



**Desarrollo de aplicación de servicios web basado en estándares de informática médica para el preprocesamiento y visualización de registros EEG**

Luisa María Zapata Saldarriaga

Trabajo de grado presentado para optar al título de Bioingeniero

Asesor

John Fredy Ochoa Gómez, Doctor (PhD) en Ingeniería Electrónica

Línea de investigación:

Neuroingeniería

Grupo de investigación:

Grupo de Neuropsicología y Conducta

Universidad de Antioquia

Facultad de Ingeniería

Bioingeniería

Medellín, Antioquia, Colombia

2022

**Referencia**

- [1] L. M. Zapata Saldarriaga, “Desarrollo de aplicación de servicios web basado en estándares de informática médica para el preprocesamiento y visualización de registros EEG”, Trabajo de grado profesional, Bioingeniería, Universidad de Antioquia, Medellín, Antioquia, Colombia, 2022.

Estilo IEEE (2020)



Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** John Jairo Arboleda Céspedes.

**Decano/Director:** Jesús Francisco Vargas Bonilla.

**Jefe departamento:** John Fredy Ochoa Gómez.

## **Dedicatoria**

Metamorfosis.

El proceso de florecer y marchitar, de crecer y levantar, de sonreír y llorar.

¡Papá, abuela, gracias por formar la mujer que hoy en día soy!

Hércules y Fabiola.

## **Agradecimientos**

Alma mater gracias por permitirme formar, crecer y sobre todo ser. Gracias a mi familia, en especial a mi madre, por enseñarme a perseverar ante los momentos más difíciles de la vida y nunca dejarme caer. Gracias a mis amigos, mi otra familia, por enseñarme el valor de la amistad, por siempre apoyarme y enseñarme tantas cosas bonitas de la vida. Gracias a mis maestros, en especial a John Fredy Ochoa Gómez y a Jonathan Gallego Londoño, por mostrarme el amor a la academia, por enseñarme el valor de la humildad y por siempre instruirme a ser cada día mejor.

Gracias e infinitas gracias al grupo LDJY por ser tan incondicionales en la vida.

¡Gracias a todos los que pasaron por este pedacito de la vida!

## TABLA DE CONTENIDO

RESUMEN.....	11
ABSTRACT .....	12
I. INTRODUCCIÓN .....	13
II. OBJETIVOS.....	16
A. Objetivo general .....	16
B. Objetivos específicos .....	16
III. MARCO TEÓRICO .....	17
Electroencefalografía .....	17
Estándar BIDS .....	19
Procesamiento de registros EEG .....	21
Control de calidad de los datos .....	22
Análisis de estadística descriptiva.....	23
Mecanismo arquitectónico de trabajo .....	24
Backend.....	24
Frontend .....	25
Gestión de servicios .....	26
Usabilidad.....	27
IV. METODOLOGÍA .....	29
A. Datos.....	29
B. Tecnologías y herramientas.....	32
C. Control de calidad de los datasets .....	34
V. RESULTADOS Y DISCUSIÓN.....	37
A. Estandarización de los datos .....	37
B. Procesamiento de los registros EEG .....	39

C. Control de calidad de datos .....	43
Preprocesamiento .....	44
1. Análisis global.....	45
2. Análisis entre estudios.....	47
3. Análisis entre grupos .....	49
4. Análisis entre sesiones .....	51
Post-procesamiento .....	54
1. Análisis global.....	54
2. Análisis ente estudios .....	56
3. Análisis entre grupos .....	59
4. Análisis entre sesiones .....	61
D. Desarrollo de interfaces.....	65
E. Gestión de servicios Linux .....	78
F. Prueba de usabilidad.....	79
CONCLUSIONES .....	83
REFERENCIAS .....	85
ANEXOS.....	89

## LISTA DE TABLAS

Tabla I. Valoración para la prueba de usabilidad.....	28
Tabla II. Información genérica de las bases de datos de prueba. ....	31
Tabla III. Información detallada de los grupos que contiene cada una de las bases de datos de prueba. ....	31
Tabla IV. Información detallada de las sesiones que contiene cada una de las bases de datos de prueba. ....	31
Tabla V. Tecnologías y herramientas empleadas en el desarrollo de la aplicación web. ....	32
Tabla VI. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa del PREP. ....	34
Tabla VII. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa de Wavelet-ICA.....	35
Tabla VIII. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa de rechazo de épocas ruidosas.....	35
Tabla IX. Duración de tiempo para el procesamiento de la información del conjunto de datos de prueba. ....	43
Tabla X. Componentes de la interfaz ‘Carga de datos’ con su respectiva funcionalidad y propósito de diseño.....	69
Tabla XI. Componentes de la interfaz ‘Procesamiento de datos’ con su respectiva funcionalidad y propósito de diseño.....	71
Tabla XII. Duración de tiempo para la generación de los gráficos de control de calidad de datos. ....	73
Tabla XIII. Componentes de la interfaz ‘Control de calidad de datos’ con su respectiva funcionalidad y propósito de diseño.....	74
Tabla XIV. Resultados de la prueba de usabilidad .....	81

## LISTA DE FIGURAS

Fig 1. Sistema Internacional 10-20. ....	17
Fig 2. Subdivisión de los ritmos frecuenciales en estado de reposo registrados en el cuero cabelludo. Federación Internacional de Neurofisiología Clínica (IFCN por sus siglas en inglés “International Federation of Clinical Neurophysiology”); Sociedad Internacional de Fármaco-EEG (IPEG por sus siglas en inglés “International Pharmaco-EEG Society”)[16]. ....	19
Fig 3. Estructura del estándar BIDS para datos de EEG [9]. ....	21
Fig 4. Componentes comunes en los gráficos de cajas y bigotes y los gráficos de violín. ....	23
Fig 5. Documentación interactiva de la API e interfaces web de exploración. ....	25
Fig 6. Tendencia de uso de frameworks web 2021, S. O. (2 de diciembre de 2021). Stack Overflow Developer Survey 2021. Tomado de: Stack Overflow Developer Survey 2021. <a href="https://insights.stackoverflow.com/survey/2021">https://insights.stackoverflow.com/survey/2021</a> . ....	26
Fig 7. Tecnologías y herramientas para la construcción de la aplicación. ....	32
Fig 8. Esquema de metodología. <b>A:</b> 1) Adquisición de los registros EEG, 2) Estandarización formato BIDS, <b>B:</b> 3) Preprocesamiento de los registros, <b>C:</b> 4) Post-procesamiento de los registros, 5) Análisis de calidad de los datos. ....	33
Fig 9. Proyecciones de mapas del cuero cabelludo de componentes gICA encontrados etiquetados como neuronales. ....	36
Fig 10. Estructura de los datos crudos después de aplicar el estándar BIDS. ....	38
Fig 11. Sistema de paquetes para el procesamiento de datos EEG. ....	39
Fig 12. Derivados del procesamiento según la nomenclatura y la extensión del estándar BIDS. ....	41
Fig 13. Script de normalización de Huber en Python. ....	42
Fig 14. Almacenamiento de métricas de calidad. ....	42
Fig 15. Análisis global para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas. ....	46
Fig 16. Análisis entre estudios para las diversas métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas. ....	48
Fig 17. Análisis entre grupos de las diversas métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas. ....	50
Fig 18. Análisis entre sesiones en el conjunto de datos SRM para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas. ....	52

Fig 19. Análisis entre sesiones en el conjunto de datos BIOMARCADORES para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.....	53
Fig 20. Análisis global en componentes gICA para datos procesados y normalizados. A) Potencia relativa en canales B) Potencia relativa en componentes gICA.....	55
Fig 21. Análisis entre estudios de la potencia relativa de los datos procesados y normalizados en canales. El eje X representa los conjuntos de datos, y el eje Y indica el valor de potencia relativa. ....	57
Fig 22. Potencia relativa entre estudios en componentes gICA en datos procesados y normalizados. El eje X indica los diferentes conjuntos de datos y el eje Y indica la potencia relativa. Las columnas hacen referencia a las bandas de frecuencia y las columnas a las componentes neuronales. ....	58
Fig 23. Análisis de la potencia relativa de los datos procesados y normalizados en canales para el conjunto de datos de Biomarcadores. El eje X representa los grupos de estudio, y el eje Y indica el valor de potencia relativa. ....	59
Fig 24. Análisis de la potencia relativa de los datos procesados y normalizados en componentes neuronales para los datos de Biomarcadores. El eje X representa las bandas de frecuencia, y el eje Y indica el valor de potencia relativa.....	60
Fig 25. Análisis de la potencia relativa del conjunto de datos SRM procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia .....	61
Fig 26. Análisis de la potencia relativa del conjunto de datos BIOMARCADORES procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. A) CTR; B) DCL; C) DTA.....	62
Fig 27. Análisis de la potencia relativa del conjunto de datos BIOMARCADORES procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia. A) G1; B) G2.....	63
Fig 28. Análisis de la potencia relativa del conjunto de datos SRM procesados y normalizados. El eje X representa las sesiones de estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia y las filas corresponde a las componentes gICA etiquetadas como neuronales.....	64
Fig 29. Interfaz de inicio. 1) Panel de navegación; 2) Ingreso al sistema.....	66



Fig 30. Sistema de alertas para notificar acceso denegado de navegación. ....	66
Fig 31. Sistema de ingreso del usuario.....	67
Fig 32. Interfaz carga de datos. 1) Cargar datos en formato BIDS; 2) Ingresar campos obligatorios acorde a la nomenclatura del estándar BIDS; 3) Visualización de las bases de datos cargadas. ...	68
Fig 33. Interfaz procesamiento de datos.1) Visualización de las bases de datos previamente cargadas. ....	70
Fig 34. Notificación de la solicitud de procesamiento de datos.....	70
Fig 35. Notificación de procesamiento completado al correo electrónico.....	71
Fig 36. Interfaz de control de calidad de datos. 1) Panel de selección de condición para analizar; 2) Panel de selección de análisis de control de calidad; 3) interfaz derivada de los botones de análisis según corresponda. ....	72
Fig 37. Gráficos de JavaScript para el análisis global de los datos.....	75
Fig 38. Panel de operaciones de las gráficas generadas con JS. ....	76
Fig 39. Texto descriptivo para la interpretación de las métricas de control de calidad del PREP.	77
Fig 40. Texto descriptivo para la interpretación de las métricas de control de calidad del wICA.	77
Fig 41. Texto descriptivo para la interpretación de las métricas de control de calidad de rechazo de épocas ruidosas.....	77
Fig 42. Texto descriptivo para la interpretación de las métricas de control de calidad del post-procesamiento.....	78
Fig 43. Servicio systemctl para el procesamiento de los datos. ....	78
Fig 44. Servicio systemctl para la visualización y gestión de los datos.....	79
Fig 45. Servicio systemctl para el frontend.....	79
Fig 46. Resultados promedio de la prueba de usabilidad. El eje 'X' representa la puntuación promedio de los participantes encuestados, y el eje 'Y' hace referencia a cada una de las preguntas de selección múltiple.....	82

## SIGLAS, ACRÓNIMOS Y ABREVIATURAS

<b>EEG</b>	Electroencefalografía
<b>wICA</b>	Wavelet- Independent Component Analysis
<b>gICA</b>	Geometrical Independent Component Analysis
<b>JS</b>	JavaScript
<b>SUS</b>	System Usability Scale
<b>RANSAC</b>	Random Sample Consensus
<b>PREP</b>	Early-stage EEG processing pipeline
<b>BIDS</b>	Brain Imaging Data Structure
<b>FAIR</b>	Findability, Accessibility, Interoperability, and Reusability

## RESUMEN

El uso de las Tecnologías de la Información y las Comunicaciones permiten el crecimiento de la información y la mejora de la calidad de los sistemas. En el área de la salud se han venido automatizando procesos y restaurando sistemas con el objetivo de crecer y optimizar flujos de trabajo que permitan la centralización de la información. En el procesamiento de señales, más específicamente en las señales electroencefalográficas (EEG) es importante direccionar la información a partir de herramientas que permitan mantener de forma organizada los datos para la gestión y manipulación de grandes volúmenes. El EEG en los últimos años ha tomado gran importancia en el área de la neurociencia cognitiva con una tendencia creciente en aplicaciones de investigación para el diagnóstico de enfermedades, por tal motivo, es necesario desarrollar metodologías y herramientas que faciliten el análisis y lecturas de estos datos. Este trabajo presenta el desarrollo de una aplicación WEB que permite el preprocesamiento y visualización de los datos a partir de la implementación de estándares de informática médica basada en tecnologías de Python y NodeJS. A partir de esta aplicación se obtienen herramientas para el procesamiento y la extracción de características con el propósito de brindar alternativas para el análisis de control y calidad de los datos empleando técnicas de inspección visual basadas en diagramas estadísticos.

***Palabras clave* — EEG, WEB, preprocesamiento, control de calidad, análisis estadístico.**

## ABSTRACT

The use of Information and Communication Technologies allow the growth of information and the improvement of the quality of the systems. In the health area, processes have been automated and systems restored with the aim of growing and optimizing workflows that allow the centralization of information. In signal processing, more specifically in electroencephalographic (EEG) signals, it is important to direct the information from tools that allow the data to be maintained in an organized way for the management and manipulation of large volumes. The EEG in recent years has taken significant importance in the area of cognitive neuroscience with a growing trend in research applications for the diagnosis of diseases, for this reason, it is necessary to develop methodologies and tools that facilitate the analysis and reading of these data. This work presents the development of a WEB application that allows the preprocessing and visualization of data from the implementation of medical informatics standards based on Python and NodeJS technologies. From this application, tools are obtained for the processing and extraction of characteristics with the purpose of providing alternatives for the analysis of control and quality of the data using visual inspection techniques based on statistical diagrams.

***Keywords*** — EEG, WEB, pipeline, quality control, statistical analysis.

## I. INTRODUCCIÓN

El uso de las Tecnologías de la Información y las Comunicaciones (TICs) permite el trabajo colaborativo entre disciplinas para brindar eficacia y el óptimo funcionamiento de estas. En los últimos años, en el área de la salud se ha evidenciado la importancia de gestionar y almacenar datos a partir de herramientas de soporte como softwares o plataformas web, lo cual da lugar a un mayor uso de la informática médica a través del campo de la tecnología de la información. La adopción de esta disciplina permite la interconectividad entre áreas, la utilización de sistemas de información en salud y el acceso a información científica técnica, proporcionando procesos óptimos de interoperabilidad y accesibilidad. No obstante, el crecimiento de volúmenes de datos en esta área ha creado la necesidad de almacenar y centralizar la información a partir de herramientas que incluyan los conceptos de Big Data y ciencia de datos, de forma que se optimicen los recursos del sistema de salud [1].

Los avances en gestión de datos permiten el desarrollo de aplicaciones para la captura, almacenamiento y manipulación de grandes volúmenes de información [2]. Sin embargo, resulta retador procesar y almacenar conjuntos de datos con características tan complejas, es por esta razón que el manejo del concepto de Big Data tiene un trasfondo más allá del conjunto de grandes flujos de datos. Este concepto está descrito por diferentes características como la de poseer alta velocidad, y variedad de información. En el área de procesamiento de señales médicas, diferentes autores muestran desarrollos utilizando herramientas que promueven la estandarización e interoperabilidad de la información. Por ejemplo, el estudio presentado por Beier et al, en el cual proponen el trabajo multidisciplinar por medio de una aplicación web enfocado al rendimiento del tamaño de archivos para la visualización y evaluación de señales biológicas; este desarrollo, apunta al trabajo colaborativo entre áreas de forma remota, lo cual fomenta el desarrollo de nuevos aportes en el curso de la investigación [3]. Rendón, en su estudio presenta avances que muestran la importancia de almacenar información biológica centralizada en la nube, específicamente de la población colombiana, como futuro índice de investigación para el análisis descriptivo de los datos haciendo énfasis en el desarrollo de una aplicación web como repositorio de bioseñales [4]. Además, Dong et al, desarrollaron una plataforma web para aplicaciones de registros electroencefalográficos (EEG) en donde proveen diferentes herramientas y flujos de preprocesamiento común, lo cual

proporciona una solución en la nube para el análisis de estos datos [5]. Estos estudios muestran la importancia de gestionar, procesar y almacenar la información de forma adecuada, empleando estándares de informática médica tanto en el área clínica como en la investigación.

Este trabajo tiene por objeto de estudio gestionar datos provenientes de señales EEG, ya que estas proporcionan información relevante que permiten la visualización y evaluación de la actividad cerebral, lo cual ha mostrado tener gran potencial en diferentes aplicaciones e investigaciones relacionadas con la conducta neuropsicológica y enfermedades neurodegenerativas [6]. Además, en la interpretación de los datos de los registros EEG, es indispensable realizar una etapa de preprocesamiento de las señales que permita obtener únicamente la información neurológica, por tal razón, en esta etapa se emplean diferentes estrategias que permitan extraer dicha información lo más limpia posible, con la menor pérdida de información.

Diferentes investigadores han mostrado los efectos de varias etapas sobre las señales, normalmente se reporta el uso de la metodología del preprocesamiento en etapa temprana (PREP) la cual está compuesta de tres criterios fundamentales. Inicialmente la remoción del ruido de línea sin necesidad de aplicar un conjunto de estrategias basados en filtros, luego realizar una referenciación robusta con respecto a una referencia promedio y finalmente realizar una detección e interpolación de canales defectuosos [7]. Así mismo, se emplean herramientas de limpieza de ruido a partir de identificación de componentes independientes, esta técnica se conoce como wICA la cual ha demostrado tener un mejor efecto sobre otros métodos para el filtrado y eliminación de artefactos [8]. Sin embargo, no existe una estandarización de la metodología de preprocesamiento para estas señales, haciendo necesario hacer seguimientos, validaciones y controles de calidad de los datos antes y durante el procesamiento para evitar realizar interpretaciones erróneas sobre las señales finales.

Una de las estrategias más comunes antes del procesamiento de los datos, es la organización de estos, la cual ha sido desarrollada a partir del estándar “Brain Imaging Data Structure” conocido por su acrónimo como BIDS. Este método fue diseñado inicialmente para imágenes de resonancia magnética y actualmente se realizó una estructuración para señales EEG debido al crecimiento de

volumen de datos de este tipo, permitiendo así tener una estructuración que facilite el compartimiento de forma más organizada los volúmenes de datos [9].

De acuerdo con lo anterior y a la problemática presentada, este trabajo propone el desarrollo de una aplicación web haciendo uso de las tecnologías disponibles de backend y frontend para el procesamiento y almacenamiento de señales biológicas aplicando estándares de informática médica como lo es el estándar BIDS para registros EEG. Con el fin de poder ser analizadas a partir de herramientas estadísticas y de interpretación visual que permitan realizar una evaluación sobre la calidad de los datos durante la fase de procesamiento y posterior a esta.

En síntesis, este documento se encuentra organizado de la siguiente manera: 1) la sección de marco teórico y antecedentes que ilustran los principios fundamentales en los que se basó el desarrollo de este trabajo, 2) el apartado de la metodología, en donde se describe y se detalla cada uno de los procesos puestos en consideración para dar cumplimiento a cada uno de los objetivos propuestos, 3) la sección de resultados y discusión, en donde se presentan cada uno de los avances realizados, de acuerdo al esquema presentado en la metodología y se realiza el análisis de los resultados y 4) donde finalmente se muestran las conclusiones sobre el desarrollo y las propuestas para investigaciones futuras.

## II. OBJETIVOS

### *A. Objetivo general*

Desarrollar una aplicación WEB que permita la centralización de la información y el preprocesamiento de señales electroencefalográficas (EEG), haciendo uso de las tecnologías disponibles de backend y frontend.

### *B. Objetivos específicos*

Diseñar una interfaz que permita el preprocesamiento de señales EEG basados en los estándares de informática médica.

Desarrollar la aplicación web para un servidor y verificar el acceso remoto a otros usuarios.

Validar la aplicación web, realizando pruebas requeridas que garanticen el correcto funcionamiento de la plataforma.



### III. MARCO TEÓRICO

#### *Electroencefalografía*

La electroencefalografía (EEG) es una técnica específica y no invasiva que permite medir la actividad eléctrica cerebral mediante la aplicación de electrodos sobre el cuero cabelludo, permitiendo medir el cambio de potencial eléctrico a través del tiempo [10]. La actividad eléctrica registrada representa principalmente la suma de los potenciales postsinápticos excitatorios e inhibitorios en las dendritas apicales de las neuronas piramidales, en donde a mayor activación del área se generan cambios de potencial más altos que permiten ser captados por los electrodos [11].

El número de electrodos y su distribución espacial varía de acuerdo con diferentes montajes, y son posicionados de acuerdo al sistema internacional 10-20 (Fig 1) [12]. Existen dos tipos de montajes básicos: bipolar y monopolar; el montaje bipolar estima las diferencias de potencial entre dos electrodos adyacentes, mientras que el montaje de tipo monopolar registra la diferencia entre un electrodo ubicado en una región cerebral activa con respecto a un electrodo de referencia, que normalmente es ubicado en el lóbulo de la oreja [13].

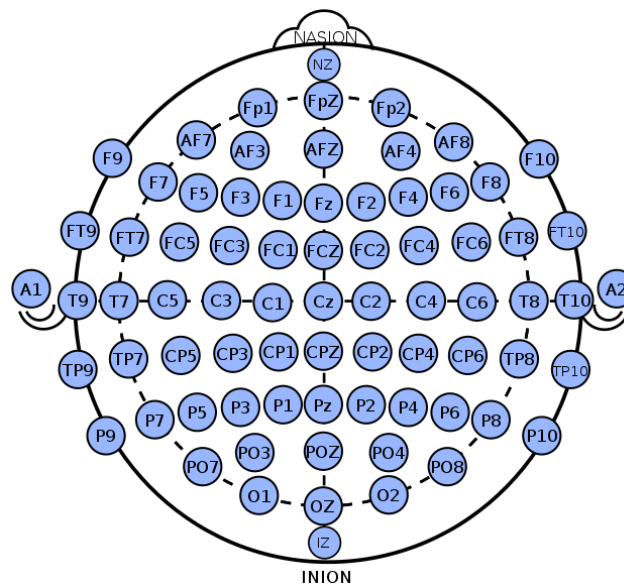


Fig 1. Sistema Internacional 10-20.

Cada canal de electrodos de EEG registra información eléctrica que se proyecta sobre este, la cual da como resultado, la medición de una mezcla lineal de un conjunto de señales que no necesariamente son de tipo neuronal, sino que son producidas por otras fuentes biológicas, a esto se le conoce como artefactos. Los artefactos son señales que provienen de fuentes ajenas a la actividad en la corteza, enmascarando la señal resultante [14].

Las señales EEG pueden analizarse en el dominio de la frecuencia utilizando la transformada de Wavelet o la descomposición de Fourier para obtener los componentes frecuenciales para su correspondiente clasificación. La transformada de Fourier utiliza funciones sinusoidales o exponenciales complejos como funciones base. En la Ecuación 1, se presenta la transformada de Fourier siendo  $x(t)$  la señal a transformar,  $\omega$  la frecuencia angular y  $F\{\cdot\}$  el operador de Fourier [15].

$$F\{x(t)\} = X(\omega) = \int_{-\infty}^{\infty} x(t) \exp(-j\omega t) dt$$

Ecuación 1. Transformada de Fourier

La transformada de Fourier permite obtener una estimación espectral y basándose en esta se puede calcular la Densidad de Potencia Espectral (PSD por sus siglas en inglés “Power Spectral Density”), la cual informa sobre la distribución de la potencia de una señal entre sus diferentes frecuencias. En EEG se usan letras del alfabeto griego para denominar rangos de frecuencia específicos. Existen diferentes clasificaciones, las cuales establecen de manera individual los límites en los que se encuentra cada banda de frecuencia. Esta clasificación es presentada por Babiloni et al, donde reportan la siguiente distribución según diferentes asociaciones mostrada en la Fig 2 [16].

