

# Análisis estadístico de señales

1<sup>ro</sup> Federico Ocampo Quiceno  
Universidad de Antioquia  
Medellín, Colombia  
federico.ocampoq@udea.edu.co

2<sup>do</sup> Daniel Palacio Alvarez  
Universidad de Antioquia  
Medellín, Colombia  
daniel.palacioa@udea.edu.co

3<sup>do</sup> Alejandro Agudelo Tamayo  
Universidad de Antioquia  
Medellín, Colombia  
alejandro.agudelot@udea.edu.co

**Abstract**—This laboratory aimed to apply basic statistical analysis techniques to biomedical signals, specifically ECG and EEG, in the time domain. Using Python and libraries such as NumPy, Pandas, SciPy, and Matplotlib, key metrics including mean, variance, standard deviation, and RMS were computed. ECG signals were analyzed both before and after filtering to observe the effect of noise reduction, and individual cardiac cycles were studied to assess variability and stationarity using tests like Mann-Whitney and the Augmented Dickey-Fuller (ADF). Although variability was observed across individual cycles, the ADF test suggested the overall signal could be considered stationary.

In the second part, EEG signals from control subjects and Parkinson's patients were compared through average energy per channel. Due to a lack of normality in the data, non-parametric tests were used. No statistically significant differences were found across most channels, though the sample size and signal variability may have influenced the outcomes. The results underline the importance of combining statistical tools and visualization techniques to effectively interpret biomedical signals and reveal subtle patterns that may not be immediately apparent.

## I. INTRODUCCIÓN

El análisis de señales biomédicas representa una herramienta fundamental en el estudio y monitoreo de procesos fisiológicos del cuerpo humano. Estas señales, como el ECG (electrocardiograma) y EEG (electroencefalograma), presentan características propias que las hacen altamente variables y dependientes del contexto fisiológico del individuo. Debido a esta variabilidad, muchas de estas señales son consideradas estocásticas, es decir, presentan un comportamiento aleatorio que requiere de técnicas estadísticas para su adecuado análisis e interpretación.

En este laboratorio se abordarán técnicas básicas de análisis estadístico de señales biomédicas en el dominio del tiempo, con el fin de comprender sus características fundamentales y extraer información significativa de ellas. A través del uso de herramientas computacionales en Python, se implementarán funciones estadísticas como el cálculo de la media, varianza, desviación estándar y el valor cuadrático medio (RMS). También se realizarán pruebas estadísticas, tanto paramétricas como no paramétricas, con el objetivo de comparar diferentes segmentos de señales y evaluar su estacionariedad.

## II. METODOLOGÍA

Para la realización de esta práctica se empleó el paquete computacional Python, utilizando principalmente bibliotecas

como NumPy, Pandas, SciPy y Matplotlib, las cuales permitieron realizar el procesamiento, análisis y visualización de señales biomédicas.

Se trabajó con dos conjuntos de datos diferentes, según el objetivo de cada parte del estudio. En la primera parte, enfocada en el análisis estadístico de señales ECG, se utilizó el archivo *signals.mat*, que contenía registros de señales adquiridas a una frecuencia de muestreo de 1024 Hz. Estas señales estaban organizadas en dos campos: *asRecording*, correspondiente a la señal original, y *filtered*, que contenía la señal procesada mediante técnicas de filtrado.

En la segunda parte, centrada en el análisis comparativo de energía en señales EEG, se utilizó un conjunto de datos con registros de EEG de múltiples canales correspondientes a dos grupos poblacionales: un grupo control compuesto por sujetos sanos y un grupo de pacientes diagnosticados con enfermedad de Parkinson.

El procedimiento experimental se estructuró en dos bloques principales:

### 1) Análisis estadístico de señales ECG:

- Se compararon las señales sin filtrar y filtradas para observar los efectos del filtrado, considerando las frecuencias características del ECG.
- Se implementaron funciones en Python para calcular métricas estadísticas como la media, varianza, desviación estándar y el valor cuadrático medio (RMS) de ciclos cardíacos extraídos manualmente.
- Se realizaron pruebas estadísticas entre los ciclos utilizando pruebas para evaluar los supuestos necesarios para una *prueba t* y, en caso de que los supuestos de aplicación no se cumplieran, se recurrió a la prueba de Mann-Whitney.
- Se comprobó el resultado mediante la prueba Dickey-Fuller.

