDenoised.zip, se convierten las señales al formato .parquet para comprimir y hacer lecturas de datos mas rapidos. Este enfoque metodológico permitió garantizar la eficiencia y reproducibilidad de las etapas analíticas del estudio [3].

B. Parte 2

. Para validar la capacidad de una herramienta de análisis de señales ECG, se seleccionaron 20 registros de ritmo sinusal (SR) y 20 de fibrilación auricular (AFIB) de bases de datos validadas. Se aplicaron dos métodos de detección de picos R: la función `ecg peaks()` de `NeuroKit2` y el detector `Pan-Tompkins` de `py-ecg-detectors`. En fibrilación auricular, ambos métodos reflejaron la irregularidad típica de los intervalos R-R, aunque `NeuroKit2` mostró más detecciones erróneas en registros. En ritmo sinusal, `py-ecg-detectors` presentó un mejor desempeño frente a complejos QRS normales. El análisis evidenció las ventajas y limitaciones de cada herramienta, información útil para validar el sistema propuesto.

C. Parte 3

. Se compararon diferentes características de la variabilidad de la frecuencia cardíaca (HRV), como SDNN, RMSSD obtenidas anteriormente. Para ello, se utilizaron pruebas de hipótesis y se apoyaron en gráficos de estadística descriptiva para visualizar las diferencias entre los grupos. Se evaluó cuál de estas características mostraba mejor capacidad de discriminación entre los grupos, y se describió detalladamente paso a paso utilizado para calcular la característica con mayor poder discriminatorio.

Formulacion de pruebas de hipotesis:

- Hipotesis general H_0 : El valor medio de RMSSD es diferente entre las arritmias de los dos grupos (SR-AF) y (AFIB-SA). H_1 : el valor medio RMSSD de los grupos (SR-AF) y (AFIB-SA) es igual.
- Hipótesis para la Normalidad H_0 : Los datos de RMSSD para los grupos de arritmias (SR-AF) y (AFIB-SA) siguen una distribución normal. H_1 : Los datos de RMSSD para los dos grupos de arritmias (SR-AF) y (AFIB-SA) No siguen una distribución normal.
- Hipótesis para la Homogeneidad H_0 : Las varianzas de RMSSD entre los grupos de arritmias (SR-AF) y (AFIB-SA) son iguales. H_1 : las arritmias (SR-AF) y (AFIB-SA) tienen varianzas de RMSSD diferentes.
- Hipótesis para la prueba de Mann-Whitney H_0 : Las distribuciones de RMSSD entre arritmias (SR-AF) y (AFIB-SA) son iguales. H_1 : Las distribuciones de RMSSD entre las arritmias (SR-AF) y (AFIB-SA) son diferentes

III. RESULTADOS

El presente estudio se llevó a cabo siguiendo un protocolo experimental diseñado para garantizar una mejor calidad, representacion y manejo de los datos analizados. Para alcanzar

este objetivo, se implementó una serie de filtros y procesos de depuración que se detallan de la siguiente manera:

A. Exploración y selección de datos

Analisis estadistico.

Filtrado inicial por género

Se realizó un primer filtrado de los datos para determinar la cantidad de registros de arritmias según el género de los sujetos de estudio. Los resultados obtenidos de este paso se resumen en la Tabla 1.

Rhythm	GenderFEMALE	GenderFEMALE
AF	188	257
AFIB	739	1041
AT	57	64
AVNRT	12	4
AVRT	3	5
SA	176	223
SAAWR	6	1
SB	1408	2481
SR	769	802
ST	1408	799
SVT	308	279

TABLE I: arritmias por genero en el "diagnostic original"

Agrupación por edad Posteriormente, se aplicó un segundo filtro basado en la edad de los individuos, agrupándolos en bloques de al menos 100 sujetos por grupo. Esta estrategia tuvo como finalidad garantizar una distribución más homogénea y representativa de la población analizada. Además se efectuó un filtrado específico basado en el tipo de arritmia registrada. La combinación de los filtros de edad y diagnóstico permitió depurar la muestra, resultando en la conservación del 76,89% del total de datos iniciales, tal como se muestra en la Tabla 2.

Rhythm	Numero
SB	3317
AFIB	1548
SR	1357
ST	1066
SVT	396
AF	370
SA	132

TABLE II: cantidad de pacientes tras ambos filtros

Análisis por rangos de edad: Se realizó un análisis detallado de los datos segmentados por rangos de edad, identificando que el intervalo de edad entre los 44 y 85 años concentraba la mayor proporción de diagnósticos de arritmia. A partir de este análisis se definió un nuevo conjunto de datos, denominado "diagnostic filtered".