Frequency (Hz)	IFCN 1999 (I)	IFCN 1999 (II)	IPEG 2012	IFCN-2017 Glossary
Delta	0.5–4	0.5–4	1.5–<6	0.1–<4
Theta	4–8	5–7	6–<8.5	4–<8
Alpha	$\alpha$ 1: 8–10 $\alpha$ 2: 10–12/13	8–12	$\alpha$ 1: 8.5–<10.5 $\alpha$ 2: 10.5–<12.5	8–13
Beta	$\beta$ 1: 12–16 $\beta$ 2: 16–20 $\beta$ 3: 20–24 $\beta$ 4: 24–28 $\beta$ 5: 28–32	$\beta$ 1: 14–20 $\beta$ 2: 21–30	$\beta$ 1: 12.5–<18.5 $\beta$ 2: 18.5–<21 $\beta$ 3: 21.0–<30	14–30
Gamma	$\gamma$ 1: 32–36 $\gamma$ 2: 36–40 $\gamma$ 3: 40–44 $\gamma$ 4: 44–48 ...	$\gamma$ 1: 30–40 $\gamma$ 2: 40–...	30–<40 $\gamma$ 1: 30–<65* $\gamma$ 2: 65–<90* $\gamma$ 3: 90–<135*	>30–80

\* Empirical subdivision.

Fig 2. Subdivisión de los ritmos frecuenciales en estado de reposo registrados en el cuero cabelludo. Federación Internacional de Neurofisiología Clínica (IFCN por sus siglas en inglés “International Federation of Clinical Neurophysiology”); Sociedad Internacional de Fármaco-EEG (IPEG por sus siglas en inglés “International Pharmacology EEG Society”)[16].

En general los ritmos presentados en la Fig 2 tienen una interpretación fisiológica. La frecuencia delta, está relacionada con el sueño profundo y es prominente en las regiones frontocentrales de la cabeza. Los ritmos theta son provocados por la somnolencia y las primeras etapas del sueño como N1 y N2, tienen prominencia en las regiones frontocentrales. La banda alfa dominante posterior, se encuentra presente de forma característica en los registros de EEG normal despiertos en la región occipital de la cabeza. Beta es prominente en las regiones frontal y central de la cabeza, suele aumentar durante la somnolencia el sueño tipo N1 y disminuye en el sueño N2 Y N3 [17].

En síntesis, esta técnica ha tenido una aplicación muy útil para describir el funcionamiento cerebral en campos como neurofisiología clínica, estudios de neurociencia cognitiva e interfaces cerebro-computadora debido a su alta resolución temporal y a que es una alternativa portátil de bajo costo que ofrece la posibilidad de llevar un control de los cambios dinámicos de la actividad eléctrica del cerebro[18][19][20].

### *Estándar BIDS*

El estándar de datos de imágenes cerebrales, más conocido como BIDS por sus siglas en inglés, tiene como objetivo la heterogeneidad de la organización de los datos con la finalidad de facilitar procesos de búsqueda y accesibilidad, que permita a los investigadores organizar y

compartir fácilmente los datos pertenecientes a un estudio, dando como resultado una mayor interoperabilidad y reutilización de los mismos [21].

Este estándar fue originalmente propuesto para datos de imágenes por resonancia magnética, sin embargo, el crecimiento de la información de señales electroencefalográficas da como resultado un volumen de información compleja que pueden ser organizado de diversas maneras, es por esta razón que Pernet et al, proponen una estructura basada en este estándar para la organización de datos de EEG [9]. Este estándar fue fundamentado en los principios “ Findability, Accessibility, Interoperability, and Reusability” más conocido por su acrónimo FAIR, el cual enfatiza en la optimización y automatización de los procesos para encontrar y reutilizar datos, que permitan estandarizar flujos de trabajo neurocientíficos para el correcto almacenamiento y procesamiento [22].

Esta estructura organiza los datos de cada sujeto en un directorio independiente el cual contiene los datos crudos, que a su vez contienen subdirectorios para cada sesión y modalidad. Dentro de la arquitectura del estándar existen dos archivos que contienen información genérica de la base de datos, los cuales son: “dataset\_description.json” y “eeg.json”. Estos archivos contienen información de los metadatos, tarea experimental y sistema de registro de EEG; cada subdirectorio y sus respectivos archivos están organizados según el patrón de nomenclatura definido por este estándar, así mismo todas las derivaciones que salgan del flujo de estándar deberán compartir este tipo de estructuración. En la Fig 3, se encuentra detallado la estructura básica en la que se organizan los datos.

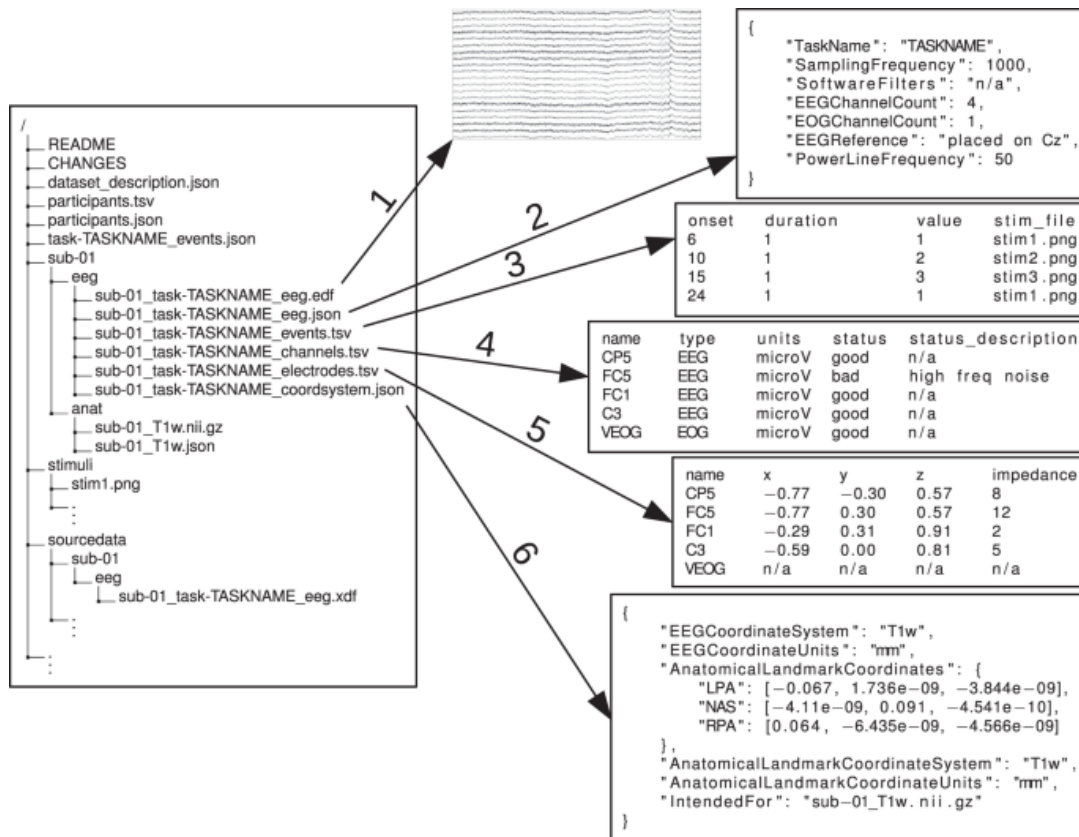


Fig 3. Estructura del estándar BIDS para datos de EEG [9].

### Procesamiento de registros EEG

El preprocesamiento de las señales EEG es un paso indispensable para la extracción de características y análisis posteriores de la señal. Sin embargo, existen ciertas limitaciones como lo es la estandarización, automatización y cuantificación del procesamiento [23]. Durante los últimos años se han venido reportando los efectos que tienen diferentes etapas de preprocesamiento sobre las señales con el objetivo de mantener únicamente la información asociada a la actividad cerebral.

Las etapas principales que se han venido reportando son la referenciación y la eliminación de canales defectuosos a partir de un flujo de procesamiento de EEG estandarizado en etapa temprana (PREP), la cual es una herramienta automatizada de código abierto que permite la evaluación y comparación de canales malos [7]. Otro paso importante en el procesamiento es la remoción e identificación de artefactos, utilizando técnicas como análisis de componentes independientes más conocido como ICA por sus siglas en inglés; esta técnica permite estimar

señales de fuente independientes a partir de un conjunto de grabaciones, en las que las señales de diferentes fuentes se mezclaron en proporciones desconocidas. El supuesto de esta técnica es que las señales se pueden descomponer en procesos independientes entre las fuentes neuronales de las señales no cerebrales, cancelando así el proceso de mezcla de los datos [24]. Esta técnica se ha venido utilizando junto con la transformada Wavelet, dando a lugar el método de limpieza wICA, la cual es una técnica mejorada para la remoción de artefactos oculares y ruido, no de manera general, sino en los componentes aislados que entrega la técnica del ICA para así evitar la mayor pérdida de información [25][26].

Un enfoque que se ha venido implementando en el procesamiento, es utilizar modelos de escala y normalización que permitan disminuir la variabilidad introducida durante la adquisición de los datos que puede ser dada por la tecnología de registro, ubicación de los electrodos, las características del cuero cabelludo del sujeto u otro factor que aumente dicho problema, para esto Nima et al [27], evaluaron diferentes técnicas de normalización en registros electroencefalográficos y el que presentó tener mejores resultados fue el estimador de Huber, más específicamente la media, la cual es una ponderación de tres medidas: mediana, primer y tercer cuartil. Luego de esta etapa se suele extraer las características según los modelos y tipos de estudios que se estén realizando.

### *Control de calidad de los datos*

El control de calidad de los datos es un proceso elemental en el desarrollo de flujos de trabajo, para dar validez y seguimiento a los procesos realizados. Cada investigador define protocolos de seguimiento acorde al tipo de estudio y procesamiento que se esté realizando, por ejemplo, Valdés et al [28], para el control de calidad de registros de EEG proponen tres etapas principales, inicialmente anotación de los movimientos u otras incidencias durante los registros, luego una inspección visual de los datos con la finalidad de detectar cuáles registros debía ser repetidos. Además, se realizan criterios de exclusión de la muestra normativa y finalmente la selección de segmentos de EEG libres de información no neuronal para el posterior análisis. Sin embargo, este tipo de revisiones carecen de automatización, lo cual puede ser poco óptimo en el control de calidad de alto flujo de datos. En general, las metodologías reportadas para el control de

calidad están basadas en inspección visual, protocolos de adquisición según equipo y estudio, validaciones dentro de los flujos de trabajo.

### *Análisis de estadística descriptiva*

Para medir las magnitudes y tendencias existen diferentes representaciones gráficas como lo son los diagramas de barras, diagramas de cajas, diagramas de violín, y estimaciones del Kernel de densidad [29]. Estas gráficas se agrupan en función de la información presente en los metadatos. De acuerdo con la información y a las características que se deseen analizar, un tipo de gráfico es más representativo que el otro. Por ejemplo, los gráficos de violín son representaciones que permiten medir la distribución de los datos a partir de la forma que muestra la gráfica, este contorno representa la distribución que puede ser de tipo Gaussiana dependiendo de la cantidad de datos atípicos que se encuentren en la muestra. Esta representación, combina la funcionalidad de los diagramas de cajas y bigotes que permiten detallar la presencia de *outliers* conocidos como datos atípicos y cada uno de los cuartiles de la distribución. Sin embargo, este tipo de gráfico solo es útil cuando se tienen metadatos de tipo robusto, ya que con un conjunto de datos pequeños la estimación del Kernel puede no detallar el perfil real de los datos. En la Fig 4, se encuentra una representación general en donde se describe cada una de las características y especificaciones que nos entrega esta métrica.

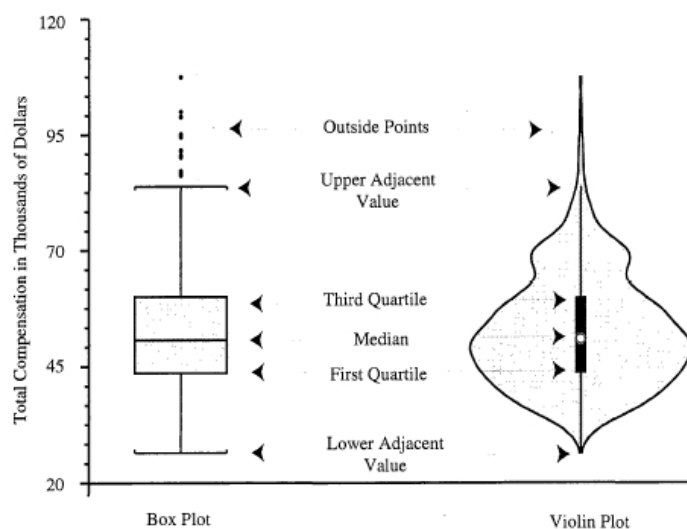


Fig 4. Componentes comunes en los gráficos de cajas y bigotes y los gráficos de violín.

### *Mecanismo arquitectónico de trabajo*

Para la implementación de la aplicación web es fundamental el desarrollo de dos componentes primarios, en los cuales, en primer lugar, se desarrollan los procesos operativos del software no visibles para posteriormente ser presentados de forma interactiva al usuario final a partir de una interfaz gráfica.

#### *Backend*

El backend es la capa de acceso a datos de un software o dispositivo, que no es accesible para el usuario final, es decir, es aquel que se encarga de interactuar y administrar las bases de datos para el manejo de la información, además, de poder acceder a los datos de tal manera que puedan ser consumibles en cualquier servidor. Este módulo es de vital importancia ya que de este depende el correcto funcionamiento de la aplicación, el cual involucra aspectos relevantes como la rapidez de carga, seguridad y acceso a navegación. Para la captura de información y el intercambio de datos entre la capa visible y la capa de acceso se manejan el uso de las Interfaces de Programación de Aplicaciones, APIs por sus siglas en inglés las cuales permiten el manejo y comunicación de la información a partir de ciertos formatos como lo son XML o JSON.

FastAPI, es un framework desarrollado en Python 3.6+ diseñado para desarrollar API RESTfull compatible con programación asíncrona la cual busca la ocurrencia de eventos independientes del flujo de trabajo principal para trabajar sobre solicitudes mientras el procesador continúa ejecutando el programa [30]. Este framework es característico de poseer alto rendimiento como entornos de ejecución multiplataforma como lo es Node y se encuentra basado en estándares compatibles para APIS como lo es JSON. Una de las principales ventajas que presenta con respecto a otros frameworks, es que posee una documentación interactiva en donde se pueden realizar pruebas y consultas organizadas y de fácil acceso a partir de herramientas como “Swagger UI” desde el navegador.

FastAPI al contar con Swagger permite realizar pruebas de llamada sobre los métodos CRUD (Create, Read, Update, Delete) lo cual resulta útil para identificar y atacar los errores de



forma rápida y eficaz. En la Fig 5, se encuentra un ejemplo de la estructuración que presenta esta herramienta. Este framework es relativamente nuevo, inició en el año 2018, sin embargo, ha conseguido escalar de forma rápida siendo uno de los más populares de Python entre Django y Flask.

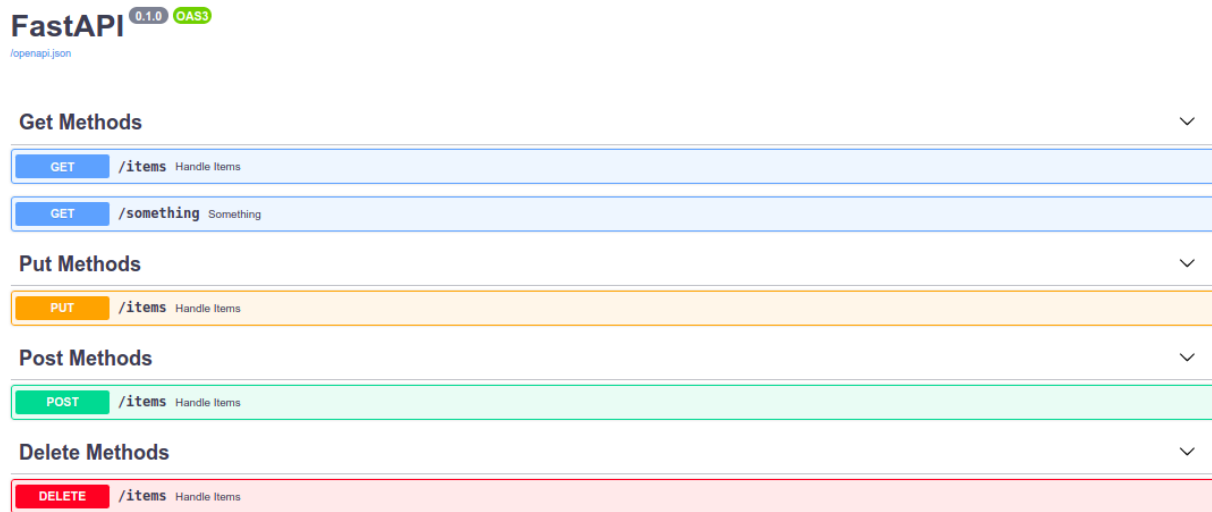


Fig 5. Documentación interactiva de la API e interfaces web de exploración.

### *Frontend*

El frontend se conoce como un conjunto de tecnologías de desarrollo web del lado del cliente, donde el usuario puede navegar e interactuar con la plataforma [31]. El desarrollo de esta parte del sitio web está enfocada a brindar una buena experiencia al usuario, usabilidad e inmersión. Actualmente, existen una gran variedad de frameworks, preprocesadores y librerías que permiten el desarrollo y diseño eficaz de los diferentes componentes visuales que dan animación e interactividad al sitio web.

El desarrollo del front-end está constituido principalmente por tres lenguajes: HTML (Hyper Text Markup Language), CSS (Cascade StyleSheets) y JavaScript más conocido por sus siglas como JS. HTML es un lenguaje de marcado que permite ordenar y etiquetar diversos documentos dentro de una lista, sin embargo, se han diseñado otros lenguajes computacionales para facilitar la gestión de los datos relacionados con el diseño visual de las plataformas [32]. Uno

de estos lenguajes es CSS el cual permite crear y definir el diseño gráfico para la presentación de un documento estructurado de forma que se puedan ofrecer contenidos de una página de forma atractiva para el usuario. Por otro lado, JavaScript es un lenguaje de programación que se utiliza para implementar más opciones de interacción y un mayor dinamismo de sus contenidos, a los sitios web [33]. La implementación conjunta de estos tres lenguajes mencionados anteriormente, permiten el desarrollo eficaz y óptimo diseño de la interfaz gráfica más conocida como frontend.

React es una biblioteca declarativa, dinámica y flexible de JavaScript utilizado para construir interfaces de usuario interactivas que se encuentra estructurado mediante un paradigma declarativo que promueve la creación de componentes que contienen un módulo lógico y un módulo para la visualización de este. Al ser desarrollado mediante esta metodología permite la reutilización de componentes dentro de otros, lo cual es una de las características importantes. En el año 2021 según el reporte anual que realiza Stack Overflow, React ha superado como el framework web más utilizado, esta información se muestra en la Fig 6 en donde se muestra los cinco primeros lugares ocupados.

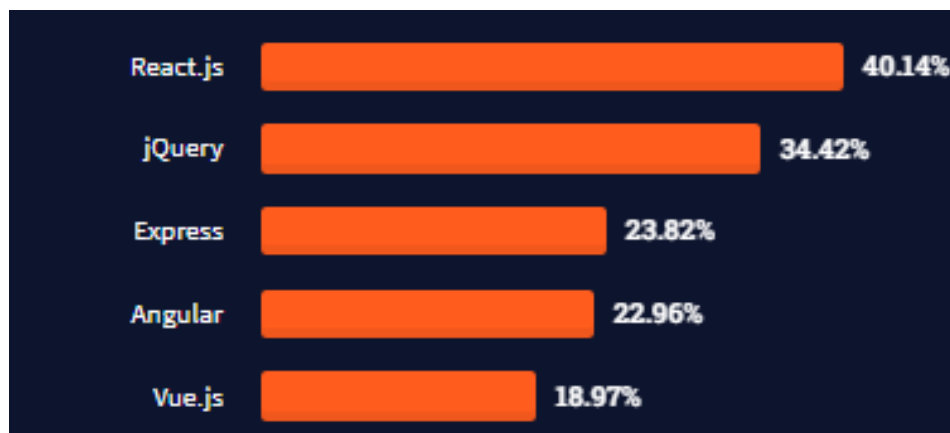


Fig 6. Tendencia de uso de frameworks web 2021, S. O. (2 de diciembre de 2021). Stack Overflow Developer Survey 2021. Tomado de: Stack Overflow Developer Survey 2021. <https://insights.stackoverflow.com/survey/2021>.

### *Gestión de servicios*

Un servicio es un programa que se ejecuta en segundo plano, fuera del control interactivo de los usuarios del entorno. En la distribución de Linux se conocen como 'Daemons', el cual es un proceso de ejecución prolongado que opera en el fondo del sistema [34].

El sistema operativo de Linux, para arrancar todos los procesos por debajo del Kernel utiliza SystemD, el cual es un administrador de servicios que permite ejecutar varios procesos de forma simultánea. Este gestor de servicios permite crear dependencias fundamentadas en el control lógico con el propósito de optimizar el uso de recursos.

Los servicios controlados por systemd son monitoreados y evaluados a partir de Systemctl, el cual es una interfaz de control que conecta varios aspectos del sistema para dar seguimiento del estado de cada servicio, o para activar/desactivar el mismo. Una vez se inicia un servicio, es posible conocer la trayectoria, y las causas por las cuales un sistema podría estar fallando, es por esta razón, que la gestión de estos permite cumplir y solucionar las demandas de los clientes durante el proceso de ejecución.

### *Usabilidad*

La escala de usabilidad del sistema (SUS) es una herramienta que permite medir la percepción de los sistemas tecnológicos del lado del cliente, para así conocer la capacidad que tiene un usuario de interpretar, aprender y navegar un software. SUS, fue diseñado mediante cinco enfoques que permiten medir la calidad de las herramientas tecnológicas, los cuales son: capacidad de aprendizaje, eficiencia, memorabilidad, errores y satisfacción [35]. La capacidad de aprendizaje está asociada con la facilidad que tiene el usuario de asimilar nueva información y aplicarlo de forma eficaz; la eficiencia, está relacionada con la capacidad del usuario de ejecutar una función; la memorabilidad esta enlazada con el proceso de aprendizaje de un usuario para volver a utilizar una aplicación después de un periodo de tiempo; los errores y satisfacción están estrechamente conectados, ya que entre menor cantidad de errores cometa un usuario al utilizar un sistema, mayor será el nivel de agrado del cliente.

Esta prueba consiste en un cuestionario de diez ítems con cinco opciones de respuesta que va desde '*Totalmente de acuerdo*' hasta '*Totalmente en desacuerdo*'; es una forma de medir de manera rápida los sistemas informáticos desde el punto de vista de la usabilidad. En la Tabla I, se encuentra descrito cada valoración de la prueba y su significado.

Tabla I. Valoración para la prueba de usabilidad

<b>Puntuación</b>	<b>Significado</b>
1	Totalmente en desacuerdo
2	En desacuerdo
3	Neutral
4	De acuerdo
5	Totalmente de acuerdo

Las puntuaciones finales de esta prueba esta entre un rango de 0-100, y para realizar el cálculo de esta, se siguen los siguientes pasos [35]:

1. Dividir las preguntas entre pares e impares.
2. Las preguntas impares se reducen a un valor de 'x-1'.
3. Las preguntas pares, se aplica la ecuación: '5-x'.
4. Realizar una sumatoria de los valores obtenidos por todas las preguntas y multiplicar los resultados por 2,5.

La interpretación de las puntuaciones SUS son descriptas por una escala adjetiva, en donde, los valores superiores a 80.3 puntos, se clasifican como grado **A**, siendo '**excelente**'; valores entre 68 y 80.3 se categorizan como grado **B**, siendo '**bueno**'; una puntuación de 67 es grado **C**, descripto como '**okey**'; Entre 51-66 se denota como grado **D** y una calificación como '**pobre**' y aquellos valores por debajo de 51 con grado **F** y se califica como '**horrible**'.

## IV. METODOLOGÍA

Para dar cumplimiento a los objetivos establecidos en el presente trabajo se llevó a cabo la siguiente metodología la cual está dividida en tres etapas principales. Primero se inició la conformación de la base de datos de prueba de registros electroencefalográficos para aplicar el procesamiento y hacer el análisis de post-procesamiento, luego de esto se definió las tecnologías y herramientas de desarrollo y finalmente se definieron las métricas y la taxonomía que permitían hacer una evaluación sobre la calidad de los datos desde lo más general hasta lo más específico utilizando técnicas de exploración de datos visual.