### 2) Comparación estadística de energía en señales EEG:

- Se diseñó una función que permitió calcular la energía promedio por canal a partir de múltiples épocas de cada sujeto.
- Los valores de energía fueron organizados en DataFrames separados para cada grupo poblacional, con canales como columnas y sujetos como filas.
- Se realizaron pruebas estadísticas entre los grupos utilizando pruebas para evaluar los supuestos necesarios para una *prueba t* y, en caso de que

los supuestos de aplicación no se cumplieran, se recurrió a la prueba de Mann-Whitney.

La metodología empleada permitió caracterizar y comparar diferentes segmentos de señales biomédicas tanto desde el punto de vista gráfico como estadístico. La elección de este enfoque se justificó por la naturaleza estocástica de las señales y la necesidad de utilizar herramientas estadísticas robustas para su correcta interpretación.

### III. RESULTADOS Y DISCUSIÓN

#### A. Análisis estadístico de señales ECG

Las señales analizadas fueron adquiridas a una tasa de muestreo de 1024 Hz. En los archivos entregados. En este se registran las diferentes señales que fueron adquiridas y luego filtradas para su posterior análisis. Primero se tomó la señal original y se comparó con la señal filtrada en la Figura 1

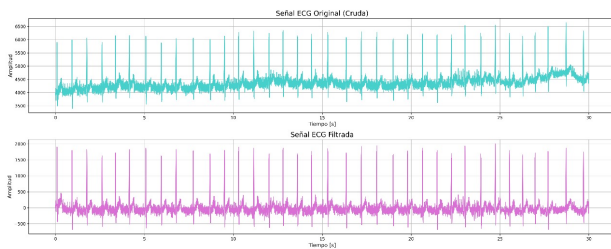


Fig. 1: Gráfica de la señal ECG original y filtrada

La señal original presenta un elevado nivel de ruido, evidenciado por oscilaciones adicionales a las propias del latido cardíaco, así como un desplazamiento de base centrado en torno a los 4000. Este ruido podría estar relacionado con el ruido electromiográfico (EMG), que se genera a partir de la actividad muscular y se superpone a las señales del ECG. Según un estudio publicado por MDPI, el ruido muscular suele encontrarse en el rango de 20 a 500 Hz, lo que puede interferir con la correcta identificación de los picos del ECG [1]. De igual manera se pudo utilizar un filtro notch a 60 o 50 Hz para eliminar el ruido eléctrico.

Tras el proceso de filtrado, la señal muestra una notable reducción del ruido y se encuentra centrada en cero, lo que indica que el filtro aplicado ha corregido el desplazamiento de base. Dado que las frecuencias de interés para el análisis de ECG se sitúan entre 0.5 a 100 Hz [2], es probable que se haya utilizado un filtro pasa-bajo o pasa-banda, diseñado específicamente para conservar las frecuencias asociadas a la actividad cardíaca, mientras atenúa tanto el ruido de alta frecuencia como el offset. Los picos correspondientes a los latidos del corazón se mantienen presentes en la señal filtrada, lo que sugiere que el filtrado ha sido efectivo para eliminar componentes no deseados sin comprometer la información relevante del ECG.

Con la señal ya procesada, se procedió a detectar los picos R, los cuales representan los eventos de despolarización ventricular y constituyen el componente más prominente del ECG. (Figura 2)

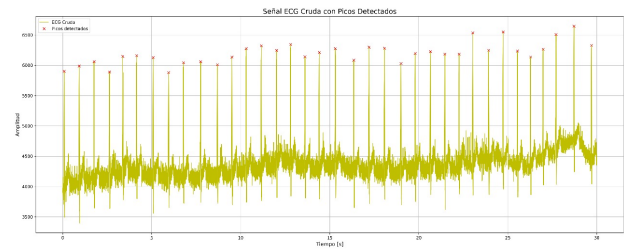


Fig. 2: Gráfica de señal cruda con picos R

Para este paso se utilizó la función `find_peaks` de la librería `scipy`, ajustando cuidadosamente parámetros como la altura mínima y la distancia entre picos, para evitar tanto doble marcación pico como la omisión de latidos (Figura 2). Este dato permitió deducir que la frecuencia cardíaca del paciente era de aproximadamente 72bpm lo cual sugiere un estado de reposo a la hora de ser tomada la señal.

La correcta identificación de los picos en el ECG permite separar la señal por ciclos.