Optimización del almacenamiento de datos: Para mejorar la eficiencia en el almacenamiento y procesamiento, cada archivo de datos fue convertido al formato .parquet, un esquema orientado a columnas que facilita el acceso y manejo de grandes volúmenes de información.

Organización y compresión de archivos: Finalmente, los archivos fueron comprimidos y organizados en grupos según el

Arritmia	conteo	mean	std	min	max%
AF	370.0	19.659459	5.506609	6.0	43
AFIB	1548.0	16.608527	4.712467	3.0	37
SA	264.0	13.863636	3.955595	7.0	27
SB	3317.0	12.833886	6.051684	1.0	36
ST	1066.0	18.344278	2.404739	3.0	43
SVT	374.0	26.877005	3.946812	11.0	48.0

TABLE III: Caption

tipo de arritmia. Esta organización tuvo como objetivo facilitar el análisis detallado de las señales electrocardiográficas en las fases siguientes del estudio.

- señal agrupacion archivo del tipo parquet.

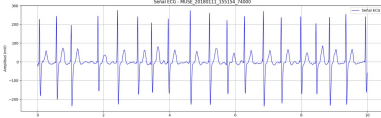


Fig. 1: señal de verificación de que cada archivo.parquet

B. Resultados del análisis para SR y AFIB

Se seleccionaron dos tipos de arritmias, SR y Fibrilación auricular (AFIB), con 20 registros de ECG para cada una, asegurando patrones característicos según la literatura clínica. Se detectaron los picos R usando dos métodos: NeuroKit2 y py-ecg-detectors. En fibrilación auricular, ambos métodos capturaron la irregularidad esperada, aunque NeuroKit2 mostró mayor susceptibilidad al ruido. En taquicardia ventricular, py-ecg-detectors tuvo mejor desempeño adaptándose a los complejos QRS anchos y rápidos. El análisis de los 40 registros evidenció fortalezas y limitaciones de ambas librerías para detectar patrones anormales en ECG.(tabla III)(tabla IV)

nombres	picos _{prom}	tiempo _{prom}
neurokit	12.05	0.000685
neurokit _{phantomkins} 1908	12.25	0.000896
PyECG _{phantomkins}	12.80	0.001718
PyECG _{hamilton} 2002	12.95	0.002534
neurokit _{hamilton}	13.25	0.003366
PyECG _{engzee}	11.10	0.019482
PyECG _{christov}	14.30	0.049152

TABLE IV: analisis diferentes librerias de neurokit y detectors para SR

Entre los métodos evaluados, el algoritmo PyECG christov de la librería Detectors mostró un desempeño destacado. En promedio, este método detectó 14.3 picos R por registro, con un tiempo promedio de procesamiento de 0.0492 segundos.

C. Variabilidad de la frecuencia cardiaca (HRV)

- La variabilidad de los datos

Nombres	picos _{prom}	tiempo _{prom}
neurokit _{phantomkins} 1908	18.30	0.001673
PyECG _{phantomkins}	17.95	0.001802
PyECG _{hamilton} 2002	18.00	0.002164
neurokit _{hamilton}	18.30	0.002295
PyECG _{engzee}	16.20	0.015722
PyECG _{christov}	18.55	0.046595

TABLE V: Analisis diferentes librerias de neurokit y detectors para AFIB

arritmia	mean _{r-r} (mean)	std	sdn _{r-r} s(mean)	std std	rmssd (mean)
AF	545.010270	182.664551	114.212351	168.941372	143.694333
AFIB	641.560988	228.996195	173.494927	155.610836	230.087483
SA	757.359982	182.578694	191.932466	108.859979	214.213500
SB	890.764484	266.922987	239.756209	135.566420	240.952445
SR	744.804463	137.367183	119.032044	101.721855	133.704138
ST	540.497891	62.240810	37.570927	80.519124	48.611059
SVT	371.603386	62.431206	21.754522	96.933833	30.376577

TABLE VI: Resumen de metricas HRV por arritmia

D. Variabilidad de la frecuencia cardiaca (HRV)

Se compararon diferentes características de la variabilidad de la frecuencia cardíaca (HRV), como SDNN, RMSSD. Se evaluó cuál de estas características mostraba mejor capacidad de discriminación entre los grupos.(figura 3)