### A. Datos

El conjunto de datos de prueba consiste en tres bases de datos robustas las cuales se utilizarán como los datos de estudio para las pruebas de procesamiento y análisis de calidad de datos, para esto fue necesario utilizar estudios que posean características similares entre sí, como lo es el paradigma de estudio. Las bases de datos utilizadas tienen en común principalmente la adquisición en condición de reposo con los ojos cerrados. La primera base de datos utilizada pertenece al grupo de Neuropsicología y conducta (GRUNECO) de la Universidad de Antioquia, la cual posee un repositorio de datos de electroencefalografía que no es de fuente abierta. Este repositorio está conformado por cinco grupos que corresponden a sujetos controles (CTR), sujetos con deterioro leve cognitivo (DCL), sujetos con demencia tipo Alzheimer (DTA), y sujetos que son miembros de la familia colombiana variante PSEN1-E280A, no portadores (G1) y portadores (G2).

Las señales de EEG se adquirieron en estado de reposo con los ojos cerrados (OC) durante 5 minutos. Los datos se registraron a partir de un amplificador Neuroscan unif (Neuroscan Medical System, Neurosoft Inc. Sterling, VA, EE. UU.), donde se utilizaron 58 canales de estaño de acuerdo con el sistema internacional 10-10 y una frecuencia de muestreo de 1000 Hz, se utilizó un filtro de línea que comprendía un filtro pasa-bandas (0,05 a 200 Hz) y un filtro de rechazo de banda (60 Hz) para eliminar el ruido de la fuente de alimentación. Además, la adquisición de referencia consistió

en un electrodo ubicado en el lóbulo de la oreja derecha, y se utilizó como tierra un electrodo ubicado entre el electrodo Cz y Fz [36].

Por otro lado, el centro de neurociencias de Cuba reportó un proyecto de registros tomados durante 1988-1990, en donde la muestra poblacional era de más de 600 sujetos, los cuales fueron seleccionados a partir de un procedimiento cuasialeatorio que buscaba garantizar una representación de edad adecuada de todo el rango de la muestra, la cual quedó conformada por 211 sujetos de los cuales 105 eran hombres y 106 eran mujeres, con un rango de edad entre los 5-80 años. Para la adquisición de estos datos se utilizó un electroencefalograma digital (MEDICID-3M) con ganancia de 10.000 y filtros de 0,3-30 Hz, filtro Notch de 60 Hz y un período de muestreo 10 Hz. Además, se utilizó el sistema internacional 10-20 para la distribución de los electrodos de plata sobre el cuero cabelludo, y el electrodo de referencia fue ubicado en el oído de forma monopolar [37].

Finalmente, se utilizaron los datos reportados en el proyecto de Modulación de respuesta Selectiva en el Departamento de Psicología de la Universidad de Oslo, Noruega. La adquisición de los registros de EEG fueron tomados en estado de reposo con los ojos cerrados durante cuatro minutos. Los datos fueron registrados a partir de un sistema BioSemi ActiveTwo, utilizando 64 electrodos siguiendo el esquema posicional del sistema 10-20 extendido (10-10). El conjunto de datos incluye 111 sujetos control sanos, de los cuales se tomó inicialmente una sesión, y posterior a la fecha de finalización se adiciono una segunda sesión, motivo por el cual no todos los sujetos se registraron las dos sesiones. Los datos fueron registrados a una frecuencia de muestreo de 1024 Hz, y únicamente se realizó un filtro de suavizado que poseía el hardware [38]. En la Tabla II, se encuentra la información sintetizada de las bases de datos de prueba, en donde se presentan las características principales de los datos y en la Tabla III y Tabla IV se encuentra información más específica de los grupos y sesiones que tiene cada una de las bases de datos de prueba.

Tabla II. Información genérica de las bases de datos de prueba.

	<b>BIOMARCADORES</b>	<b>CHBMP</b>	<b>SRM</b>
Fuente de adquisición	Grupo de Neuropsicología y Conducta (GRUNECO).	Centro de Neurociencias de Cuba, La Habana, Cuba.	Departamento de Psicología de la Universidad de Oslo, Noruega.
Número de sujetos	145	248	111
Número de EEGs	447	248	153
Edad (años)	21-65	18-68	17-71
Tarea	Reposo	Reposo Reactividad Hiperventilación Recuperación	Reposo
Sistema de montaje	10-10	10-10	10-20
Frecuencia de muestreo	1000 Hz	200 Hz	1024 Hz
Número de grupos	5	1	1
Número de sesiones	6	1	2

Tabla III. Información detallada de los grupos que contiene cada una de las bases de datos de prueba.

<b>Estudios/ Grupos</b>	<b>CTR</b>	<b>G1</b>	<b>G2</b>	<b>DCL</b>	<b>DTA</b>	<b>Total</b>
Biomarcadores	36	34	38	29	8	145
CHBMP	248	X	X	X	X	248
SRM	111	X	X	X	X	111
Número de sujetos						504

Tabla IV. Información detallada de las sesiones que contiene cada una de las bases de datos de prueba.

<b>Estudios/ Sesiones</b>	<b>V0</b>	<b>V1</b>	<b>V2</b>	<b>V3</b>	<b>V4</b>	<b>V4P</b>	<b>Total</b>
Biomarcadores	120	96	109	80	36	6	447
CHBMP	248	X	X	X	X	X	248
SRM	111	42	X	X	X	X	153
Número de registros EEG							848

## B. Tecnologías y herramientas

A continuación, se presenta las tecnologías definidas en el desarrollo del proyecto en donde se establecieron un conjunto de herramientas para la construcción de la aplicación, mostradas en la Fig 7 y justificadas en la Tabla V, en donde se encuentra una breve descripción del criterio final de selección de cada una de las tecnologías.

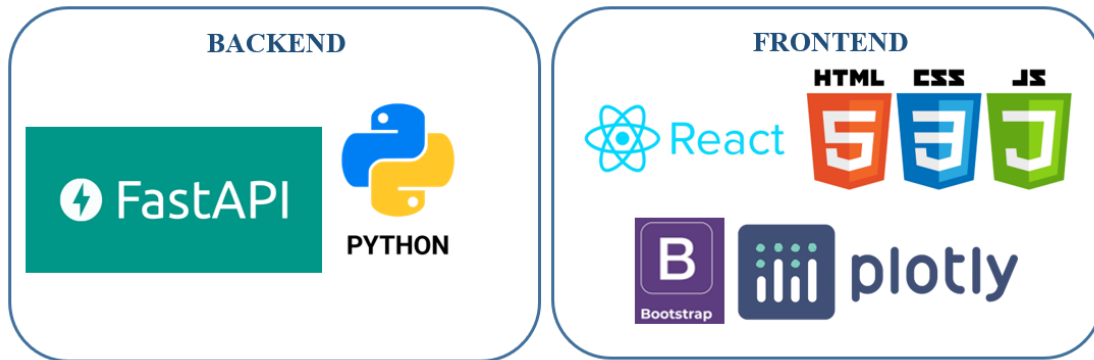


Fig 7. Tecnologías y herramientas para la construcción de la aplicación.

Tabla V. Tecnologías y herramientas empleadas en el desarrollo de la aplicación web.

Tecnología	Justificación
Python 3.x	Actualmente, es el lenguaje de programación más utilizado y compatible en el área de procesamiento de señales lo cual lo hace adecuado para el desarrollo de la plataforma.
FastAPI	FastAPI es un web framework de alto rendimiento que permite construir APIs basados en el estándar de Python, lo cual lo hace compatible para el desarrollo. Además, este framework está continuamente soportado, lo cual hace que sea muy estable.
JavaScript	JavaScript es un lenguaje de programación ligero e interpretado que permite una velocidad de ejecución rápida con el navegador y es altamente compatible y soportado.
React	ReactJS es una librería escrita en JavaScript de código abierto enfocada a la visualización para facilitar la creación de componentes interactivos e independientes al poseer una programación modular que ofrece funciones en forma de bloques libres llamados módulos.
CSS	Es un lenguaje de marcado que permite manejar el diseño y presentación de las páginas web. Además, permite almacenar diferentes estilos por separado permitiendo una mejor dinamización de la página.
Bootstrap	Framework CSS que permite estilizar los elementos de una página HTML, proporcionando interactividad a partir de componentes para facilitar la comunicación con el usuario.
Plotly	Librería de Python que permite generar gráficos interactivos, además se puede utilizar con código de JS lo que permite construir una interfaz del lado del frontend interactiva.



Por otro lado, en la Fig 8 se presenta el esquema de la metodología dividida en los dos componentes principales para la construcción de la plataforma, en donde el backend está subdividido en cinco subcomponentes principales, que van desde la adquisición, estandarización, preprocesamiento y post-procesamiento de los datos hasta el análisis de calidad de datos. A partir de estas etapas se definieron las interfaces necesarias que permiten usabilidad e interoperabilidad al usuario final.

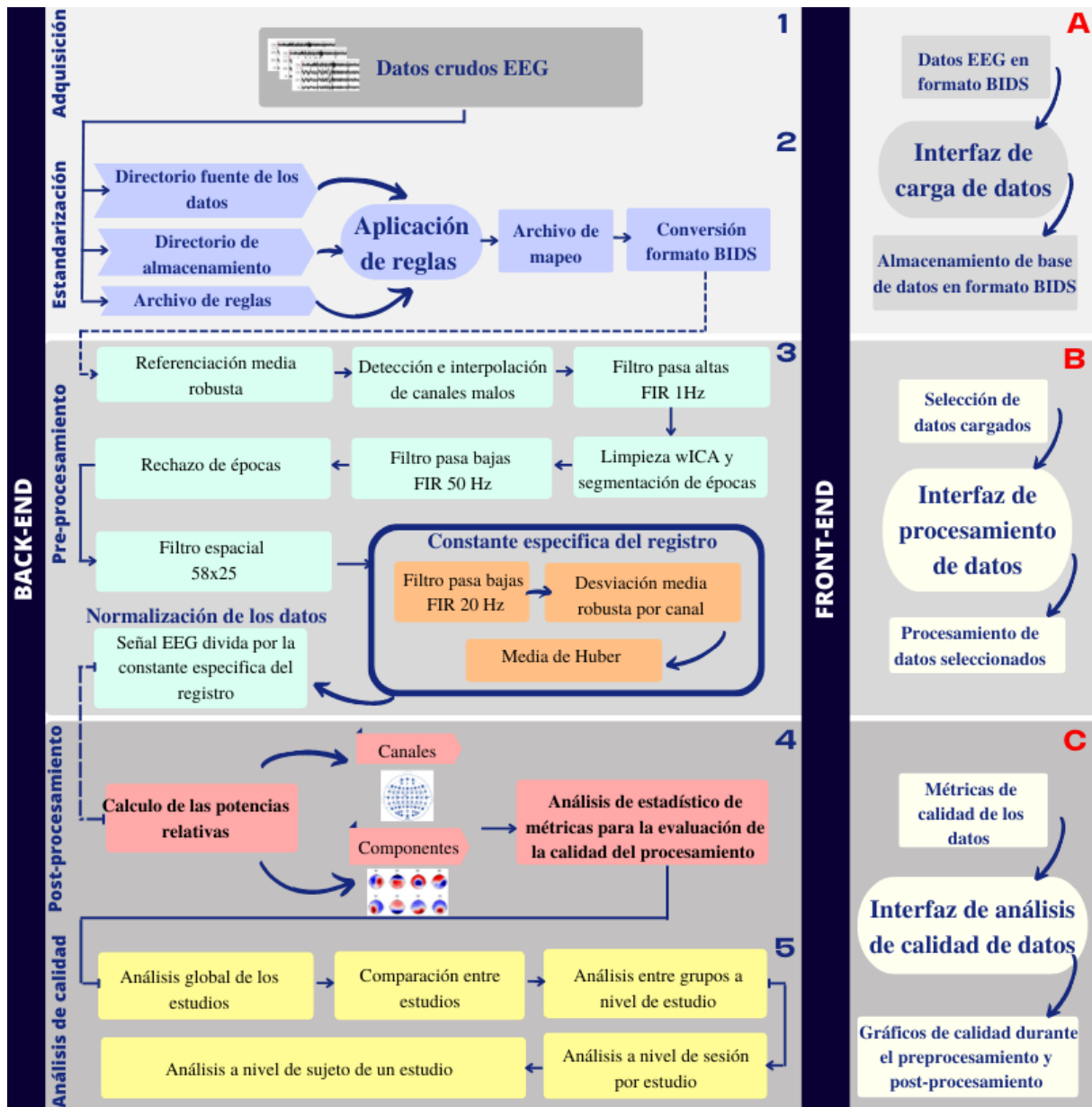


Fig 8. Esquema de metodología. **A:** 1) Adquisición de los registros EEG, 2) Estandarización formato BIDS, **B:** 3) Preprocesamiento de los registros, **C:** 4) Post-procesamiento de los registros, 5) Análisis de calidad de los datos.

### C. Control de calidad de los datasets

Para la evaluación y cuantificación de la calidad de los datos se tomaron en cuenta las siguientes métricas presentadas en la Tabla VI, VII y VIII, en donde se hace una evaluación cuantitativa de las principales etapas del procesamiento las cuales son: etapa del PREP, Wavelet-ICA y finalmente la evaluación del rechazo de épocas. Además, se encuentra el apartado de una columna, en donde se expande la explicación de que es lo que mide cada una de estas.

Tabla VI. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa del PREP.

<b>FLUJO DE PROCESAMIENTO EN ETAPA TEMPRANA (PREP)</b>	
<b>Métrica</b>	<b>Justificación</b>
Canales defectuosos tipo NAN <i>'Bad by NAN'</i>	Detección de canales que contengan datos tipo NAN.
Canales defectuosos tipo planos <i>'Bad by flat'</i>	Localización de canales cuando aparece plana la señal en relación con los otros canales cuando se presenta un gráfico de desplazamiento.
Canales defectuosos por desviación <i>'Bad by deviation'</i>	Se estima las diferencias de amplitud entre los conjuntos de datos para clasificar como un canal malo por desviación. Sin embargo, no identifica los canales que capturan los parpadeos y actividad muscular como ruidosos.
Canales defectuosos por ruido de alta frecuencia <i>'Bad by hf noise'</i>	Se detectan aquellos canales que tengan ruido de alta frecuencia.
Canales defectuosos por correlación <i>'Bad by correlation'</i>	Detección de canales malos de acuerdo con la correlación máxima, si es inferior a un umbral (0,4 por defecto) para un determinado porcentaje de las ventanas (1% por defecto) se identifica como un canal malo.
Canales defectuosos por relación señal-ruido <i>'Bad by SNR'</i>	Se identifican canales defectuosos como aquellos que tengan una baja grabación de SNR.
Canales defectuosos por abandono <i>'Bad by dropout'</i>	Identificación de canales defectuosos por abandono.
Canales defectuosos por consenso de muestra aleatoria <i>'Bad by ransac'</i>	Detección de canales defectuosos a partir del método iterativo RANSAC (Random Sample Consensus), el cual a partir de un conjunto de observaciones calcula los parámetros de un modelo matemático.
Canales defectuosos en general <i>'Bad all'</i>	Canales que presentan fallas en general.

Tabla VII. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa de Wavelet-ICA.

<b>LIMPIEZA WAVELET JUNTO CON LA TÉCNICA DE ANÁLISIS DE COMPONENTES INDEPENDIENTES (WICA)</b>	
<b>Métrica</b>	<b>Justificación</b>
Relación de componentes filtradas sobre componentes totales	Detección y estimación del porcentaje filtrado de componente en relación con el número de componentes totales. Esta métrica informa que tanto se están filtrando los datos en esta etapa.

Tabla VIII. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa de rechazo de épocas ruidosas.

<b>RECHAZO DE ÉPOCAS RUIDOSAS</b>	
<b>Métrica</b>	<b>Justificación</b>
Curtosis	Medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.
Amplitud	Medida física que muestra la variación máxima del desplazamiento.
Tendencias lineales	Línea recta que se usa con conjuntos de datos lineales. Los patrones lineales pueden indicar la tendencia lineal, en donde se puede presentar un incremento o decremento de un ritmo constante.
Potencia espectral	Medición que informa la distribución de potencia espectral sobre los ritmos de frecuencia de una señal.

Además, se incluye una etapa de calidad de los datos después del preprocesamiento, en donde se mide la calidad de acuerdo con la cantidad de outliers y la variabilidad según la característica analizada. Para este caso se evaluó la potencia relativa en cada una de las bandas de frecuencia de acuerdo con la clasificación IPEG 2012 mostrada previamente en la Fig 2. La extracción de esta característica se realizó utilizando dos enfoques. El primero es analizar sobre las componentes neuronales clasificadas y etiquetadas en el estudio de García et al (Fig 9) [36] y luego realizar un enfoque sobre la distribución de la energía a nivel general de los canales.

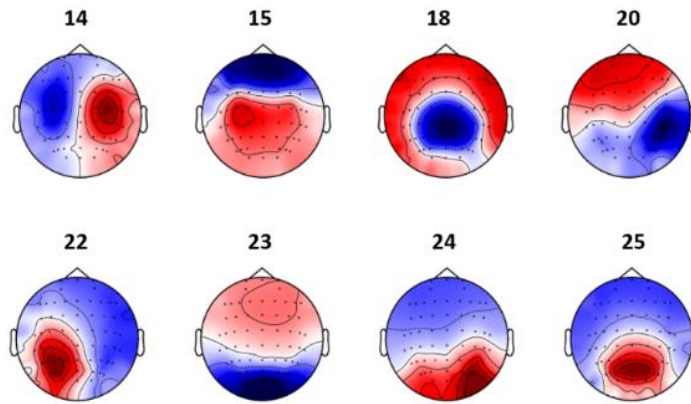


Fig 9. Proyecciones de mapas del cuero cabelludo de componentes gICA encontrados etiquetados como neuronales.

## V. RESULTADOS Y DISCUSIÓN

### A. Estandarización de los datos

Inicialmente, para el procesamiento de los datos, es necesario realizar la estandarización de los metadatos de prueba aplicando la conversión a BIDS, para esto se utilizó el paquete de Python “sovabids” que se encuentra alojado en un repositorio de GitHub de acceso libre. Este permite la conversión de datos EEG de forma automática, y para la aplicación de este paquete se requiere la configuración de un archivo de “reglas” en la cual se establece la nomenclatura de acuerdo con la codificación que tenga cada conjunto de datos. A continuación, se presenta un ejemplo en donde se configuró este archivo, específicamente para la base de datos de *Biomarcadores*.

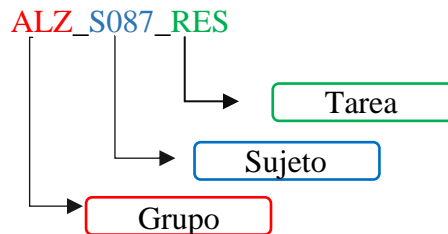
```
dataset_description:                               # Configuring the
dataset_description.json file                     #
  Name : Biomarcadores                            # Name of the dataset, set up
as a fixed string                                #
  Authors:                                        # Here I put the personnel
involved in the acquisition of the dataset       #
  - Veronica Henao Isaza                          # See
http://fcon\_1000.projects.nitrc.org/indi/retro/MPI\_LEMON.html

sidecar:                                          # Configuring the sidecar eeg
file                                              #
  PowerLineFrequency : 50                         # Noted from the visual
inspection of the eeg spectrum                   #
  EEGReference : Mastoide derecha                # As mentioned in
https://www.nature.com/articles/sdata2018308

channels:                                        # Configuring the channels tsv
  type :                                          # This property allow us to
overwrite channel types inferred by MNE         #
  VEO : VEOG                                     # Here the syntax is <channel
name> : <channel type according to bids notation>
  HEO : HEOG                                     # Here we set the type of F3,
it was already correctly inferred by mne but it is included to illustrate
retying of various channels.

non-bids:                                        # Additional configuration not
belonging specifically to any of the previous objects
  eeg_extension : .cnt                           # Sets which extension to read
as an eeg file
  path_analysis:                                 # Some bids properties can be
inferred from the path of the source files
  pattern : Codificado EEG V0 V1 V2 V3 / (.+) / (.+ .{3}) (.+) (.+) .cnt # For
example here we extract from the path the "subject" child of the "entities"
object
  fields :
    - entities.session
    - entities.subject
    - ignore
    - entities.task
```

Los campos principales que se configuran para esta base de datos son: *sesiones*, *sujetos* y *tareas*. Para dar claridad a estos campos establecidos, se muestra la codificación de un archivo en donde ‘ALZ’ corresponde al grupo de prueba, ‘S087’ corresponde al participante con el número correspondiente, y ‘RES’ hace referencia a la tarea que se encontraba realizando el sujeto en el momento de la captura de los datos; en este caso hace referencia a la adquisición de la señal en condición de reposo.



Una vez configurado, se procede hacer la conversión como lo indica la documentación de la librería. La instalación de este paquete se realiza ejecutando las siguientes líneas de código proporcionadas por la guía de instalación del repositorio.

```
# Installation package sovabids
git clone https://github.com/yjmantilla/sovabids.git
cd sovabids
pip install -r requirements-user.txt
```

Al finalizar la conversión se obtiene el siguiente esquema presentado en la Fig 10, en donde se obtiene un folder por sujeto y en cada uno de estos se encuentra el número de sesiones que tenga registrado dicho participante. En cada sesión existe una carpeta ‘eeg’ en donde se encuentra los canales en formato ‘tsv’ y los registros en diferentes extensiones.

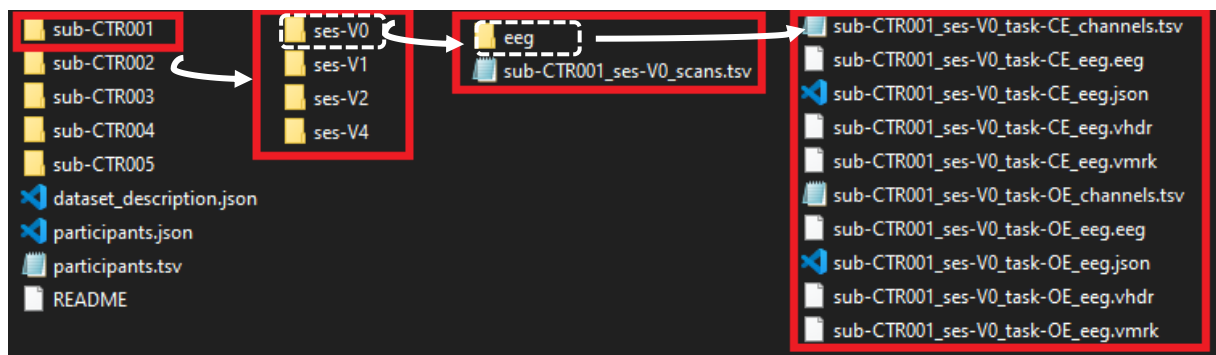


Fig 10. Estructura de los datos crudos después de aplicar el estándar BIDS.

De acuerdo con los resultados obtenidos mediante este proceso de estandarización, es posible estructurar los datos de una forma más organizada para el intercambio de la información y así fomentar procesos de preparación de los datos para la construcción de bases de datos de EEG a gran escala. Esta herramienta posibilita el desarrollo del metanálisis multimodal para realizar flujos de trabajo neurocientíficos que permitan una exploración de datos EEG de una forma más versátil.

### B. Procesamiento de los registros EEG

Para implementar el procesamiento en la plataforma, fue necesario instalar ciertos paquetes específicos del protocolo que se trabaja dentro del laboratorio de GRUNECO. Estos paquetes se encuentran en el repositorio de GitHub y la mayoría de estos son de acceso privado, por tanto, en la Fig 11 se muestra esquemáticamente cuales paquetes se utilizaron y las funciones principales que cada uno de estos posee dentro del flujo de preprocesamiento mencionado anteriormente.