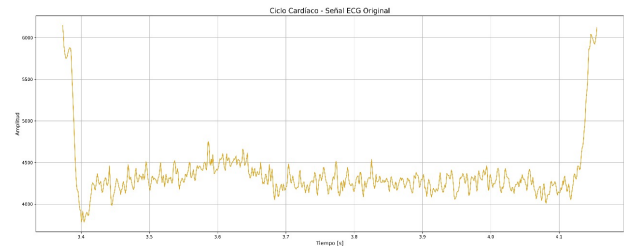


Fig. 3: Gráfica de ciclo cardíaco original

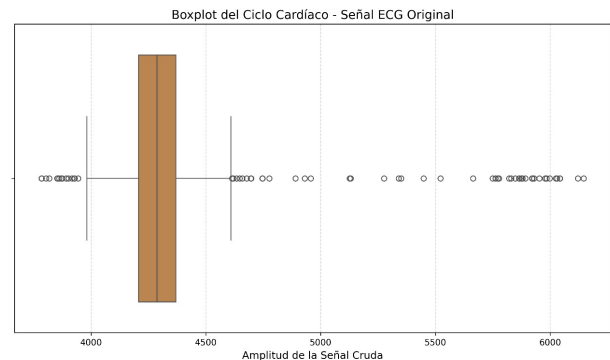


Fig. 4: Gráfico de cajas y bigotes de señal original

La mayor parte de los valores de amplitud están agrupados alrededor del promedio 4131. Esto sugiere que hay un desplazamiento de base o offset en la señal, ya que no está centrada en cero. La proximidad entre el promedio 4352 y el valor RMS 4365 indica que la señal está relativamente estable, con pequeñas fluctuaciones en torno al valor promedio. La alta varianza 116036.37 refleja la dispersión significativa en la señal, influenciada por los picos altos de los latidos (complejos QRS en un ECG típico) y las zonas más bajas (segmentos PR y ST entre los latidos). Este comportamiento es común

en las señales de ECG debido a la naturaleza de los ciclos cardíacos [3]. La desviación estándar de 340.64 indica una gran variabilidad, lo cual es coherente con la visualización del boxplot, que muestra varios valores atípicos. Esto también sugiere fluctuaciones notables en la señal, causadas por los picos relacionados con los latidos y posiblemente influenciados por ruido o artefactos. Luego se realizan los mismos gráficos para la señal filtrada.

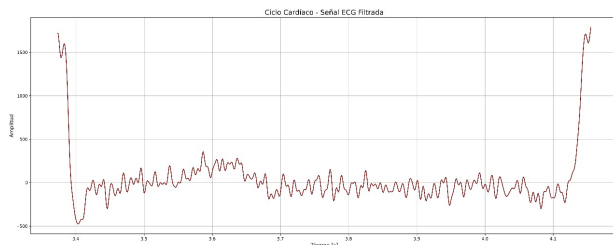


Fig. 5: Gráfica de ciclo cardíaco original

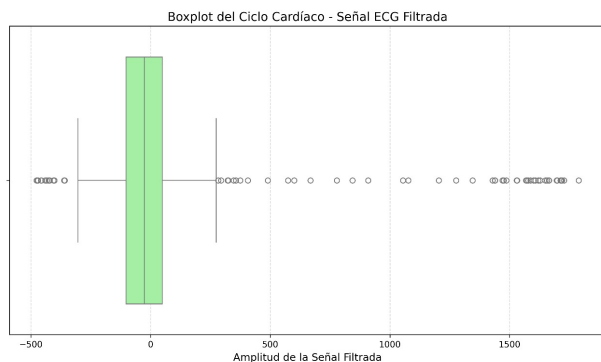


Fig. 6: Gráfico de cajas y bigotes de señal original

El promedio ahora es 40.8, lo que muestra que el filtrado ha corregido ese desplazamiento y ha centrado la señal en torno a cero.

El RMS en la señal filtrada es 341, lo que indica una reducción significativa en la variabilidad de la señal. Esto se debe a que el proceso de filtrado ha eliminado componentes no deseados (como el ruido o las variaciones de baja frecuencia), lo que resulta en una señal más "limpia" y con menos fluctuaciones abruptas no relacionadas con el comportamiento fisiológico de los latidos del corazón. Esto puede indicar que la señal ahora está más concentrada en torno a valores de amplitud más pequeños como se puede ver en la Figura 5 donde el pico R se da en el 1700 y sin filtrar estaba en 3500, además, que las fluctuaciones más extremas (incluyendo señal eléctrica) han sido suavizadas.