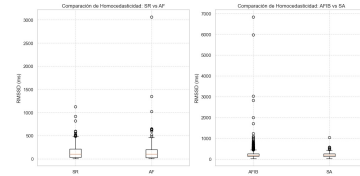


Fig. 4: Comparacion

IV. DISCUSIÓN

En este estudio, se llevó a cabo un filtrado de los datos clínicos con el objetivo de analizar la prevalencia de arritmias en función del género y la edad. Se garantizó una muestra representativa al agrupar a los pacientes en intervalos de edad, asegurando que cada grupo contara con un mínimo de 100 individuos. Se definió un conjunto de datos "diagnostic filtered", centrado en el rango de edades entre 44 a 85 años, que abarca el 76,89% de los casos. Los datos fueron convertidos a formato .parquet y organizados por tipo de arritmia para optimizar su manejo [2]. El análisis reveló que la bradicardia sinusal es la arritmia más común, con una mayor prevalencia en hombres y en pacientes mayores de 60 años, lo cual coincide con la información disponible en la literatura médica. Para el entrenamiento de algoritmos de aprendizaje automático, se sugiere utilizar datos de pacientes con edades comprendidas entre 40 y 80 años, centrándose en las arritmias más prevalentes. Por último, se propone profundizar el análisis mediante el estudio de la variabilidad del intervalo RR y la aplicación de técnicas de reducción de dimensionalidad, como el PCA, con el fin de mejorar la caracterización de los datos [3].

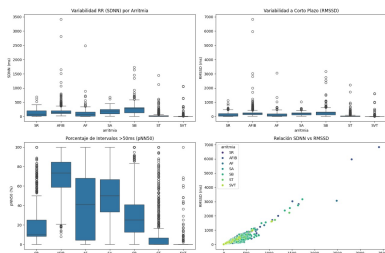


Fig. 2: Variabilidad

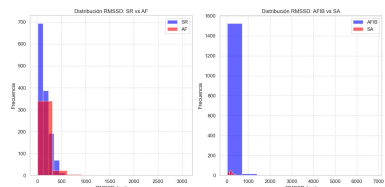


Fig. 3: Distribucion

Entre los métodos evaluados, el algoritmo PyECG christov de la librería Detectors mostró un desempeño destacado. Esta eficiencia en la detección y procesamiento se traduce en un rendimiento superior frente a los otros algoritmos evaluados. La ventaja de PyECG christov radica en su alta precisión y velocidad, lo que es particularmente importante para aplicaciones en tiempo real o en escenarios donde el procesamiento rápido y confiable de señales ECG es crítico. A diferencia de otros métodos que presentaron mayor sensibilidad al ruido o errores en la identificación de picos, PyECG christov logró mantener una detección consistente y robusta, incluso en registros con irregularidades características de arritmias como AFIB. Por estas razones, PyECG christov se perfila como el método más adecuado para integrarse en la herramienta de análisis [2].

Dado que los datos de RMSSD no se distribuyeron de manera normal, las pruebas de normalidad, como la de Shapiro Wilk, rechazaron la hipótesis nula que sostiene que estos datos seguían una distribución normal. En este escenario, se utilizó pruebas no paramétricas para analizar las diferencias entre grupos, tales como la prueba de Mann Whitney, que no requiere que se cumpla la suposición de normalidad. Por lo tanto, la hipótesis nula para la prueba de Mann Whitney se mantuvo bajo la suposición de que las distribuciones de RMSSD no eran normales, permitiendo la comparación de las medianas entre los grupos. Como resultado, para evaluar la homogeneidad de varianzas (homocedasticidad), si la prueba de Levene se aplicó y mostró que las varianzas eran significativamente diferentes, la hipótesis nula fue rechazada, lo que implicó que se debieron usar métodos estadísticos que no supusieran homogeneidad de varianzas. Este enfoque permitió ajustar el análisis a las características no normales de los datos.

V. CONCLUSIONES

- El algoritmo PyECG Christov se ha consolidado como el más eficiente y preciso en la detección y el procesamiento de señales de ECG, superando a otros métodos evaluados.

Su gran precisión y rápida capacidad de procesamiento lo posicionan como una opción ideal para aplicaciones en tiempo real, donde la confiabilidad en la detección de arritmias, como la fibrilación auricular (AFIB), resulta fundamental. Además, su rendimiento robusto, incluso ante la presencia de ruido o irregularidades en las señales.

REFERENCES

- [1] E. Pérez, J. Higuera, J. Gómez, and P. Martínez, *Casos ECG de concurso eCardio 2017*. Sociedad Española de Cardiología, 2017. [Online]. Available: <https://secardiologia.es/images/publicaciones/libros/casos-ecg-de-concurso-ecardio-2017.pdf>
- [2] A. Varios, *Electrocardiografía: Manual para la práctica de las arritmias*. Editorial de la Universidad Nacional de La Pampa, 2020. [Online]. Available: <https://www.unlpam.edu.ar/cultura-y-extension/edunlpam/catalogo/libros-de-textos-para-estudiantes-universitarios/electrocardiografia-manual>
- [3] S. Hurtado Requena, "Diseño y montaje de un simulador de ecg," Master's thesis, Universidad de Málaga, 2018. [Online]. Available: <https://riuma.uma.es/xmlui/bitstream/handle/10630/15473/SandrohurtadorequenaMemoria.pdf?sequence=1>