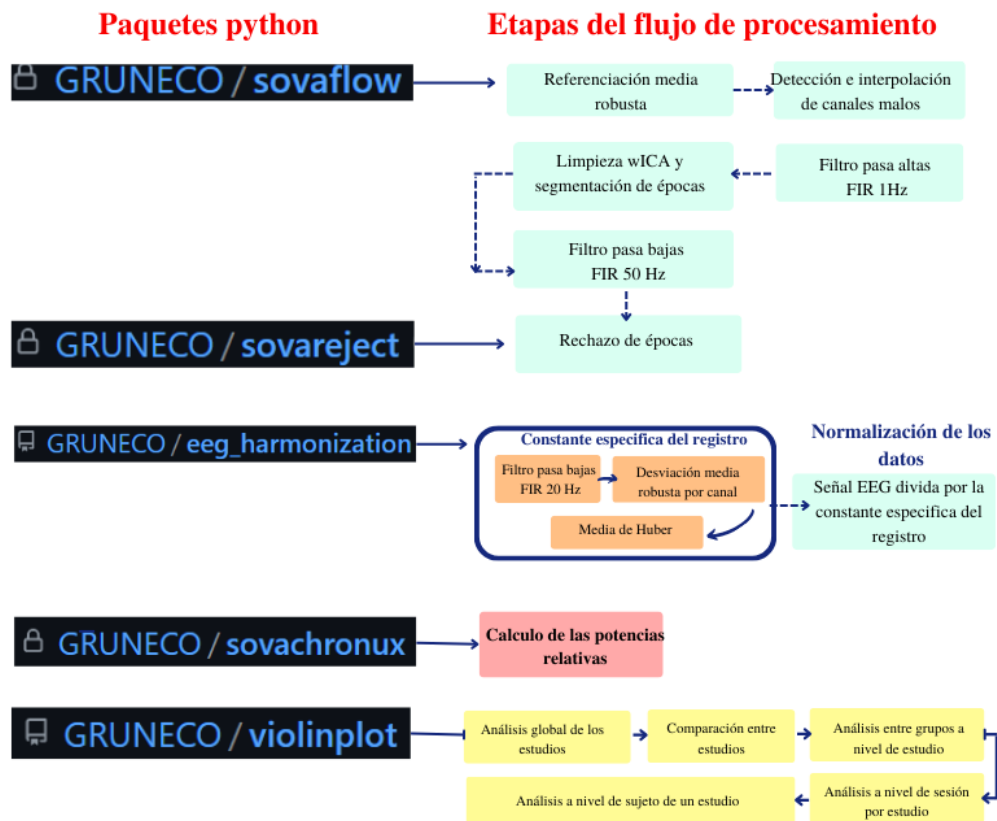


Fig 11. Sistema de paquetes para el procesamiento de datos EEG.

La instalación de cada uno de estos paquetes se realizó utilizando las siguientes líneas de código presentadas a continuación. Los paquetes ‘eeg\_harmonization’ y ‘violinplot’ se instalan en modo desarrollo para la construcción de cada uno de los procesos necesarios.

```
# Installation package sovaflow
pip install git+https://enlace_repositorio/sovaflow.git

# Installation package sovareject
pip install git+https://enlace_repositorio/sovareject.git

# Installation package eeg_harmonization
git clone https://enlace_repositorio/eeg_harmonization.git
cd eeg_harmonization
pip install -r requirements-install.txt

# Installation package sovachronux
pip install git+https://enlace_repositorio/sovachronux.git

# Installation package violinplot
git clone https://enlace_repositorio/violinplot.git
cd violinplot
pip install -r requirements-install.txt
```

La unificación de este flujo se implementó a partir de la función ‘*harmonize*’, la cual realiza cada una de las importaciones y genera los archivos derivados de cada una de estas etapas relevantes del procesamiento: PREP, Wavalet-ICA y rechazo de épocas ruidosas. El almacenamiento de los derivados se realizó conservando el estándar BIDS almacenado en los formatos ‘*json*’ y ‘*txt*’ siguiendo la nomenclatura del modelo, de igual modo, se incluye las señales con la extensión ‘*fif*’ durante cada una de las etapas más representativas para el análisis posterior. El objetivo de esta metodología es mantener la interoperabilidad y la reutilización de los datos.

En la Fig 12, se presenta un ejemplo como resultado de los derivados del procesamiento en donde se encuentra detallado las etapas de interés del preprocesamiento siguiendo el estándar BIDS, además de extraer características como potencia relativa con y sin la etapa de normalización.



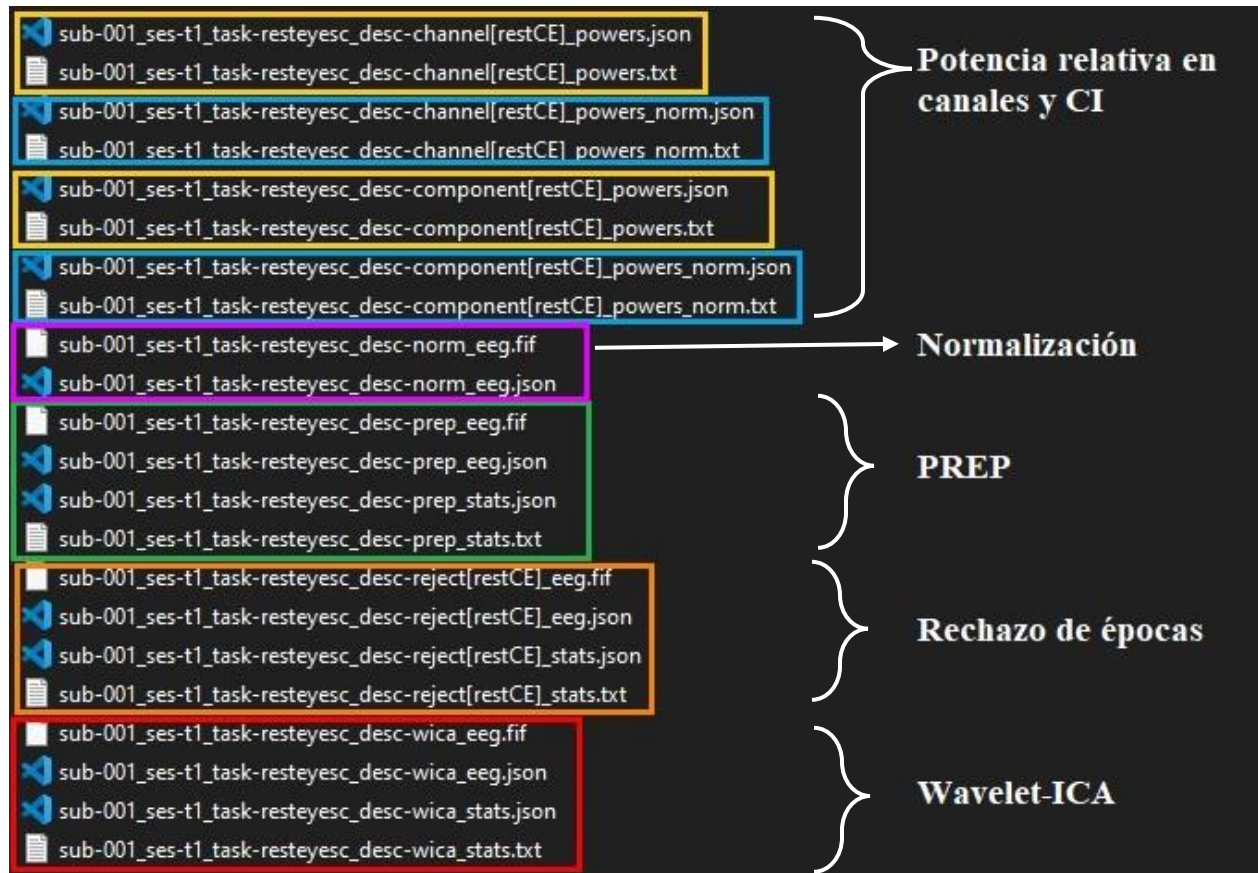


Fig 12. Derivados del procesamiento según la nomenclatura y la extensión del estándar BIDS.

La etapa de normalización fue una etapa adicional que se añadió al flujo de trabajo con el que se venía procesando las señales. Esta etapa se siguió bajo la metodología presentada por Nima et al [27], donde evaluaron diferentes tipos de normalizaciones, obteniendo con mejores resultados la normalización de Huber.

En la Fig 13, se presenta el script implementado para aplicar dicha normalización sobre los datos, donde inicialmente se realiza un filtro pasa bajas de 20 Hz a la señal de la etapa anterior, es decir, la señal al finalizar el proceso de rechazo de épocas ruidosas; luego de esto se calcula la desviación estándar por canal y finalmente se encuentra la media de Huber; este conjunto de pasos da a lugar a la constante específica del registro. La normalización consiste en dividir la señal sobre la constante previamente encontrada.

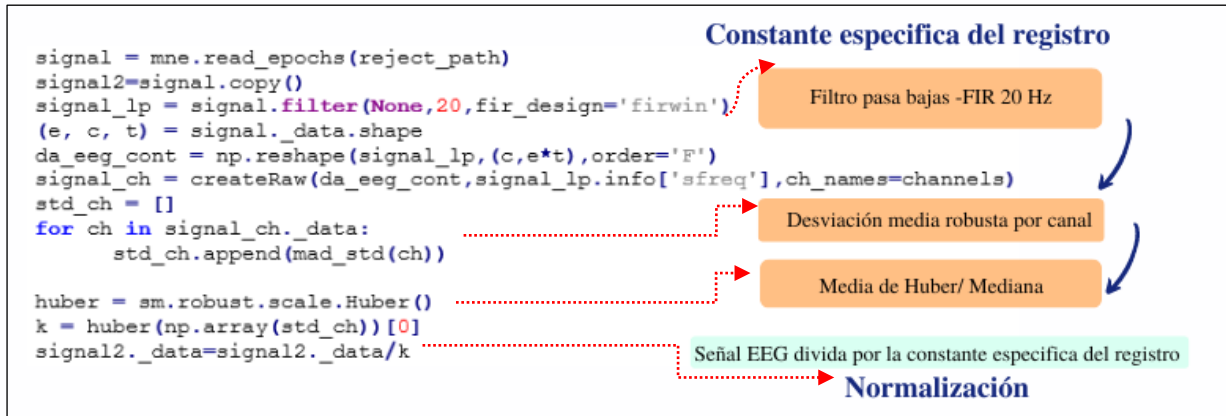


Fig 13. Script de normalización de Huber en Python.

La información fue almacenada para cada una de las etapas de preprocesamiento y post-procesamiento en dataframes en formato *feather*, teniendo en cuenta las mejores características para almacenamiento de búfer. El enfoque de almacenamiento en esta extensión permitió tener una mayor rapidez de lectura de los datos en comparación con formatos como 'csv' o 'xmls', lo cual fue necesario dado que esta información se emplea para generar los diferentes análisis estadísticos y el control de calidad de los datos que se generan sobre demanda.

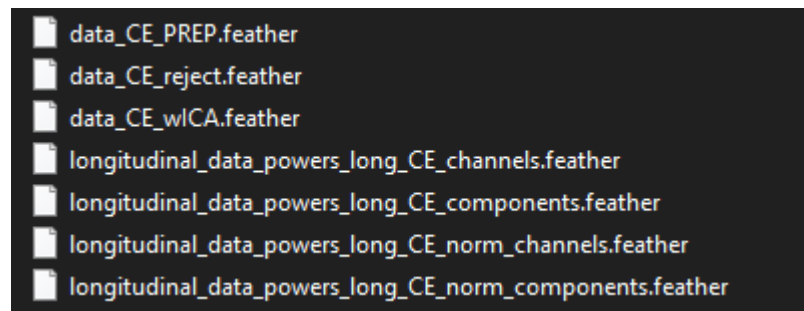


Fig 14. Almacenamiento de métricas de calidad.

Una vez organizado cada uno de los paquetes y definido el formato de almacenamiento, se procesan las bases de datos de prueba. Sin embargo, los registros EEG son datos muy robustos que toman tiempo para procesar dependiendo de las características, como lo es la duración del paradigma, el número de condiciones registradas y otros aspectos que influyen en el procesamiento de la información. Para ilustrar el tiempo que tarda en realizarse este proceso, se presenta la Tabla IX, en donde se muestra la duración de las diferentes bases de datos tanto de la etapa del

procesamiento como de la creación de los dataframes que contiene la información de las métricas para el análisis de control de calidad de los datos.

Tabla IX. Duración de tiempo para el procesamiento de la información del conjunto de datos de prueba.

	Tiempo de procesamiento [s]	Tiempo de creación de dataframes [s]
Biomarcadores	182002,114757	888,875
CHBMP	91093,5266378	509,99293720000423
SRM	93745.9407302	542.80145739999716

De acuerdo con los tiempos obtenidos para el procesamiento, es posible identificar que estos procesos son muy extensos, motivo por el cual es necesario tener en cuenta en el desarrollo de la aplicación, en donde se busquen estrategias para la gestión de datos y para la notificación de procesos terminados, ya que estos pueden tardar incluso días.

### C. *Control de calidad de datos*

El control de calidad de datos se implementó realizando un análisis desde lo más global hasta lo más específico, en donde se buscaba analizar el proceso de calidad del flujo de procesamiento trabajado en el laboratorio, además de comparar como es la calidad de los registros tomados en el grupo de investigación en comparación con otras bases de datos.

Para realizar este análisis se utilizó diagramas de violín y diagramas de cajas y bigotes en donde se grafica cada una de las métricas presentadas en la Tabla VI. Definición de las métricas para evaluación de la calidad de la base de datos en la etapa del PREP. Tabla VI, VII y VIII. Las secciones están clasificadas en cuatro enfoques:

1. **Análisis general:** Aquí se unifican todas las bases de datos en un solo diagrama de violín, para ver las tendencias y si la distribución de los datos comparte diferencias o semejanzas entre sí.
2. **Análisis entre estudios:** Se discrimina cada una de las bases de datos, para identificar similitudes o diferencias entre los estudios.

3. **Análisis entre sesiones de un estudio:** Este análisis solo es válido para estudios longitudinales, ya que permite evaluar en el tiempo que tan diferentes son los datos.
4. **Análisis entre grupos de un estudio:** Se compara los datos entre los diferentes grupos con la finalidad de encontrar diferencias. Este enfoque suele ser útil para comparar datos control en relación con otro grupo que se esté analizando.

El paquete ‘sovaviolin’ de Python fue desarrollado con la finalidad de crear un módulo de forma organizada y genérica pensando en la reutilización de este. Este paquete está dividido según las etapas de análisis del procesamiento de datos y a su vez cada etapa contiene funciones de los cuatro enfoques presentados previamente. Para mostrar los diferentes gráficos en la interfaz se codificaron las imágenes con el formato base64 y una vez la interfaz lee los datos se decodifica la información. De esta manera se puede generar gráficos bajo demanda, ya que en este caso almacenar la información no era una opción óptima, debido a que los gráficos dependen de varios parámetros que son variables. A continuación, se presentan los gráficos definidos para el análisis y control de calidad de datos en los enfoques descritos anteriormente.

### *Preprocesamiento*

Para el análisis del preprocesamiento se evalúan las tres etapas principales del flujo, en donde inicialmente se evalúa la calidad de los datos al aplicar el PREP, luego al utilizar herramientas como wICA que utilizan modelos de aprendizaje profundo para la identificación y clasificación de artefactos, y finalmente se aplica el rechazo sobre las épocas ruidosas la cual está basada en cuatro criterios principales: tendencias lineales anormales, espectros de potencias anormales, actividad estadísticamente atípica y amplificaciones de la señal extrema.

En los diferentes enfoques de análisis planteados se encuentra inicialmente la transición de los datos una vez realizado la etapa del PREP, en donde se muestra la información de los datos sin procesar, antes de realizar la interpolación y luego de esta. Se observa que a partir de este proceso hay un patrón que muestra una tendencia de disminución de la media de la señal cuando se da el proceso de interpolación. El objetivo de este paso es poder identificar si existen muchos canales

dañados en la etapa de la adquisición, por lo cual es ideal que se mantenga la longitud en cero de las métricas evaluadas en esta fase.

Luego de esta etapa, se aplica un filtro pasa altas de 1Hz el cual tiene como objetivo remover las tendencias de baja frecuencia de la señal, debido a que la técnica del ICA es sensible a estas. Seguido de este filtrado se realiza la etapa del wICA el cual busca filtrar artefactos para conservar únicamente la información neuronal. Para analizar la calidad del procesamiento en esta etapa se planteó calcular el porcentaje de componentes filtradas determinada como el número de componentes filtradas sobre el número de componentes totales. En esta métrica se espera que el porcentaje de filtrado no sea muy alto dado que en esta etapa ya se aplicaron previamente procesos de mejoramiento de la señal buscando obtener únicamente la información neurológica.

Posterior a esta etapa se realiza un filtro pasa bajas de 50 Hz el cual se realiza para atenuar el rango de frecuencia de interés que se encuentra entre 0.5 y 45 Hz. La etapa de rechazo de épocas ruidosas se aplica una vez la señal ha sido filtrada, en donde se calculan los criterios antes de aplicar la etapa y después de esta, con la finalidad de observar su efecto sobre la señal. Las épocas se segmentaron con una ventana de tiempo de cinco segundos, y la selección de su tamaño tiene en cuenta el proceso de cuantificación de los ritmos lentos y la estacionariedad de la señal.

### *1. Análisis global*

La Fig 15 comprende cada una de las etapas que hacen parte del análisis global de los datos durante el preprocesamiento, en donde cada una de las métricas contiene información de la mezcla de las tres bases de datos de prueba. Estas gráficas, permiten identificar que tan diferentes son los datos entre ellos a nivel general, en donde por medio de las representaciones estadísticas de violín obtenemos información de donde se encuentra la media, si existen muchos datos atípicos y la distribución de estos.

En la Fig 15A, se encuentra la etapa del PREP, en donde el eje X representa el estado de la señal durante la aplicación de esta etapa; el eje Y es el valor de la longitud de cada métrica, es decir es la magnitud de canales encontrados por las diferentes variables evaluadas. Luego se encuentra

la siguiente etapa, wICA, donde el eje X hace referencia a las componentes, y el eje Y es el número de componentes filtradas sobre totales. La etapa de rechazo de épocas está dada por el valor de cada métrica en el eje Y, y cada una de las columnas hace referencia a los criterios de evaluación.

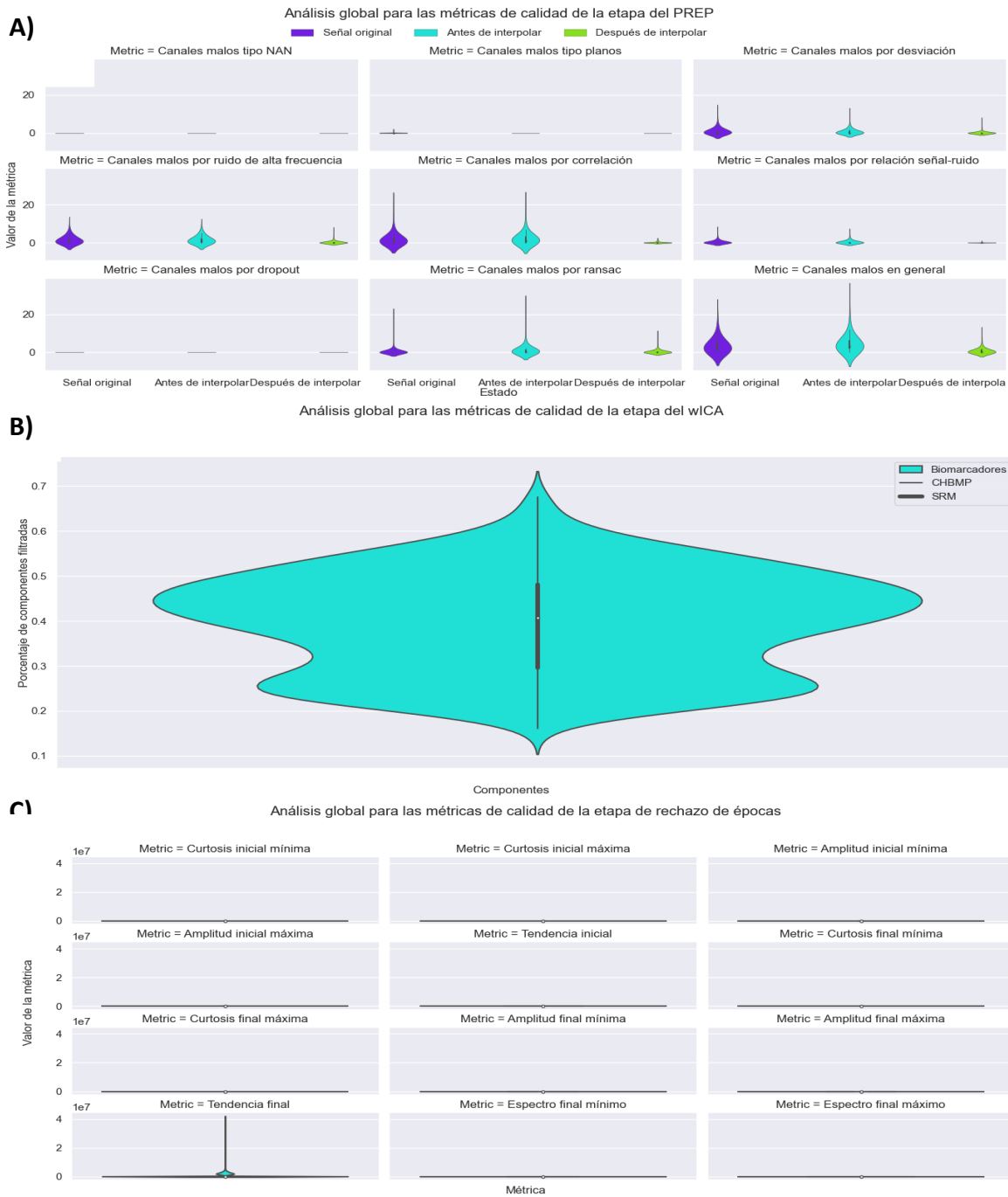


Fig 15. Análisis global para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.

Los resultados previamente mostrados, indican que las bases de datos a pesar del proceso de mezcla poseen una alta concentración de los sujetos sobre una misma media, lo cual demuestra que las bases de datos a pesar de factores como año de registro, población, dispositivo de captura, rango de edad y demás variables demográficas presentan un rango en común. En la etapa del PREP, se observa que se identificaron canales malos por desviación, ruido de alta frecuencia, correlación, SNR, RANSAC y canales en general malos, cada métrica tiene la longitud de estos canales.

En la Fig 15B la fase del wICA, muestra dos concentraciones diferentes, una en donde la mayoría de los sujetos se filtró las componentes un 45% y en otra un 25%. Este gráfico nos permite conocer las diferencias entre el rango de componentes filtradas, sin embargo, no se especifica la concentración de las componentes a cuál de las bases de datos mezclada corresponde. Por otro lado, en la Fig 15C se presenta las métricas de rechazo de épocas en donde se muestra que la mayoría están cercanos a cero.

## *2. Análisis entre estudios*

En la Fig 16 se encuentra la comparación entre las bases de datos en donde se discrimina cada una de las poblaciones registradas. Este enfoque permite comprender que tan diferente son las señales capturadas procesándolas con un mismo flujo de trabajo, ya que cada una de estas fue tomada tanto con poblaciones y dispositivos diferentes.