La pequeña diferencia en la varianza entre la señal original y la filtrada (114850.02) puede deberse a que, aunque ha cambiado la señal en términos de amplitud y offset, se han conservado los datos más característicos de la señal de ECG, como es el complejo PQR donde se dan las ondas más predominantes. Aunque en general se esperaba una reducción

en la varianza, este valor sugiere que la filtración permitió la conservación de los eventos significativos.

Después del filtrado, la desviación estándar es 338.89, un valor ligeramente inferior. Esto indica que, si bien el filtrado ha eliminado ciertos componentes de baja y alta frecuencia no deseados, la señal aún conserva variaciones significativas relacionadas con los eventos fisiológicos importantes, como los picos de los latidos del corazón. Estos picos representan variaciones bruscas en la amplitud, que podrían haber contribuido al aumento en la desviación estándar.

Se grafican 15 ciclos para su análisis.

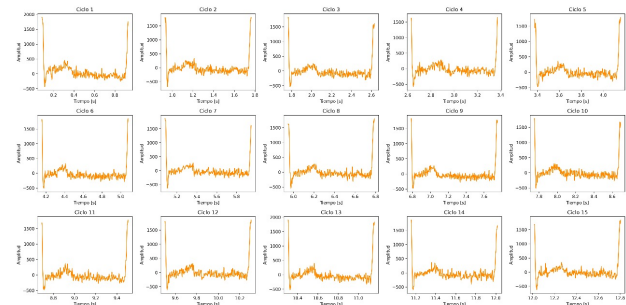


Fig. 7: Gráfico de los 15 ciclos en ECG

Los 15 ciclos muestran una estructura muy similar, sin embargo, las fluctuaciones entre los picos varían en amplitud, forma en cada ciclo y tiempo.

Algunos ciclos tienen mayores oscilaciones, mientras que otros son más estables (con amplitudes más bajas), además entre pico y pico difieren en el tiempo. Estas diferencias visuales entre los ciclos sugiere que la señal no es constante y debido al análisis en tiempo parece que es no estacionaria.

Los promedios de los ciclos varían considerablemente. Van desde valores negativos hasta valores positivos. Estos cambios en el promedio indican que la señal no se mantiene estable a lo largo de todos los ciclos.

La varianza, que mide la dispersión de los valores de cada ciclo, también muestra una alta variabilidad. Algunas varianzas son mucho mayores que otras, por ejemplo, el ciclo 5 tiene una varianza de 114850.0201, mientras que el ciclo 4 tiene una varianza mucho menor de 46839.0197. Esto sugiere que algunos ciclos tienen fluctuaciones mucho más extremas que otros, lo que contribuye a que la señal sea menos estable.

No se puede considerar que la señal es estacionaria debido a la variabilidad en las estadísticas de cada ciclo y la diferencia en el comportamiento gráfico entre ciclos. Tanto los gráficos como los valores del promedio y la varianza demuestran que la señal cambia con el tiempo, lo que es indicativo de una señal no estacionaria.

Para determinar si la señal es estacionaria o no estacionaria se realizan pruebas estadísticas. Primero se realiza una prueba de Kolmogorov-Smirnov para conocer la normalidad de 2 ciclos y luego compararlos.

Se eligen los ciclos 2 y 5. Para ambos ciclos, los datos no parecen seguir una distribución normal, ya que se rechazó la hipótesis nula ( $p = 0.000$  para ambos ciclos). Esto implica que

la prueba t de Student no es adecuada para comparar estas muestras, lo que justifica el uso de la prueba U de Mann-Whitney.

Luego se realiza una prueba de homocedasticidad. La prueba de Levene mostró que las varianzas entre los ciclos son iguales ( $p = 0.436$ ). Aunque los datos no son normales, esta homogeneidad de varianza permite realizar comparaciones entre las muestras.

Por último se asume que los grupos son independientes. Estos mismos 3 análisis se realizan sobre otros 2 ciclos diferentes (ciclo 13 y 15) y se llegaron a las mismas conclusiones que con los primeros ciclos analizados con valores ligeramente diferentes.

Se aplicó la prueba de Dickey-Fuller Aumentada (ADF) para evaluar la estacionariedad de la serie temporal. El estadístico resultó considerablemente más negativo que los valores críticos, y el p-valor fue inferior a 0.05, lo que permite rechazar la hipótesis nula con un alto grado de confianza. Esto indica que la serie es estacionaria en términos globales.