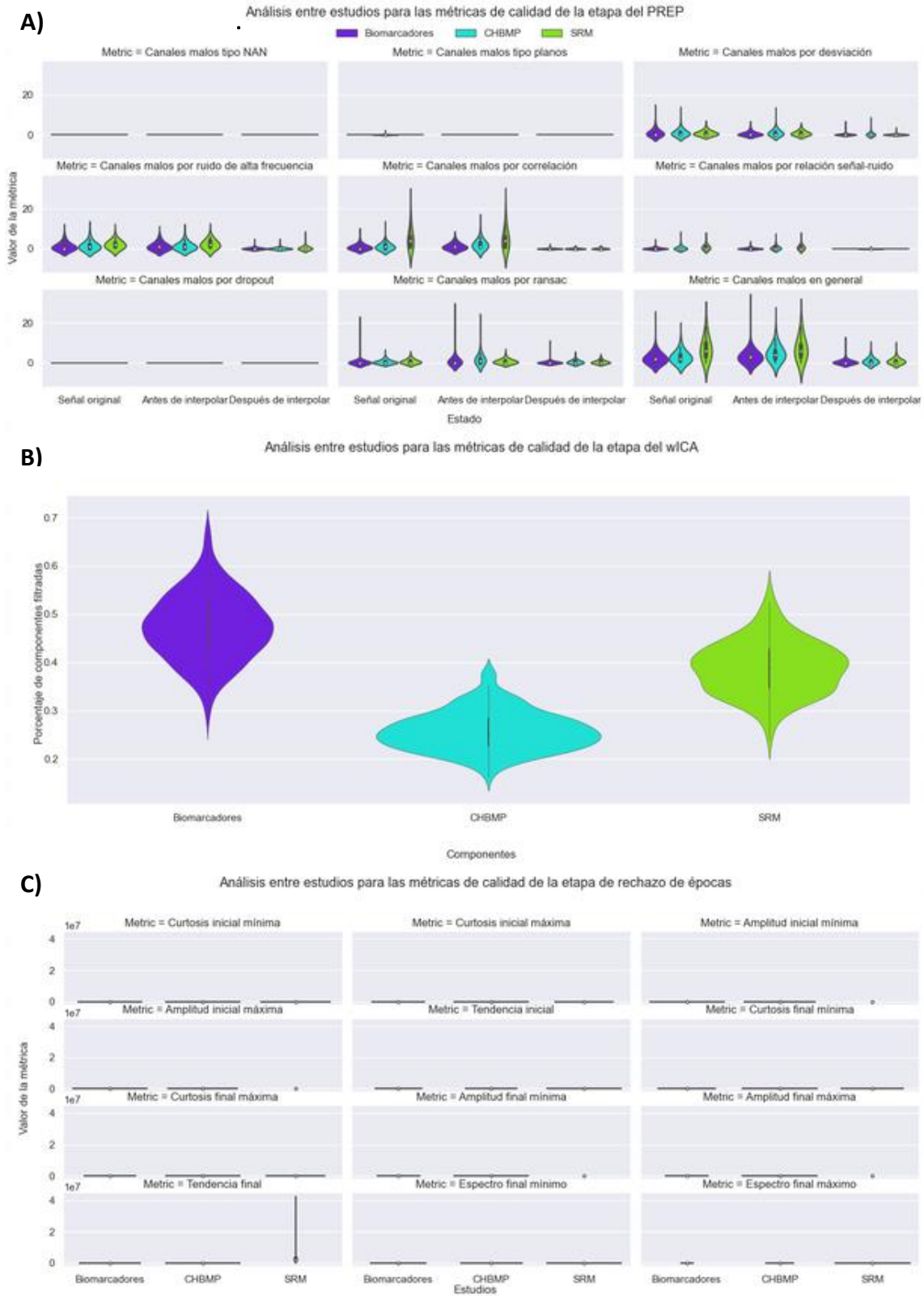


Fig 16. Análisis entre estudios para las diversas métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.



De acuerdo con los resultados obtenidos mediante este enfoque es posible identificar que a pesar de que las poblaciones de registro son diferentes, la adquisición según la distribución de los datos en la etapa del PREP son similares. Además, en la etapa del wICA se observa que alrededor del 45% en la mayoría de los sujetos de la base de datos de *Biomarcadores* fue filtrado las componentes, mientras que para la base de datos CHBMP, tan solo fue de aproximadamente el 25% y para SRM un valor cercano al 40%. Esto quiere decir que el clasificador que utiliza la técnica del wICA identificó un mayor número de componentes artefactuales en la población registrada en Colombia. A diferencia del análisis global, este enfoque permite identificar el porcentaje filtrado a cuál conjunto de datos corresponde.

Por otra parte, en el proceso de rechazo de épocas ruidosas según las métricas evaluadas, tanto al inicio como al final de esta, se mantiene valores muy pequeños con respecto a la escala del eje Y, el cual indica el valor de cada métrica evaluada.

### 3. *Análisis entre grupos*

De acuerdo con la información mostrada en la Tabla III, la única base de datos analizada que posee diferentes grupos es la base de datos de *Biomarcadores*, las demás solo tienen un grupo de participantes conocido como grupo control, por tal razón, únicamente se ilustra las gráficas que hagan parte de este conjunto de datos. Sin embargo, en la aplicación se puede seleccionar cualquiera de las bases de datos cargadas para que el usuario interactúe con la interfaz e identifique si este tipo de análisis es adecuado para el conjunto de datos que desee analizar. Las etapas de este enfoque se encuentran organizadas en la Fig 17.

El conjunto de datos de *Biomarcadores* está constituido por cinco grupos registrados, lo cual permite identificar si el tipo de patología o enfermedad afecta en el proceso de calidad de los datos, para esto se comparan los diferentes grupos con respecto a una referencia, que para este caso corresponde al grupo ‘CTR’ conocido como grupo control.

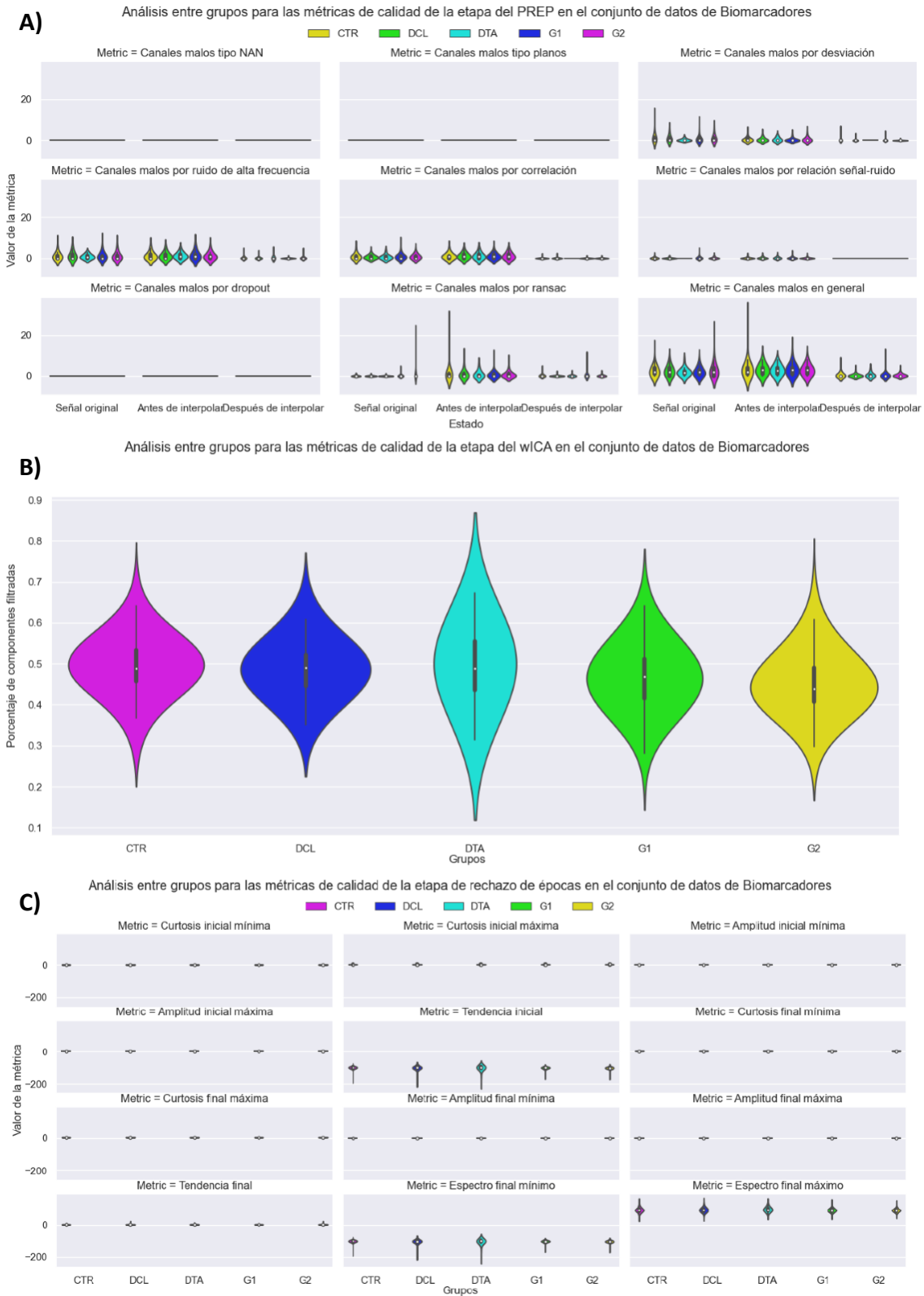


Fig 17. Análisis entre grupos de las diversas métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.

A partir de estas técnicas de inspección visual es posible determinar que los grupos siguen patrones muy similares entre cada etapa de procesamiento. Una de las etapas más notorias es la etapa del wICA (Fig 17B), en donde se observa que el número de componentes que se filtró para cada uno de los grupos se mantiene homogéneo entre el 40% y el 50% de componentes artefactuales a eliminar.

Los resultados de este enfoque muestran que el procesamiento, independiente del grupo de población registrada sigue un comportamiento similar si se comparan entre estos, lo cual es un buen resultado ya que hay grupos de la población más complicadas de registrar. Sin embargo, es necesario confirmar la información mediante pruebas estadísticas cuantitativas que informen la significancia estadística del reporte por inspección visual o realizar modelos de automatización que permitan determinar si un registro de una población con respecto al otro tiene valores de porcentajes cercanos de calidad, con la finalidad de descartar información atípica y evitar sesgar los resultados.

#### *4. Análisis entre sesiones*

Las bases de datos que reportan datos longitudinales son Biomarcadores y SRM, motivo por el cual únicamente se presenta la información asociada a estos datos (Fig 18 y Fig 19). La captura de registros en el tiempo permite medir la reproducibilidad y la repetibilidad de los procesos, por lo cual es necesario realizar validaciones y controles sobre cómo cambia procesar los datos de una señal de un sujeto que se capturo en un intervalo de tiempo definido por cada estudio.

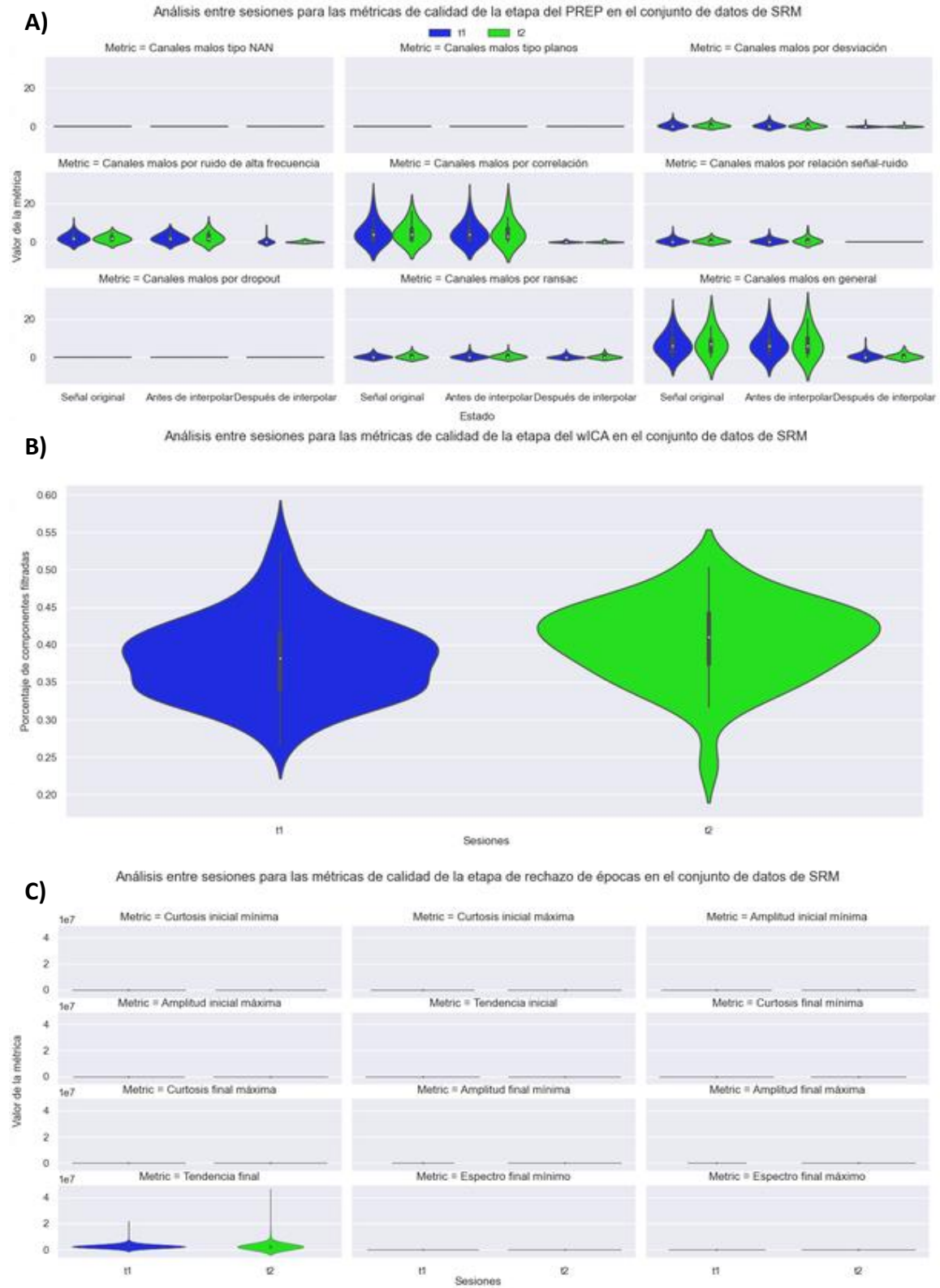


Fig 18. Análisis entre sesiones en el conjunto de datos SRM para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.

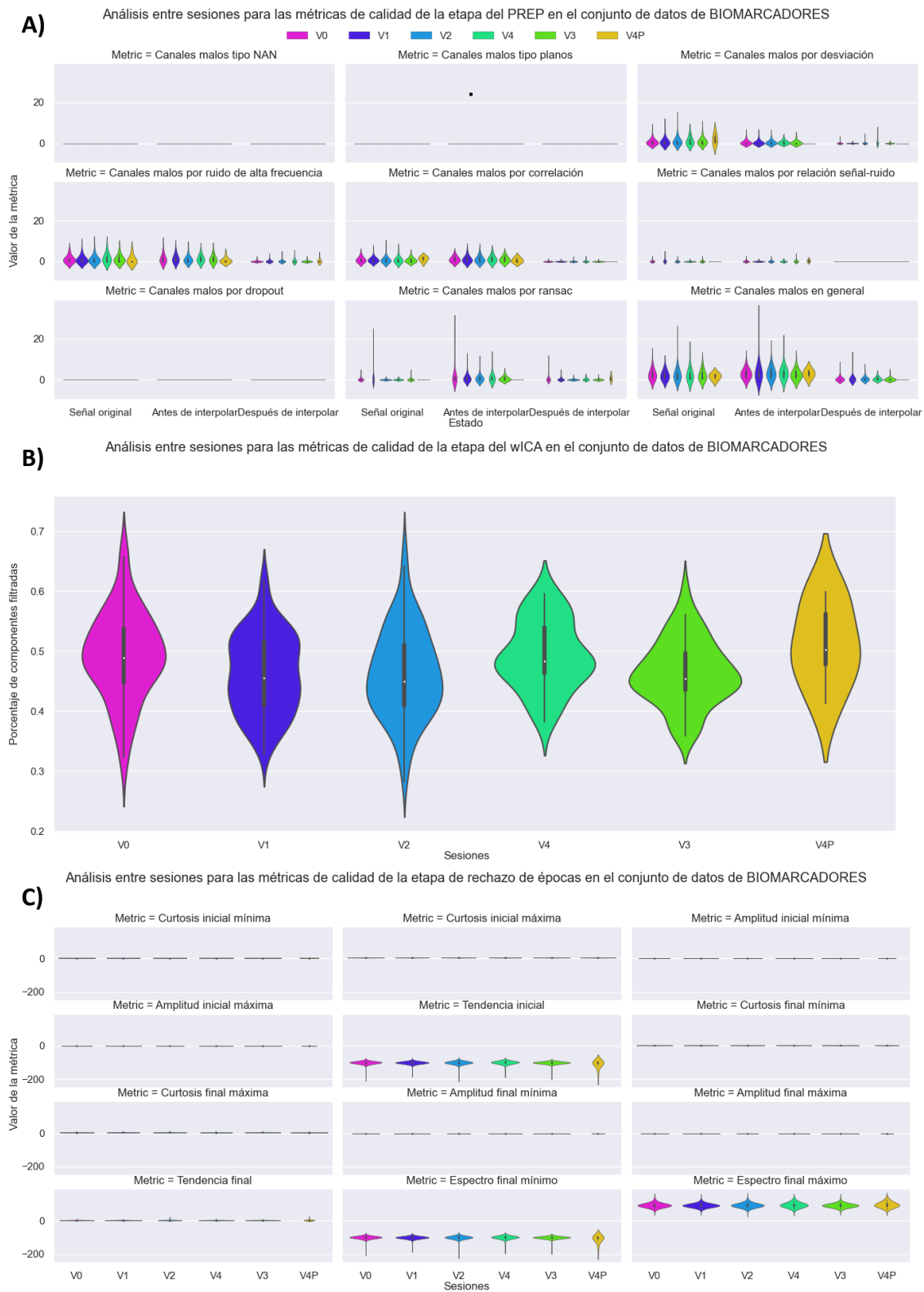


Fig 19. Análisis entre sesiones en el conjunto de datos BIOMARCADORES para las diferentes métricas de calidad. A) La etapa del PREP; B) etapa del wICA; C) etapa de rechazo de épocas ruidosas.

Con base a los resultados mostrados en las Fig 18 y Fig 19, es posible observar que ente sesiones no hay cambios relevantes, sin embargo, es necesario hacer pruebas estadísticas cuantitativas que permitan medir la reproducibilidad y repetibilidad de los resultados, con la finalidad de obtener la significancia estadística, donde se demuestre la validez y el porcentaje de error de estos.

En general, en la etapa del PREP se siguen obteniendo canales malos por desviación, ruido de alta frecuencia, correlación, SNR, RANSAC. Sin embargo, la longitud de estos no supera 20 canales y los sujetos que tienen este valor son datos atípicos en el conjunto de datos.

### *Post-procesamiento*

Una vez realizado el procesamiento de los datos, es posible realizar análisis de características. En este caso, se evalúa la potencia relativa en las diferentes bandas de frecuencia mencionadas anteriormente. No obstante, también se estima las diferencias entre una etapa adicional de normalización y el flujo de trabajo previamente realizado en dos enfoques; Inicialmente se hace un análisis por canales, en donde se tiene en cuenta la energía que aporta los diferentes electrodos en cierto rango de frecuencia y análisis a nivel de componentes etiquetadas y clasificadas como neuronales. Este tipo de observaciones resultan muy útiles para identificar patrones o encontrar biomarcadores que permitan avanzar en los desarrollos de investigación de alguna patología en específico.

En esta fase también se realizan los mismos enfoques de análisis desde lo más general hasta lo más específico, es decir, primero se analizan los datos de forma global, luego se comparan a nivel individual entre estos, y según el conjunto de datos, si posee grupos o sesiones se realizan las debidas observaciones.

#### *1. Análisis global*

La Fig 20, representa la energía capturada sobre el cuero cabelludo por los electrodos después de realizar una serie de filtros para la obtención de la actividad cerebral. En esta gráfica se

compara en cada una de las bandas de frecuencia el proceso de normalización con respecto al procesamiento realizado. En la Fig 20A se presenta la potencia relativa en canales, mientras que en la Fig 20B, se presenta la misma información organizada por columnas, en donde cada elemento de la matriz corresponde a una componente gICA etiquetada como neuronal.

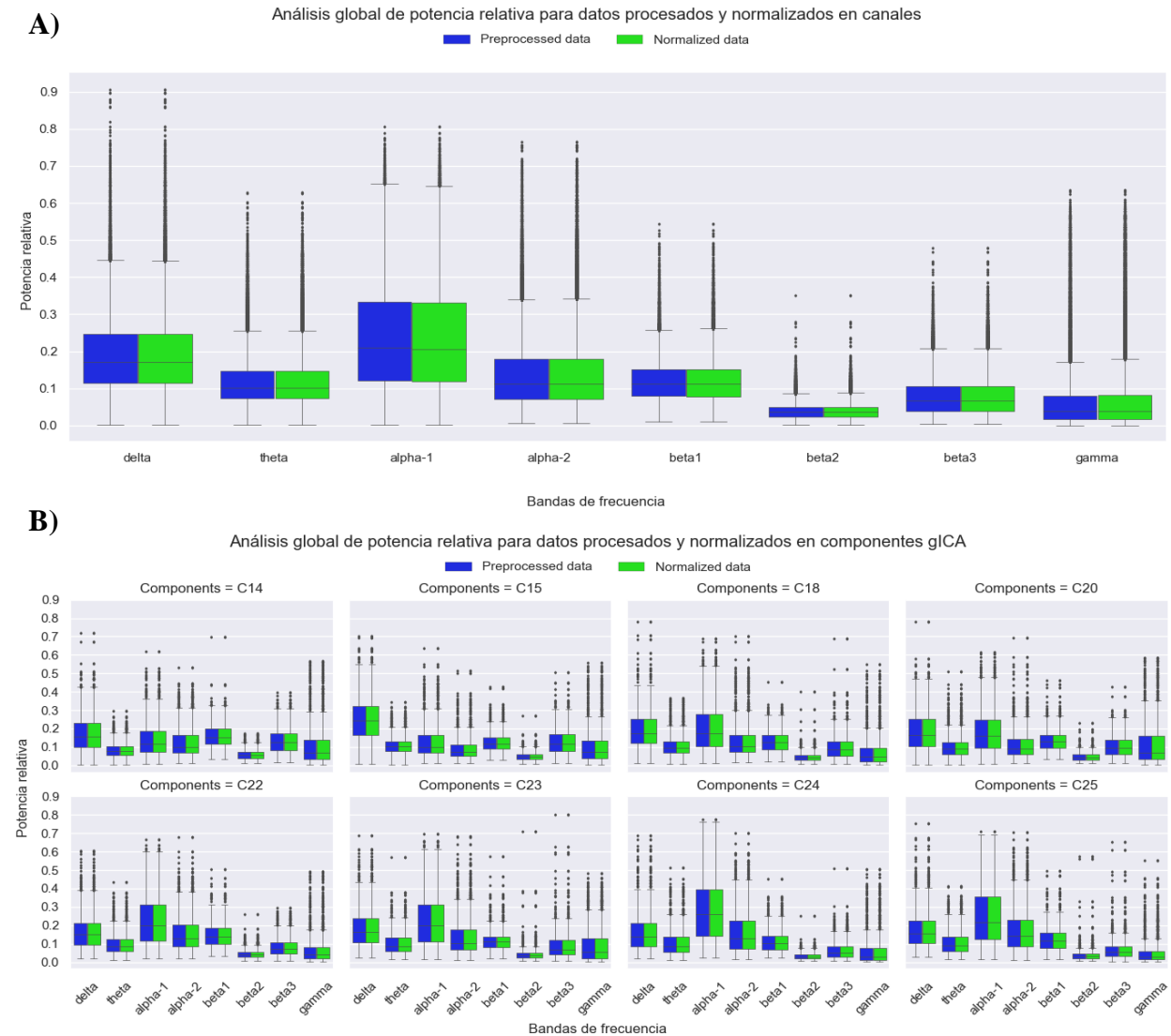


Fig 20. Análisis global en componentes gICA para datos procesados y normalizados. A) Potencia relativa en canales B) Potencia relativa en componentes gICA.

Los resultados obtenidos muestran que la etapa de normalización con respecto a la procesada no existe diferencias, lo cual puede deberse a que el cálculo de potencia relativa tiene implícito un proceso de normalización, por lo que volver a normalizar los datos no va a tener

cambios significativos. Al comparar los datos con y sin normalización se evidencia que los cambios se presentan a partir del décimo decimal aproximadamente, esto en términos matemáticos representan el mismo valor, razón por la que, al realizar la norma sobre los datos, no se encuentran diferencias al graficar las diferentes representaciones de cajas y bigotes.