Este hallazgo contrasta con los análisis exploratorios por ciclos, en los que se observaron variaciones en media, varianza y forma, sugiriendo no estacionariedad local. La discrepancia se explica por el enfoque de la prueba ADF, que evalúa la estructura global de la serie. Además, el filtrado previo probablemente eliminó componentes de baja y alta frecuencia, estabilizando las propiedades estadísticas de la señal. En conjunto, los resultados sugieren que, además de la homogeneidad observada a nivel de ciclos individuales, la serie temporal presenta propiedades estadísticas estables cuando se analiza en su totalidad, lo que respalda su caracterización como estacionaria bajo los supuestos del test ADF.

### B. Comparación estadística de energía en señales EEG

Se desarrolló una función que recibe una señal con la forma *shape(canales, épocas, núm. de puntos)* y es capaz de devolver la energía promedio de cada canal como se muestra en la Tabla I.

Tabla I: Energía promedio de cada canal del sujeto 2

Canal	Energía Promedio
Canal 1	15966.40
Canal 2	17617.81
Canal 3	20804.94
Canal 4	19654.40
Canal 5	16678.98
Canal 6	93894.05
Canal 7	66862.50
Canal 8	75685.13

Ahora se realizó lo mismo pero tomando todos los sujetos asociados a cada grupo poblacional obteniendo los resultados mostrados en la Tabla II y Tabla III

Tabla II: Energía promedio del grupo control

	Canal <sub>1</sub>	Canal <sub>2</sub>	Canal <sub>3</sub>	...
Sujeto 1	21465.65	20985.91	22760.15	...
Sujeto 2	15966.40	17617.81	20804.94	...
Sujeto 3	14148.67	18284.00	28749.93	...
Sujeto 4	35311.30	34916.69	38800.43	...
Sujeto 5	18510.83	19738.49	20911.79	...
Sujeto 6	13180.11	13925.22	16218.99	...
Sujeto 7	11197.55	10948.37	12737.00	...
Sujeto 8	28551.12	26204.84	17384.00	...
Sujeto 9	9133.04	9214.16	11626.41	...
Sujeto 10	47166.56	55107.80	52286.88	...
Sujeto ...	...	...	...	...

Tabla III: Energía promedio del grupo parkinson

	Canal <sub>1</sub>	Canal <sub>2</sub>	Canal <sub>3</sub>	...
Sujeto 1	12438.24	11261.18	10819.63	...
Sujeto 2	17995.66	12001.60	12286.34	...
Sujeto 3	38092.10	43575.38	41979.99	...
Sujeto 4	23742.33	22070.01	24540.32	...
Sujeto 5	48574.52	51806.53	73171.95	...
Sujeto 6	16202.42	13124.25	13988.67	...
Sujeto 7	10692.95	10841.19	12154.39	...
Sujeto 8	12157.23	13398.66	17668.88	...
Sujeto 9	9581.81	14008.57	9589.23	...
Sujeto 10	23658.74	23990.26	30633.75	...
Sujeto ...	...	...	...	...

Con el objetivo de evaluar la homogeneidad de varianzas entre los grupos poblacionales, se aplicó la prueba de Levene como se evidencia en la Tabla VI. Los resultados indicaron que no se encontró evidencia suficiente para rechazar la hipótesis nula, lo que sugiere que las varianzas de los canales EEG entre los grupos control y con enfermedad de Parkinson pueden considerarse aproximadamente iguales. No obstante, al detectarse una violación del supuesto de normalidad mediante la prueba de Kolmogorov-Smirnov como se visualiza en la Tabla IV y Tabla V, se decidió continuar el análisis utilizando métodos no paramétricos, más adecuados bajo estas condiciones.

Tabla IV: Prueba de normalidad para el grupo control

Canal	p-value	Decisión
Canal <sub>1</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>2</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>3</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>4</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>5</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>6</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>7</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>8</sub>	0.0000	Se rechaza H <sub>0</sub>

Tabla V: Prueba de normalidad para el grupo parkinson

Canal	p-value	Decisión
Canal <sub>1</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>2</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>3</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>4</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>5</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>6</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>7</sub>	0.0000	Se rechaza H <sub>0</sub>
Canal <sub>8</sub>	0.0000	Se rechaza H <sub>0</sub>

Tabla VI: Prueba de Levene

Canal	p-value	Decisión
Canal <sub>1</sub>	0.8848	No se rechaza H <sub>0</sub>
Canal <sub>2</sub>	0.9591	No se rechaza H <sub>0</sub>
Canal <sub>3</sub>	0.9776	No se rechaza H <sub>0</sub>
Canal <sub>4</sub>	0.7593	No se rechaza H <sub>0</sub>
Canal <sub>5</sub>	0.9046	No se rechaza H <sub>0</sub>
Canal <sub>6</sub>	0.4720	No se rechaza H <sub>0</sub>
Canal <sub>7</sub>	0.6179	No se rechaza H <sub>0</sub>
Canal <sub>8</sub>	0.5941	No se rechaza H <sub>0</sub>

En particular, se empleó la prueba de Mann-Whitney U para comparar las distribuciones de las señales entre ambos grupos. En varios canales analizados, no se rechazó la hipótesis nula, lo cual indica la ausencia de diferencias estadísticamente significativas en esos casos. Este resultado sugiere que, al menos para dichos canales, las características de las señales EEG podrían presentar comportamientos similares entre los grupos.

Tabla VII: Prueba U de Mann-Whitney

Canal	p-value	Decisión
Canal <sub>1</sub>	0.4057	No se rechaza H <sub>0</sub>
Canal <sub>2</sub>	0.5705	No se rechaza H <sub>0</sub>
Canal <sub>3</sub>	0.4604	No se rechaza H <sub>0</sub>
Canal <sub>4</sub>	0.2345	No se rechaza H <sub>0</sub>
Canal <sub>5</sub>	0.5600	No se rechaza H <sub>0</sub>
Canal <sub>6</sub>	0.2801	No se rechaza H <sub>0</sub>
Canal <sub>7</sub>	0.1183	No se rechaza H <sub>0</sub>
Canal <sub>8</sub>	0.1505	No se rechaza H <sub>0</sub>

Sin embargo, es fundamental considerar que la falta de significancia estadística no implica necesariamente la inexistencia de diferencias reales. Factores como el tamaño muestral reducido, la alta variabilidad inherente a las señales EEG, así como la presencia de ruido o artefactos fisiológicos y ambientales, pueden limitar la sensibilidad de las pruebas estadísticas utilizadas. Adicionalmente, variables como el estado emocional, el nivel de atención de los participantes, o condiciones externas al protocolo experimental podrían haber influido en los registros, comprometiendo así la potencia del estudio y la capacidad para detectar efectos sutiles pero relevantes.

#### IV. CONCLUSIONES

- Los resultados muestran que existen diferencias significativas entre los ciclos de ECG, lo que refleja variabilidad en características como la amplitud y la forma de los picos, posiblemente por factores fisiológicos o condiciones durante la medición. Sin embargo, al analizar la señal completa, se observó un comportamiento estacionario, probablemente por el efecto de promediar los datos. Lo que resalta la importancia de analizar tanto los ciclos individuales como la señal global para no perder información relevante sobre la variabilidad de la señal.
- En el análisis de energía promedio de señales EEG, aunque no se hallaron diferencias significativas en varios canales, no se descarta la presencia de cambios sutiles. El uso de pruebas no paramétricas fue clave ante la falta de normalidad en los datos. Sin embargo, el tamaño muestral y la variabilidad de las señales limitan los resultados, esto puede ser verificado mediante otras características o métodos para obtener conclusiones más claras.
- El análisis estadístico aplicado a las señales permitió comprender mejor su comportamiento y variabilidad en el tiempo. A partir de este enfoque, fue posible identificar diferencias y patrones que no serían evidentes solo mediante la observación gráfica.

#### REFERENCES

- [1] N.-T. Bui and G.-s. Byun, "The comparison features of ecg signal with different sampling frequencies and filter methods for real-time measurement," *Symmetry*, vol. 13, no. 8, 2021. [Online]. Available: <https://www.mdpi.com/2073-8994/13/8/1461>
- [2] L. G. Tereshchenko and M. E. Josephson, "Frequency content and characteristics of ventricular conduction," *Journal of Electrocardiology*, vol. 48, no. 6, pp. 933–937, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022073615002915>
- [3] Python Software Foundation, "Módulo statistics — python 3.9.17 documentación," 2021, consultado el 14 de abril de 2025. [Online]. Available: <https://docs.python.org/es/3.9/library/statistics.html>