## *2. Análisis ente estudios*

Para la comparación entre estudios se realizó un gráfico similar al enfoque realizado en el análisis global. Inicialmente se muestra la potencia por canales en la Fig 21 y luego se saca los gráficos en componentes. Sin embargo, para mostrar la potencia en componentes fue necesario diseñar una estrategia para presentar toda la información. Las opciones posibles para ilustrar eran por cada una de las componentes y todas las bandas comparando los 'n' estudios, o hacer un esquema por cada una de las bandas y todas las componentes realizando la distinción de las bases de datos, cualquiera de las elecciones constaba de una matriz 8x8 o en su defecto 8x1 en donde se grafica por banda.

El análisis que se determinó como la mejor forma de representar la información es analizando para una sola banda que tan diferentes son los estudios, ya que comparar entre rangos de frecuencia diferentes puede que no sea el análisis más adecuado en este caso, en donde se busca es realizar una comparación entre las bases de datos. Esta información se presenta entre la Fig 21 y la Fig 22, en donde cada elemento de la matriz representa una banda de frecuencia en una componente gICA.



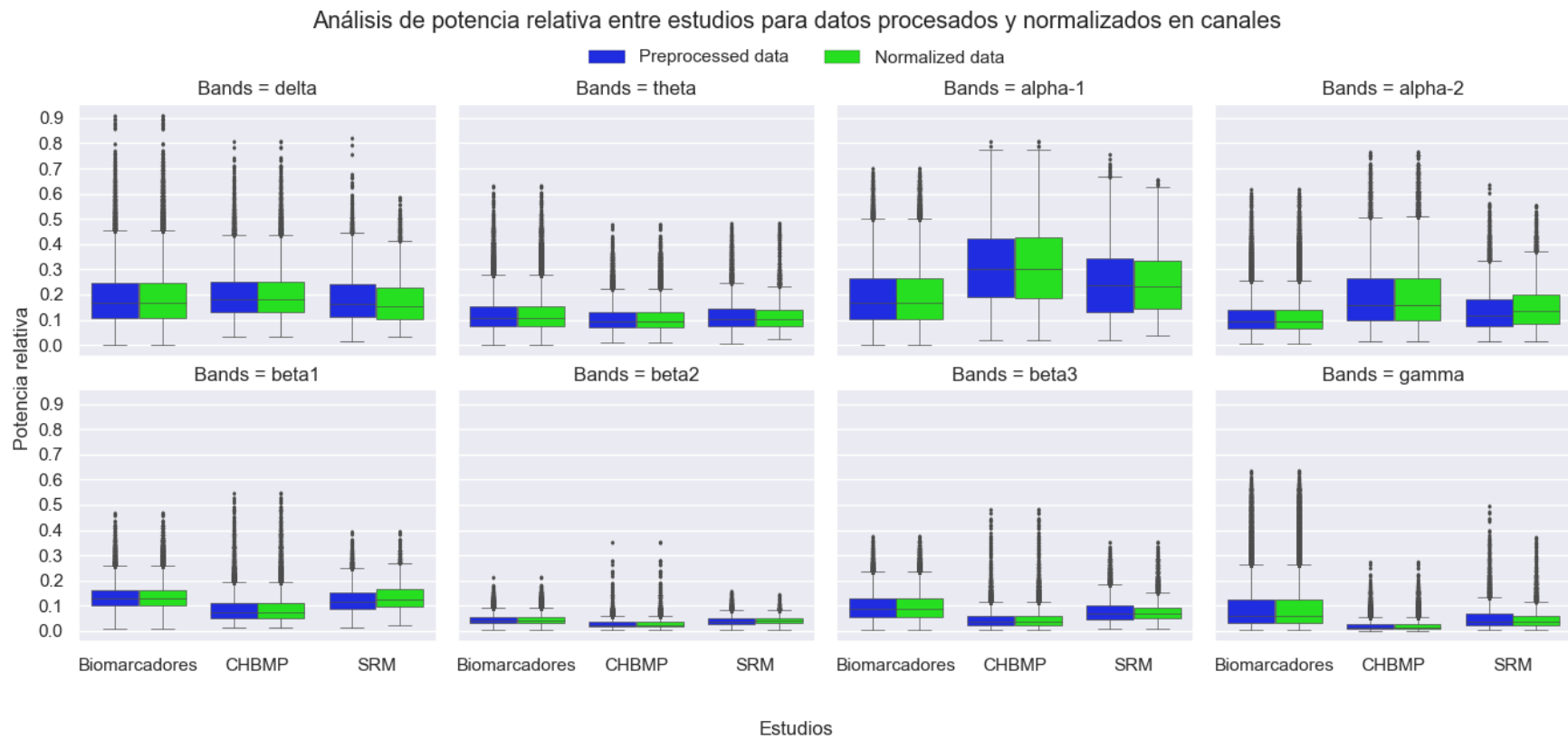


Fig 21. Análisis entre estudios de la potencia relativa de los datos procesados y normalizados en canales. El eje X representa los conjuntos de datos, y el eje Y indica el valor de potencia relativa.

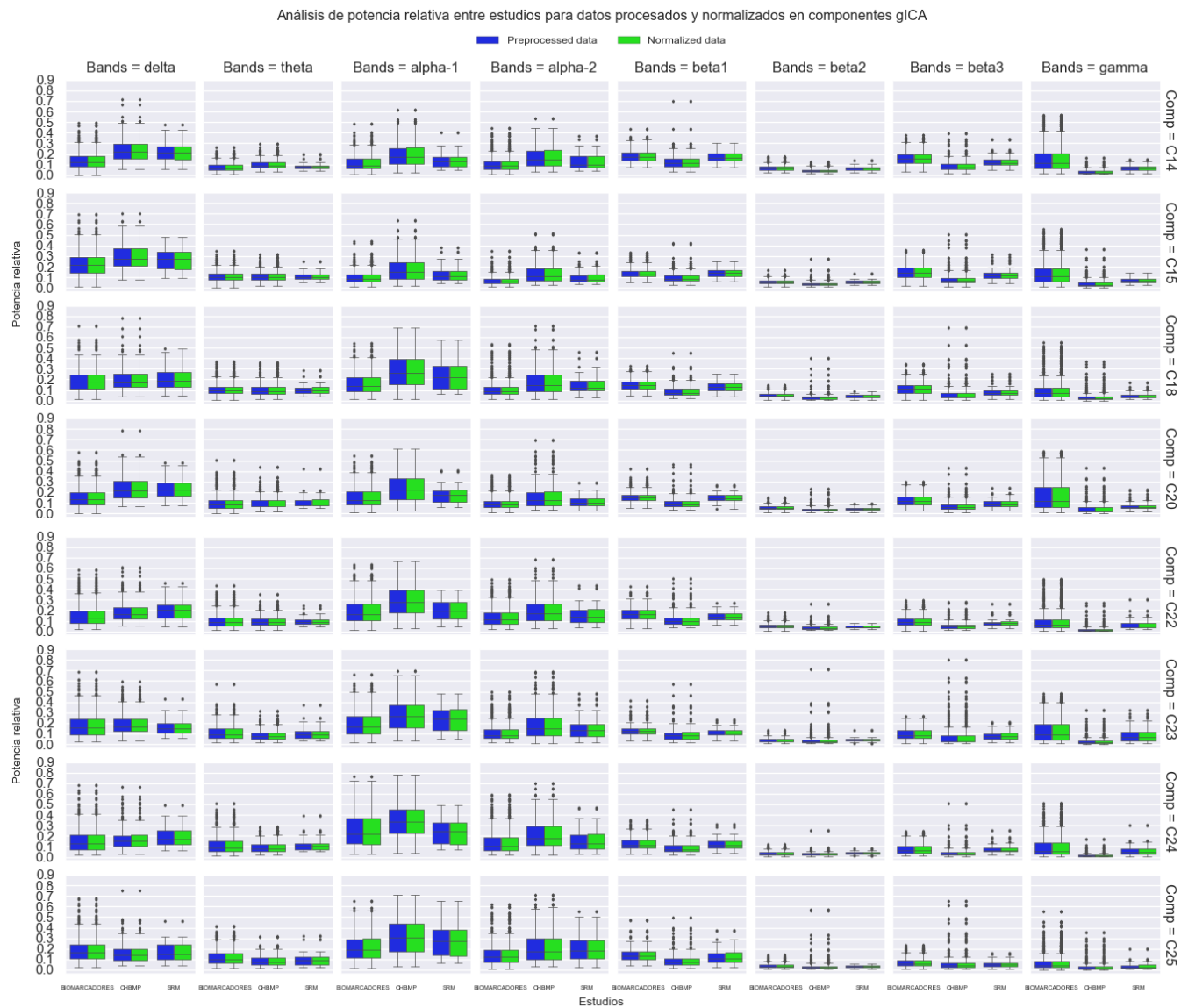


Fig 22. Potencia relativa entre estudios en componentes gICA en datos procesados y normalizados. El eje X indica los diferentes conjuntos de datos y el eje Y indica la potencia relativa. Las columnas hacen referencia a las bandas de frecuencia y las columnas a las componentes neuronales.

A pesar de que son muchas gráficas para sacar conclusiones y análisis de los datos, estas permiten entender cómo cambian los resultados según el ritmo que se esté analizando. En algunos estudios o en algunas poblaciones predominan ritmos específicos; en tal caso, este tipo de gráficas permiten analizar los ritmos de interés y sacar relaciones o patrones que se puedan presentar en una población.

### 3. Análisis entre grupos

Al igual que en la fase de preprocesamiento, aquí solo se realiza un análisis para la base de datos de *Biomarcadores* presentado en la Fig 23 y en la Fig 24. Sin embargo, el gráfico en componentes el eje X muestra son las bandas de frecuencia y los grupos se categorizan por cada ritmo de frecuencia.

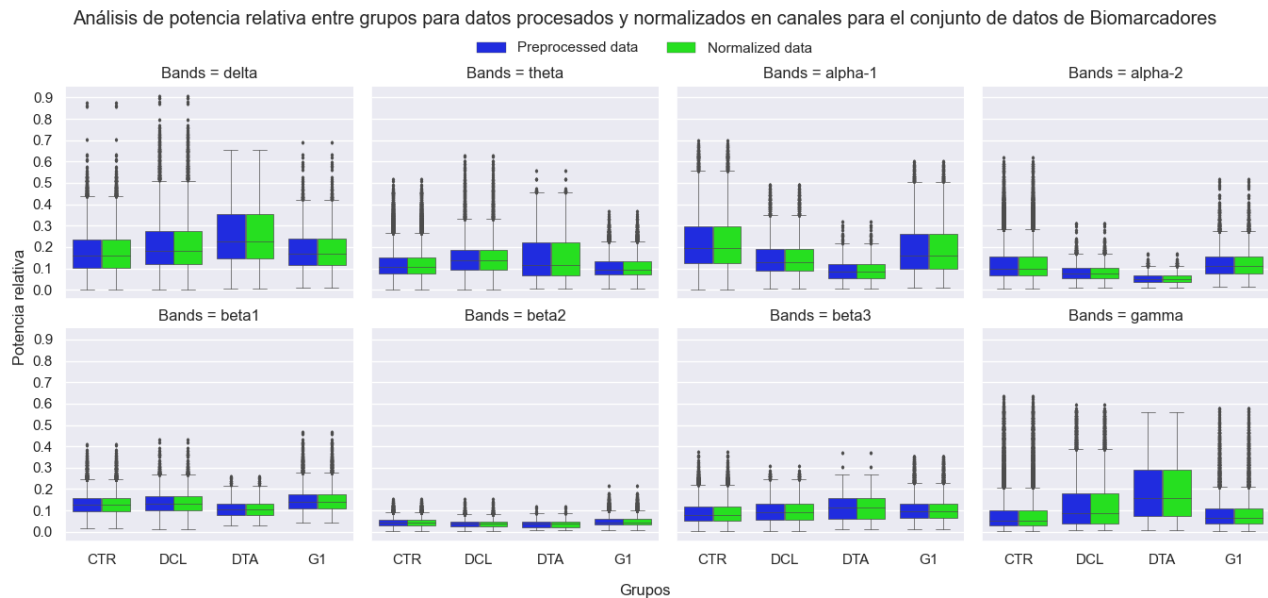


Fig 23. Análisis de la potencia relativa de los datos procesados y normalizados en canales para el conjunto de datos de Biomarcadores. El eje X representa los grupos de estudio, y el eje Y indica el valor de potencia relativa.

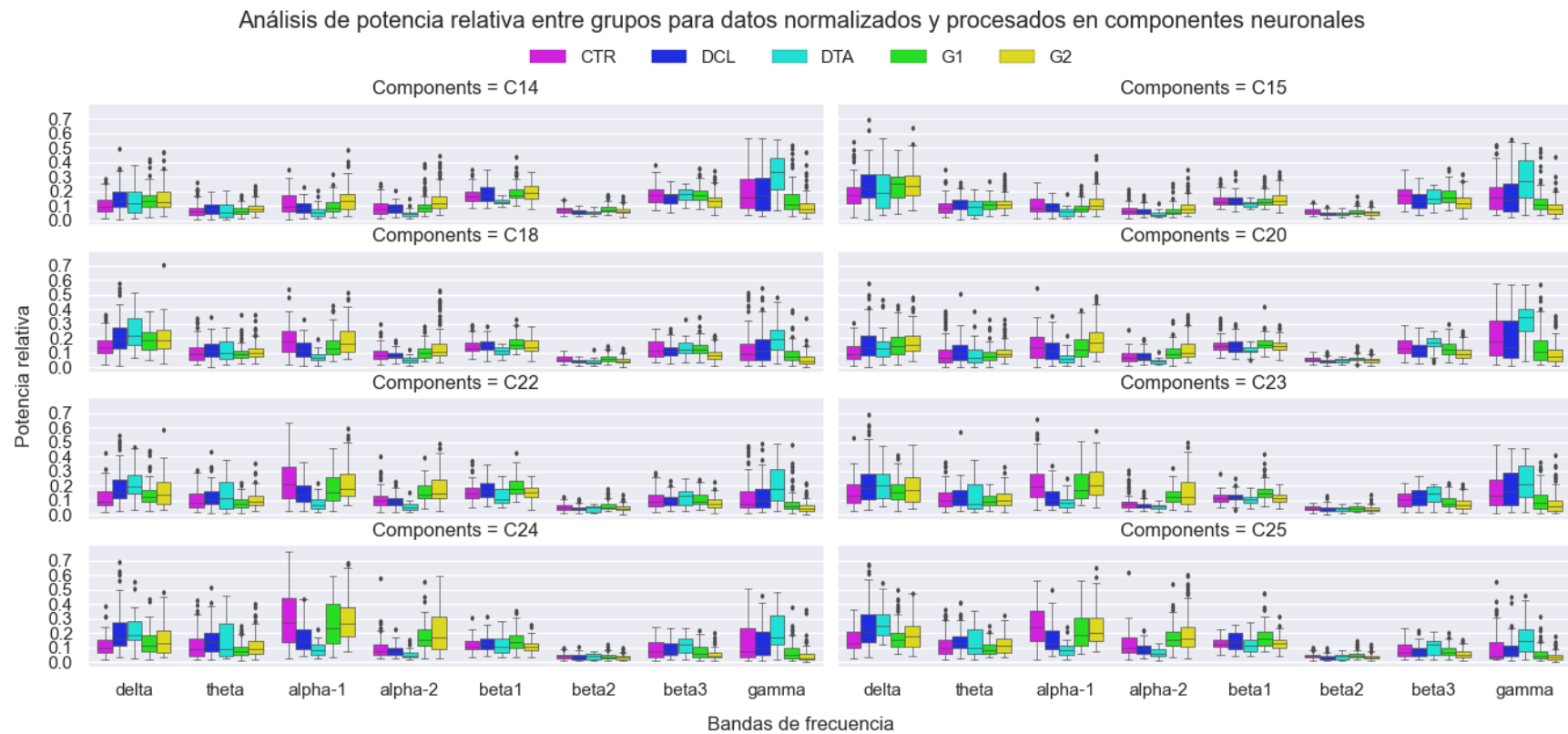


Fig 24. Análisis de la potencia relativa de los datos procesados y normalizados en componentes neuronales para los datos de Biomarcadores. El eje X representa las bandas de frecuencia, y el eje Y indica el valor de potencia relativa.

Los resultados de este enfoque permiten identificar las diferencias que existen entre los grupos de la base de datos de *Biomarcadores*, en donde se observa que el análisis por componentes presenta mayor número de outliers que en el enfoque por componentes. Este es un comportamiento que se presenta a nivel general en la etapa de post-procesamiento. En las diferentes gráficas los rangos de frecuencia para la banda de frecuencia beta en sus diferentes clasificaciones (beta-1, beta-2 y beta-3), es el ritmo con menor energía a nivel general.

#### 4. Análisis entre sesiones

A nivel de sesiones, se analiza la base de datos *SRM* y *Biomarcadores*. La Fig 25 muestra la potencia por canales para la base de datos SRM, en donde cada elemento de la matriz hace referencia a cada una de las bandas de frecuencia, mientras que, las gráficas comprendidas entre la Fig 26- 28 informa el cálculo de potencia relativa en componentes.

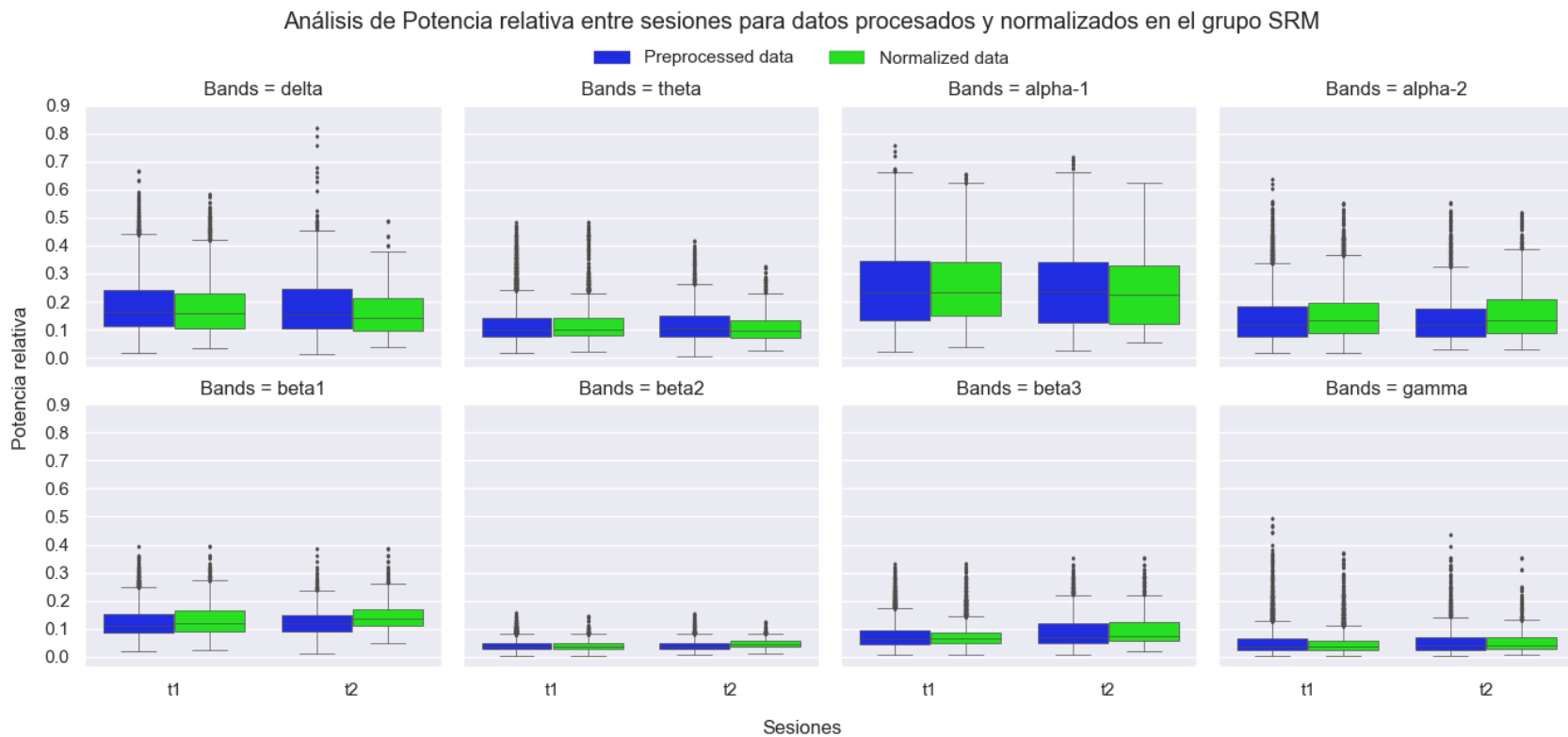


Fig 25. Análisis de la potencia relativa del conjunto de datos SRM procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia

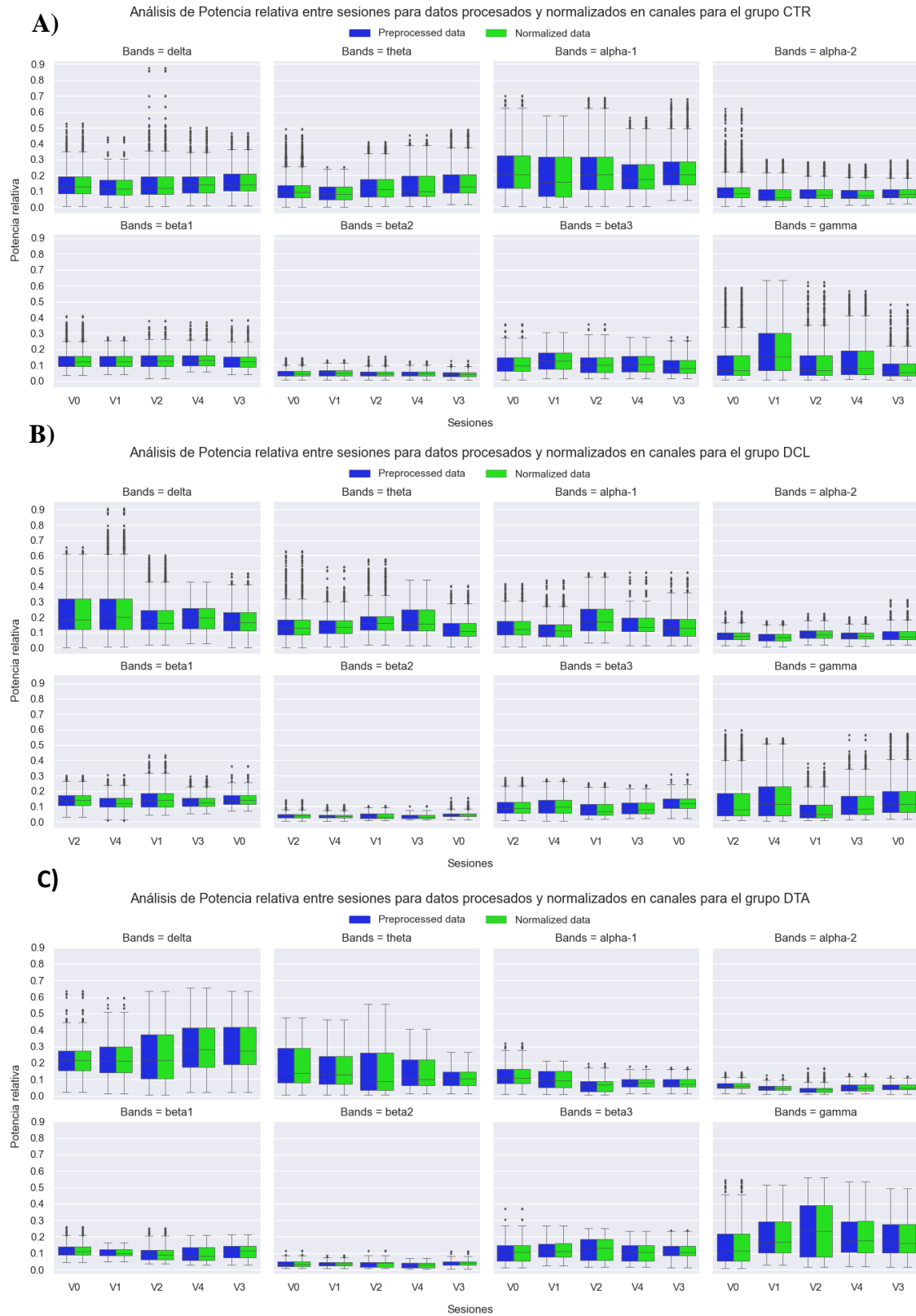


Fig 26. Análisis de la potencia relativa del conjunto de datos BIOMARCADORES procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. A) CTR; B) DCL; C) DTA.

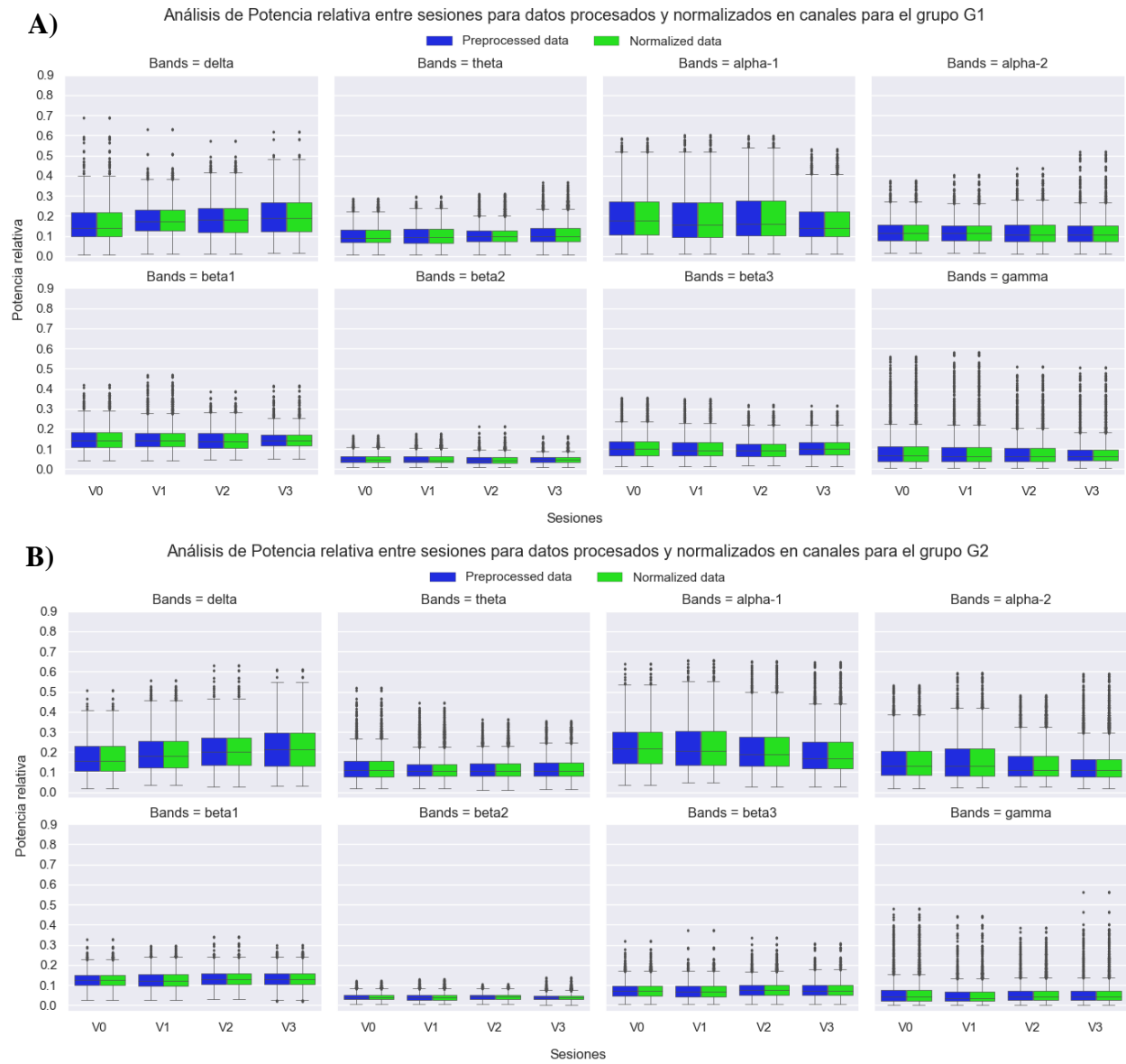


Fig 27. Análisis de la potencia relativa del conjunto de datos BIOMARCADORES procesados y normalizados en canales. El eje X representa las sesiones del estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia. A) G1; B) G2.

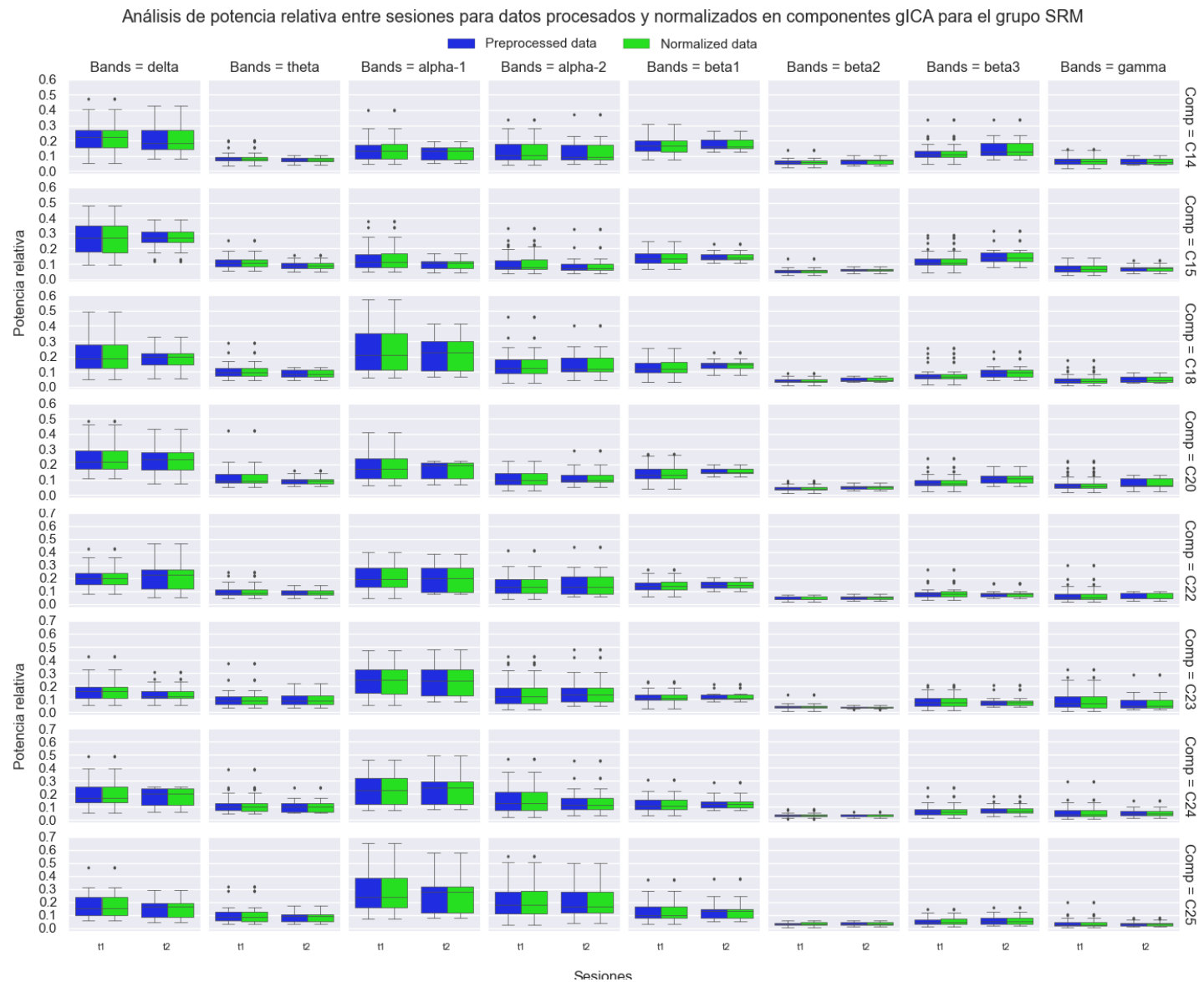


Fig 28. Análisis de la potencia relativa del conjunto de datos SRM procesados y normalizados. El eje X representa las sesiones de estudio, y el eje Y indica el valor de potencia relativa. Cada columna representa las bandas de frecuencia y las filas corresponde a las componentes gICA etiquetadas como neuronales.



De acuerdo con los resultados obtenidos mediante este enfoque se observa que hay un comportamiento similar entre lo discutido anteriormente, ya que se conserva cambios muy mínimos entre la fase procesada y normalizada. Además, desde el punto de vista de las sesiones, estas presentan comportamientos no completamente iguales debido a que en la base de datos SRM no hay uniformidad entre el tiempo de adquisición de la segunda sesión, debido a que el estudio reporta que no fue una etapa planeada en el momento de la organización y programación de los registros, es decir, no necesariamente la misma cantidad de sujetos registrados en la primera sesión son la misma cantidad de participantes en el segundo registro.

Los gráficos para la base de datos de Biomarcadores, no se presenta debido a que son muchas gráficas, ya que son cinco grupos, y por cada grupo son ocho bandas y ocho componentes, es decir cuarenta gráficos en total. Sin embargo, el análisis para este conjunto de datos es similar que para la base de datos SRM analizada previamente.

#### *D. Desarrollo de interfaces*

Para el desarrollo del producto mínimo viable, se desarrollaron cuatro interfaces principales. La pestaña de inicio en donde se encuentra información para cualquier tipo de usuario que ingrese a la plataforma, luego se encuentra la interfaz de carga de datos, que es de acceso único a usuarios del sistema, al igual que las interfaces de procesamiento de datos y control de calidad de datos.

La Fig 29, presenta la pestaña de inicio en donde se encuentra información general de acceso público, en donde el usuario podrá conocer los servicios que ofrece la plataforma que van desde exploración de datos de electroencefalografía hasta procesamiento y análisis de control de calidad. Así mismo, en la parte superior de esta pestaña, encontrará el menú para realizar la navegación por las diferentes interfaces, y en esta misma encontrada un ícono de login, el cual fue introducido para trabajos futuros en donde se realice gestión de usuarios.



Fig 29. Interfaz de inicio. 1) Panel de navegación; 2) Ingreso al sistema.

Las pestañas de ‘*Carga de datos*’, ‘*Procesamiento de datos*’ y ‘*Control de calidad de estudios*’ requieren el ingreso de usuario, por tanto, para cumplir con este requisito, en caso del usuario navegar por estas pestañas sin acceder de forma correcta, encontrará mensajes de alerta que indicarán acceso restringido, en donde se debe realizar inicialmente el proceso de ingreso al sistema para poder acceder a la información. La Fig 30, muestra la ventana emergente diseñada para notificar al usuario el acceso denegado, solicitando el ingreso al sistema con las credenciales correspondientes.

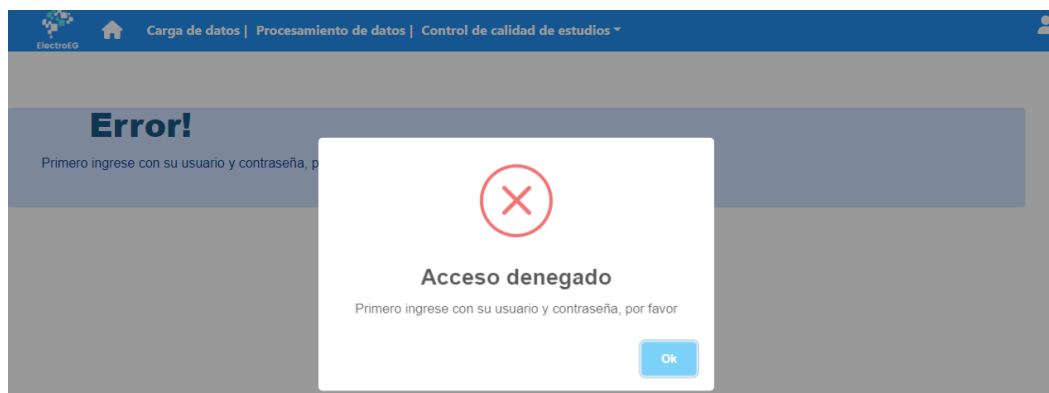


Fig 30. Sistema de alertas para notificar acceso denegado de navegación.

En la Fig 31, se muestra el panel de ingreso del usuario en donde se solicita introducir un correo electrónico y una contraseña para poder entrar al sistema. Luego de realizar este proceso se habilitan las pestañas de acceso denegado.

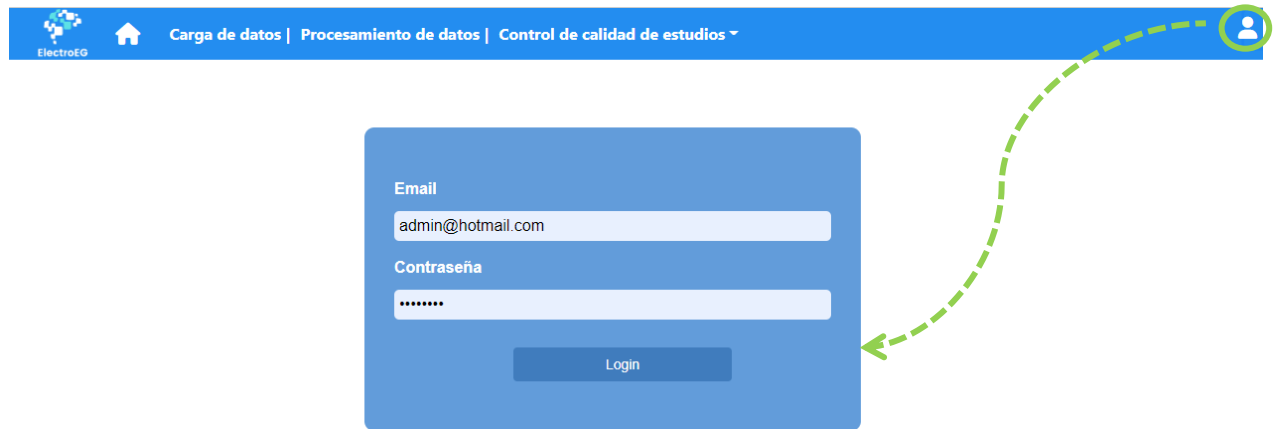


Fig 31. Sistema de ingreso del usuario.

La interfaz de carga de datos está pensada para subir datos únicamente en formato BIDS, en la Fig 32, se muestra el diseño de esta, dividido en tres partes principales: carga de archivos, formulario de la base de datos y visualización del conjunto de datos cargados. Inicialmente se ingresan los datos, en donde únicamente se permite seleccionar carpetas, ya que el formato BIDS se encuentra estructurado en folders. Después de subir los datos, se solicita al usuario que ingrese la información asociada a la base de datos de prueba cargada, en donde se ingresa el '*nombre del estudio*' que hace referencia al nombre de la base de datos según la codificación; se solicita la '*extensión*' en la que se almacenaron los datos al aplicar el estándar BIDS; se ingresa la '*condición*' que se haya realizado acorde a como se haya categorizado la nomenclatura para la estandarización de los datos; Se ingresa la '*frecuencia de línea*' para la remoción de este tipo de ruido que en general es 50 Hz o 60 Hz, sin embargo también acepta una lista de frecuencias en donde podrá listar los armónicos; se ingresa la '*Expresión regular de grupo*' que es común utilizarla cuando los datos son robustos y complejos y se tiene una expresión regular para identificar nomenclaturas definidas dentro de la codificación del estudio, en caso de que no exista para alguna base de datos simplemente se deja vacía, para ser interpretada como un dato tipo None; finalmente se ingresa los campos de los '*Eventos*' que hace referencia al código numérico establecido por la base de datos para nombrar una condición en específico.

**CARGA DE DATOS CRUDOS DE EEG**

1) Seleccionar archivo Ninguno archivo seleccionado **Cargar**

Campos obligatorios\*

2) **Nombre estudio\***  
Nombre del estudio según el estándar BIDS

**Extensión\***  
Seleccione una extensión

**Condición\***  
Condición

**Frecuencia de línea\***  
Lista de frecuencias, en su defecto la frecuencia de la red

**Expresión regular de grupo\***  
Expresión regular de grupo

**Eventos\***  
Eventos

3) **Guardar**

Nombre estudio	Número de sujetos	Número de EEGs	Número de EEGs procesados	Operaciones
Biomarcadores	1	1378	447	<b>Eliminar</b>

Fig 32. Interfaz carga de datos. 1) Cargar datos en formato BIDS; 2) Ingresar campos obligatorios acorde a la nomenclatura del estándar BIDS; 3) Visualización de las bases de datos cargadas.

Una vez ingresado los campos obligatorios se guarda la información, y será visualizado la base de datos en la parte inferior de la interfaz en donde se especifica el nombre del estudio, cantidad de participantes, número de registros crudos de EEG y cantidad de registros procesados. En esta tabla encontrará la opción de operación ‘eliminar’ para evitar almacenar información que ya no sea de interés para el usuario y así poder ir depurando el sistema.

En la siguiente tabla encontrará sintetizado cada uno de los componentes mencionados que hacen parte de la interfaz de ‘carga de datos’ en donde se especifica la funcionalidad y propósito de diseño.

Tabla X. Componentes de la interfaz ‘Carga de datos’ con su respectiva funcionalidad y propósito de diseño.

CARGA DE DATOS		
Componentes	Funcionalidad	Propósito
Seleccionar archivos	Envía la petición para seleccionar los archivos.	Este botón fue diseñado para acceder a los archivos del sistema para poder cargarlos.
Botón cargar	Envía la petición de almacenamiento al Backend.	Enviar la petición al Backend con los datos de la base de datos.
Input nombre estudio	Input Field, en donde se solicita información del nombre de la base de datos cargada.	Insertar el atributo ‘Nombre estudio’ el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Input extensión	Input Field, en donde se solicita seleccionar la extensión de los registros de la base de datos cargada.	Insertar el atributo ‘Extensión’ el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Input condición	Input Field, en donde se solicita la condición de la tarea realizada en la adquisición de los registros.	Insertar el atributo ‘Condición’ el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Input frecuencia de línea	Input Field, en donde se solicita la frecuencia de línea y sus armónicos si es el caso.	Insertar el atributo ‘Frecuencia de línea’ el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Input expresión regular de grupo	Input Field, en donde se solicita la expresión regular de grupo.	Insertar el atributo ‘Expresión regular de grupo’ el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Input eventos	Input Field, en donde se solicita la identificación numérica del evento, de acuerdo con la nomenclatura establecida por el estudio.	Insertar el atributo “eventos” el cual se añade en el archivo ‘ <i>dataset_description</i> ’ en formato json.
Botón guardar	Envía la petición de almacenamiento en formato json, simulando una base de datos al Backend.	Enviar la petición al Backend con la información de la base de datos, ingresada por el usuario en los campos de ingreso.
Botón eliminar	Envía la petición de eliminar al Backend, en donde se borra del espacio de almacenamiento y del archivo ‘ <i>dataset_description.json</i> ’	Enviar la petición al Backend con la información de la base de datos para realizar la operación de eliminación.

La interfaz de ‘*Procesamiento de datos*’ se presenta en la Fig 33, en donde se muestra una tabla de chequeo que contiene el nombre del estudio, número de sujetos, número de EEGs sin procesar y procesados. Además, se encuentra la columna de Estado y operaciones las cuales permiten identificar si la base de datos está completamente procesada y la opción de eliminar la base de datos respectivamente.

La base de datos muestra el estado ‘Procesado’ cuando el número de EEGs coincide con el número de EEGs procesados, sin embargo, puede que el usuario solo procese una condición como es el caso de la base de datos de ‘Biomarcadores’ el cual únicamente cuenta con los archivos

procesados en condición ojos cerrados. Si la condición de ojos abiertos se procesara cambiaría el tipo de estado.

The screenshot shows the 'BASES DE DATOS' section of the application. At the top, there is a navigation bar with 'ElectroEG', a home icon, and menu items: 'Carga de datos', 'Procesamiento de datos', and 'Control de calidad de estudios'. Below this is a blue header with a database icon and the text 'BASES DE DATOS'. A green instruction 'Seleccione la base de datos a procesar' is displayed above a table. The table has columns: 'Nombre estudio', 'Número de sujetos', 'Número de EEGs', 'Número de EEGs procesados', 'Estado', and 'Operaciones'. The first row, 'Biomarcadores', is selected with a radio button. Below the table is a blue button labeled 'Procesar archivos'.

Nombre estudio	Número de sujetos	Número de EEGs	Número de EEGs procesados	Estado	Operaciones
<input checked="" type="checkbox"/> Biomarcadores	1	1378	447	Sin procesar	
<input type="checkbox"/> Dataset containing Cuban Brain Mapping database	282	250	247	Sin procesar	
<input type="checkbox"/> SRM Resting-state EEG	111	153	153	Procesado	

Fig 33. Interfaz procesamiento de datos.1) Visualización de las bases de datos previamente cargadas.

Una vez seleccionado las bases de datos a procesar y seleccionado el botón procesar archivos, se le notifica al usuario mediante una ventana emergente que la solicitud ha iniciado y su finalización será comunicado mediante un correo electrónico para poder continuar con la exploración y análisis de los datos.

The screenshot shows the same interface as Fig 33, but with a white notification dialog box in the center. The dialog box has a green checkmark icon and the title 'Procesando datos'. The text inside reads: 'Una vez terminado el proceso se notificará por medio de un correo electrónico'. There is an 'Ok' button at the bottom right of the dialog. The background interface is dimmed.

Fig 34. Notificación de la solicitud de procesamiento de datos.

Por otra parte, se presenta la Tabla XI, el cual muestra un resumen de los componentes en el proceso de diseño de la interfaz, en donde encontrará información acerca de su funcionalidad y propósito de creación.

Tabla XI. Componentes de la interfaz ‘Procesamiento de datos’ con su respectiva funcionalidad y propósito de diseño.

PROCESAMIENTO DE DATOS		
Componentes	Funcionalidad	Propósito
Casilla de chequeo	Casilla de verificación para hacer múltiples selecciones para procesar los datos.	Este componente tiene el objetivo de seleccionar una o múltiples bases de datos para iniciar el procesamiento una vez se presione el botón “Procesar archivos”.
Botón procesar archivos	Envía la petición de procesamiento de datos al Backend en donde se notifica que va a inicializar el proceso y será notificado mediante un correo electrónico su finalización.	Enviar la petición al Backend con la información de las bases de datos a procesar.
Botón eliminar	Envía la petición de eliminar al Backend, en donde se borra del espacio de almacenamiento y del archivo ‘ <i>dataset_description.json</i> ’.	Enviar la petición al Backend con la información de la base de datos a eliminar.

En la Fig 35, se muestra el mensaje de notificación que deberá llegar una vez el proceso es completado. En el punto actual de diseño, se envía a un correo electrónico fijo, sin embargo, lo ideal es que en trabajos futuros cuando se realice la gestión de usuarios, con el correo electrónico que se haya registrado en el sistema, a ese deberá notificar el proceso de completado.



Fig 35. Notificación de procesamiento completado al correo electrónico.

La interfaz de ‘Control de calidad de estudios’ se muestra en la Fig 36. Esta gráfica muestra cada uno de los botones con los cuales se pueden hacer los análisis previamente mostrados en la sección V literal C. Al presionar cada uno de estos botones encontrará una próxima interfaz la cual preguntará que tipo de etapa desea analizar si la fase de preprocesamiento o la de post-procesamiento, al dar clic en cualquiera de estas opciones encontrará los mismos gráficos mostrados anteriormente ya que en este caso se tienen las mismas bases de datos cargadas en la aplicación. Sin embargo, si cambiara el número de participantes o cualquier otro parámetro que influya sobre el análisis de los datos las gráficas se modifican.

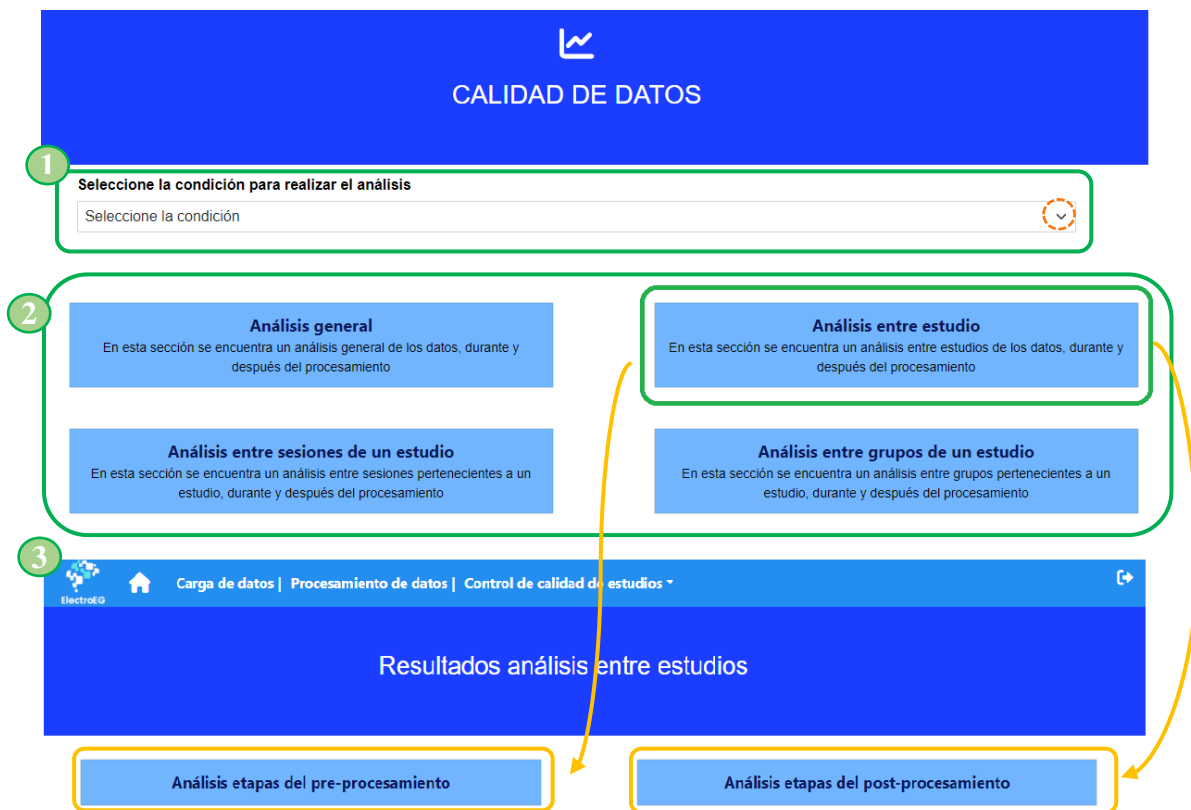


Fig 36. Interfaz de control de calidad de datos. 1) Panel de selección de condición para analizar; 2) Panel de selección de análisis de control de calidad; 3) interfaz derivada de los botones de análisis según corresponda.

En esta interfaz es de vital importancia el tiempo de ejecución para la generación de las gráficas, por tal razón en la siguiente tabla se muestra la duración de tiempo estimado que demora en generar los gráficos para los diferentes análisis, en cada una de las fases que incluyen las múltiples etapas descritas anteriormente.



Tabla XII. Duración de tiempo para la generación de los gráficos de control de calidad de datos.

<b>Análisis</b>	<b>Tiempo pre-procesamiento [s]</b>	<b>Tiempo-post-procesamiento [s]</b>
Global	15.532485500000007	
Estudios	19.746037399999977	41.670448099999999
Sesiones de estudio	22.114257500000008	521.5640822
Grupos de estudio	22.107645300000006	38.632331699999895

Los resultados presentados en la Tabla XII, son tiempos relativos que dependen de varios factores como lo es calidad de internet y características del servidor en la que se estén ejecutando los servicios.

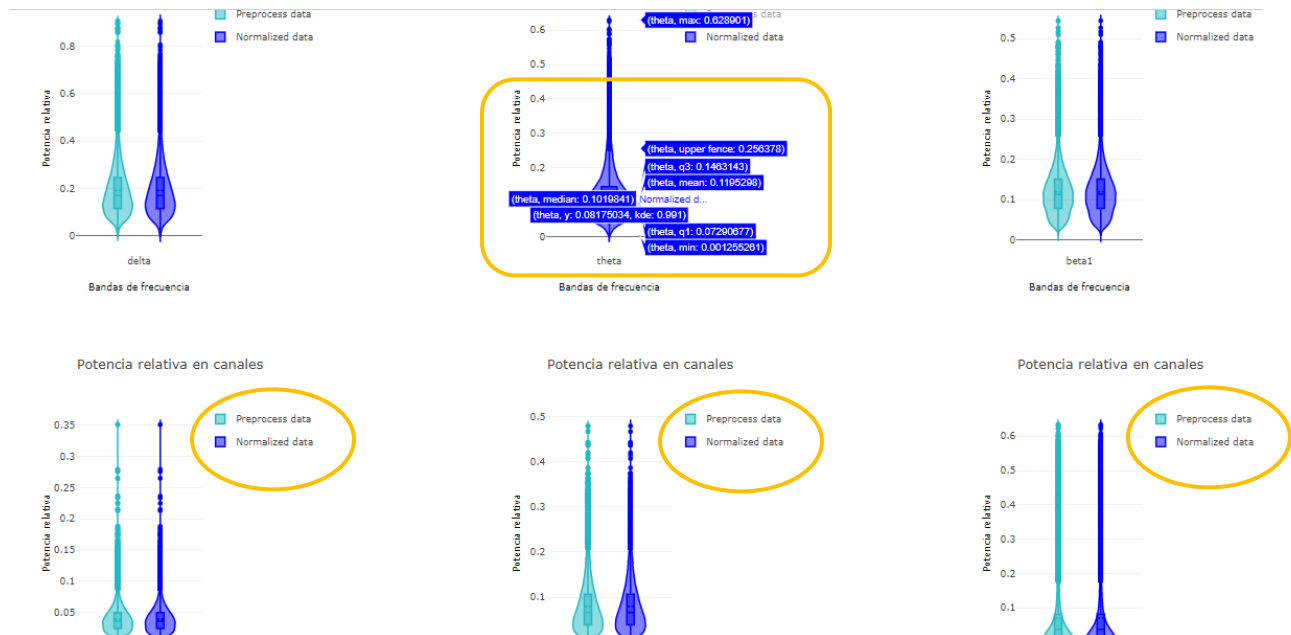
Finalmente, La Tabla XIII presenta la información de los componentes que hacen parte la interfaz de control de calidad de datos y sus derivaciones, en donde se describe brevemente cada uno de los botones que finalidad tiene en el desarrollo de la aplicación.

Tabla XIII. Componentes de la interfaz ‘Control de calidad de datos’ con su respectiva funcionalidad y propósito de diseño.

<b>CONTROL DE CALIDAD DE DATOS</b>		
<b>Componentes</b>	<b>Funcionalidad</b>	<b>Propósito</b>
Botón de selección	Envía la petición para seleccionar la condición a analizar.	Enviar la petición al Backend para realizar únicamente los gráficos asociados a la condición seleccionada.
Botón análisis general	Envía la petición para generar los gráficos con el proceso de mezcla del conjunto de bases de datos cargada.	Visualizar la información de las bases de datos mezcladas para evaluar la calidad de las diferentes etapas, tanto de preprocesamiento como de post-
Botón análisis entre estudio	Envía la petición para generar los gráficos de las etapas: PREP, wICA y rechazo de épocas discriminado entre bases de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del preprocesamiento.
Botón análisis post-procesamiento	Envía la petición para generar los gráficos de potencia relativa canales y componentes gICA, discriminado entre bases de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del post-procesamiento.
Botón análisis entre sesiones de un estudio	Envía la petición para generar los gráficos de las etapas: PREP, wICA y rechazo de épocas discriminado entre sesiones de un conjunto de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del preprocesamiento.
Botón análisis post-procesamiento	Envía la petición para generar los gráficos de potencia relativa canales y componentes gICA, discriminado entre sesiones de un conjunto de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del post-procesamiento.
Botón análisis entre grupos de un estudio	Envía la petición para generar los gráficos de las etapas: PREP, wICA y rechazo de épocas discriminado entre grupos de un conjunto de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del preprocesamiento.
Botón análisis post-procesamiento	Envía la petición para generar los gráficos de potencia relativa canales y componentes gICA, discriminado entre grupos de un conjunto de datos.	Visualizar las métricas de control de calidad de las diferentes etapas del post-procesamiento.

Los gráficos en la sección de análisis global en el diseño de la plataforma se realizaron con funciones de JavaScript para observar si cambiaban los tiempos de generación de los gráficos. Sin embargo, no hay una diferencia notoria entre estos, motivo por el cual no se realizó la modificación

de todo el módulo. No obstante, estos gráficos presentan una ventaja sobre los gráficos codificados de Python, ya que estos gráficos son dinámicos y tienen diferentes herramientas de visualización como acercar, alejar y guardar. En la Fig 37 se muestra los gráficos generados a partir de JS para la etapa de post-procesamiento, en donde se ilustra su dinamismo.



Por otro lado, en la Fig 38 se presenta el panel de operaciones que ofrece Plotly, un framework que permite generar graficas de JavaScript, permitiendo al usuario guardar de manera local cada gráfico, además de, realizar acercamiento o alejamiento sobre la gráfica.

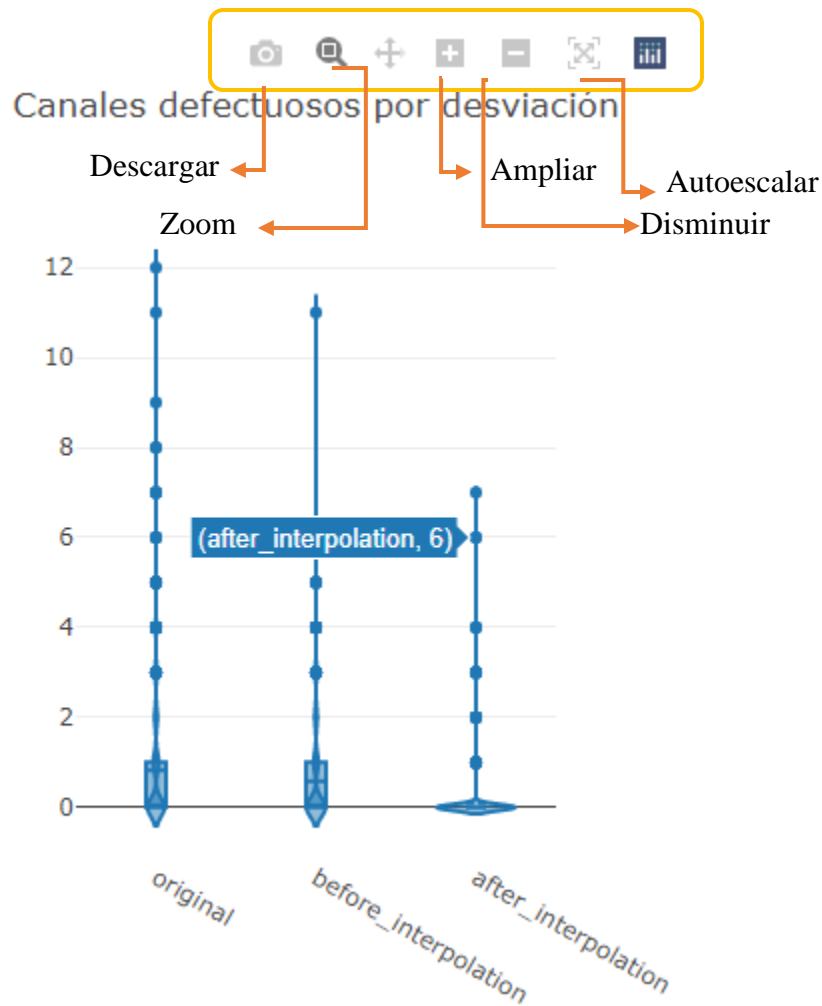


Fig 38. Panel de operaciones de las gráficas generadas con JS.

Las etapas de control de calidad contienen textos de ayuda, que permiten interpretar cada una de las métricas graficadas. En las siguientes figuras se muestran la información que se le presenta al usuario para la etapa del PREP, wICA, rechazo de épocas ruidosas y potencia relativa, respectivamente. Inicialmente se muestra una descripción de la función que realiza cada etapa sobre los datos, y luego se presenta una tabla, en donde cada fila proporciona información acerca de cada métrica evaluada.

ElectroEG
Carga de datos | Procesamiento de datos | Control de calidad de estudios

RESULTADOS ANÁLISIS GLOBAL

ETAPA PREP

PREP realiza la eliminación de ruido automatizada, la detección de canales defectuosos y la referenciación de una manera que permite a los usuarios trabajar con los datos procesados. El PREP se basa en cuatro pasos principales:

- Eliminación del ruido de línea sin utilizar estrategias de filtrado.
- Referencia sólida de la señal en relación con una estimación de la referencia promedio "verdadera".
- Detección e interpolación de canales defectuosos en relación con la referencia.
- Conservar suficiente información para permitir a los usuarios volver a hacer referencia utilizando otro método o deshacer la interpolación de un canal en particular.

Métrica	Descripción
Canales defectuosos tipo NAN	Detección de canales que contengan datos tipo NAN.
Canales defectuosos tipo planos	Localización de canales cuando aparece plana la señal en relación con los otros canales cuando se presenta un gráfico de desplazamiento.
Canales defectuosos por desviación	Se estima las diferencias de amplitud entre los conjuntos de datos para clasificar como un canal malo por desviación. Sin embargo, no identifica los canales que capturan los parpadeos y actividad muscular como ruidosos.

Fig 39. Texto descriptivo para la interpretación de las métricas de control de calidad del PREP.

ETAPA wICA

Wavelet mejorado con la técnica ICA (wICA). El método se basa en ICA y hace uso del umbral de wavelet no para eliminar el ruido del EEG sin procesar observado, sino como un paso intermedio para los componentes independientes desmezclados. La umbralización permite conservar la estructura tiempo-frecuencia de los artefactos y recuperar la actividad cerebral "filtrada" en los componentes

Métrica	Descripción
Relación de componentes filtradas sobre componentes totales	Detección y estimación del porcentaje filtrado de componente en relación con el número de componentes totales. Esta métrica informa que tanto se están filtrando los datos en esta etapa.

Transformada wavelet junto con la técnica de análisis de componentes independientes (wICA)

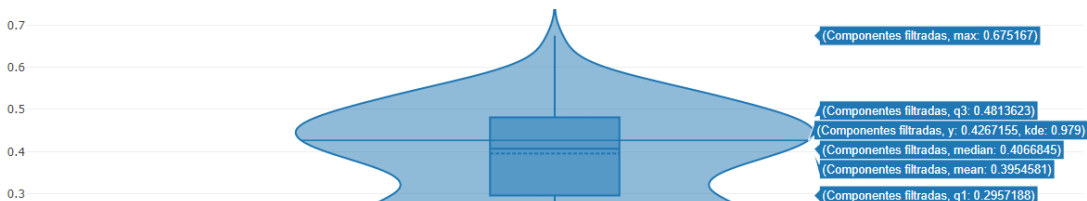


Fig 40. Texto descriptivo para la interpretación de las métricas de control de calidad del wICA.

ETAPA RECHAZO DE ÉPOCAS

Métrica	Descripción
Curtosis	Medida estadística que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.
Amplitud	Medida física que muestra la variación máxima del desplazamiento.
Tendencias lineales	Tendencia lineal es una línea recta que se usa con conjuntos de datos lineales. Muestran patrones en donde una línea de tendencia lineal frecuentemente muestra que hay algo que aumenta o disminuye a un ritmo constante.
Potencia espectral	Medición que informa la distribución de potencia espectral sobre los ritmos de frecuencia de una señal.

Fig 41. Texto descriptivo para la interpretación de las métricas de control de calidad de rechazo de épocas ruidosas.

**POTENCIA RELATIVA**

La Densidad Espectral de una señal es una función matemática que informa la distribución de la potencia o la energía sobre las distintas frecuencias de las que está formada.

Banda de frecuencia	Rango de frecuencia (Hz)	Característica
Delta	1,5-6	Sueño profundo
Theta	6-8.5	Meditación profunda
Alpha	8.5-12.5	Ojos cerrados, despierto
Beta	12.5-30	Ojos abiertos, pensando
Gamma	30-45 Hz	Estado de consciencia

Fig 42. Texto descriptivo para la interpretación de las métricas de control de calidad del post-procesamiento.

### E. Gestión de servicios Linux

Para el despliegue de la aplicación se utilizó un gestor de servicios conocido como `systemctl` el cual permite la administración central de servicios para controlar el sistema `init` de la máquina. Se inició tres servicios presentados en la Fig 43, 44, y 45, los cuales permiten el despliegue de la plataforma. Para la ejecución de estos servicios se configuró un archivo `<filename>.service`, el cual contiene información del directorio a ejecutar. La configuración de este servicio se encuentra detallado y explicado paso a paso en la guía de servicios de Linux proporcionada en la sección de los anexos.

La Fig 43 y 44, son servicios de Python de `uvicorn`, en donde se muestra el correcto funcionamiento del servicio de procesamiento de datos EEG y el de graficación. Estos se encuentran corriendo en el puerto 8080 y el 8000 respectivamente, su estado puede verificarse mediante el comando `status` cada vez que se inicie el servidor.

```
bioinstrumentation@bioinstrumentation:~/etc/systemd/system$ systemctl status backend_preprocess.service
● backend_preprocess.service - @LuisaMariaZapatasalDarraga: Servicio para procesamiento de señales EEG
   Loaded: loaded (/etc/systemd/system/backend_preprocess.service; static; vendor preset: enabled)
   Active: active (running) since Wed 2022-06-29 19:25:57 -05; 26s ago
     Main PID: 27878 (bash)
       Tasks: 18
      Memory: 443.0M
         CPU: 22.265s
    CGroup: /system.slice/backend_preprocess.service
            └─27878 /bin/bash -c cd /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_preprocess_eeg && source prep_env/bin/activate && python3 src/main.py
               └─27880 python3 src/main.py
                 └─27981 /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_preprocess_eeg/prep_env/bin/python3 -c from multiprocessing.resource_tracker import main
                   └─27982 /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_preprocess_eeg/prep_env/bin/python3 -c from multiprocessing.spawn import spawn_main; spa
Jun 29 19:26:02 bioinstrumentation bash[27878]: INFO: Uvicorn running on http://localhost:8080 (Press CTRL+C to quit)
Jun 29 19:26:02 bioinstrumentation bash[27878]: 2022-06-29 19:26:02,940 - uvicorn.error - INFO - Uvicorn running on http://localhost:8080 (Press CTRL+C to quit)
Jun 29 19:26:02 bioinstrumentation bash[27878]: INFO: Started reloader process [27880] using statreload
Jun 29 19:26:02 bioinstrumentation bash[27878]: 2022-06-29 19:26:02,940 - uvicorn.error - INFO - Started reloader process [27880] using statreload
Jun 29 19:26:08 bioinstrumentation bash[27878]: INFO: Started server process [27902]
Jun 29 19:26:08 bioinstrumentation bash[27878]: 2022-06-29 19:26:08,265 - uvicorn.error - INFO - Started server process [27902]
Jun 29 19:26:08 bioinstrumentation bash[27878]: INFO: Waiting for application startup.
Jun 29 19:26:08 bioinstrumentation bash[27878]: 2022-06-29 19:26:08,266 - uvicorn.error - INFO - Waiting for application startup.
Jun 29 19:26:08 bioinstrumentation bash[27878]: INFO: Application startup complete.
Jun 29 19:26:08 bioinstrumentation bash[27878]: 2022-06-29 19:26:08,266 - uvicorn.error - INFO - Application startup complete.
```

Fig 43. Servicio `systemctl` para el procesamiento de los datos.

```

bioinstrumentation@bioinstrumentation:~/etc/systemd/system$ systemctl status backend-graphics.service
● backend-graphics.service - @LuisaMaríaZapataSaldarriaga - Servicio para graficación y análisis de calidad de datos
   Loaded: loaded (/etc/systemd/system/backend-graphics.service; static; vendor preset: enabled)
   Active: active (running) since Wed 2022-06-29 19:29:24 -05; 39s ago
     Main PID: 27993 (bash)
       Tasks: 4
      Memory: 341.1M
         CPU: 44.814s
      CGroup: /system.slice/backend-graphics.service
              └─27993 /bin/bash -c cd /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_graphics && source my_env/bin/activate && python3 src/main.py
                  └─27995 python3 src/main.py
                    └─28017 /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_graphics/my_env/bin/python3 -c from multiprocessing.resource_tracker import main;main(4)
                      └─28018 /home/bioinstrumentation/Documents/TDG_LUISA/Backend_service_graphics/my_env/bin/python3 -c from multiprocessing.spawn import spawn_main; spawn_main(
Jun 29 19:29:29 bioinstrumentation bash[27993]: INFO: Uvicorn running on http://localhost:8000 (Press CTRL+C to quit)
Jun 29 19:29:29 bioinstrumentation bash[27993]: 2022-06-29 19:29:29,254 - uvicorn.error - INFO - Uvicorn running on http://localhost:8000 (Press CTRL+C to quit)
Jun 29 19:29:29 bioinstrumentation bash[27993]: INFO: Started reloader process [27995] using statreload
Jun 29 19:29:29 bioinstrumentation bash[27993]: 2022-06-29 19:29:29,254 - uvicorn.error - INFO - Started reloader process [27995] using statreload
Jun 29 19:29:34 bioinstrumentation bash[27993]: INFO: Started server process [28018]
Jun 29 19:29:34 bioinstrumentation bash[27993]: 2022-06-29 19:29:34,612 - uvicorn.error - INFO - Started server process [28018]
Jun 29 19:29:34 bioinstrumentation bash[27993]: INFO: Waiting for application startup.
Jun 29 19:29:34 bioinstrumentation bash[27993]: 2022-06-29 19:29:34,612 - uvicorn.error - INFO - Waiting for application startup.
Jun 29 19:29:34 bioinstrumentation bash[27993]: INFO: Application startup complete.
Jun 29 19:29:34 bioinstrumentation bash[27993]: 2022-06-29 19:29:34,612 - uvicorn.error - INFO - Application startup complete.

```

Fig 44. Servicio systemctl para la visualización y gestión de los datos.

Por otro lado, el servicio de node se muestra en la Fig 45, en donde al igual que los servicios de Python puede verificar si el servicio se encuentra activo y que se encuentre corriendo correctamente. En los mensajes proporcionados se muestra el enlace http para acceder por medio de un navegador a la dirección, este servicio está siendo ejecutado por el puerto 3000.

```

bioinstrumentation@bioinstrumentation:~/etc/systemd/system$ systemctl status frontend.service
● frontend.service - @LuisaMaríaZapataSaldarriaga - Servicio para despliegue del front-end de la aplicación
   Loaded: loaded (/etc/systemd/system/frontend.service; static; vendor preset: enabled)
   Active: active (running) since Wed 2022-06-29 19:34:15 -05; 25s ago
     Main PID: 28075 (bash)
       Tasks: 6
      Memory: 200.7K
         CPU: 0.000s
      CGroup: /system.slice/frontend.service
              └─28075 /bin/bash -c cd /home/bioinstrumentation/Documents/TDG_LUISA/Frontend && npm start
                  └─28076 npm start
                    └─28088 sh -c react-scripts start
                      └─28089 node /home/bioinstrumentation/Documents/TDG_LUISA/Frontend/node_modules/.bin/react-scripts start
                        └─28096 /usr/local/bin/node /home/bioinstrumentation/Documents/TDG_LUISA/Frontend/node_modules/react-scripts/scripts/start.js
Jun 29 19:34:19 bioinstrumentation bash[28075]: (Use `node --trace-deprecation ...` to show where the warning was created)
Jun 29 19:34:19 bioinstrumentation bash[28075]: (node:28096) [DEP_WEBPACK_DEV_SERVER_ON_BEFORE_SETUP_MIDDLEWARE] DeprecationWarning: 'onBeforeSetupMiddleware' option is
Jun 29 19:34:28 bioinstrumentation bash[28075]: Starting the development server...
Jun 29 19:34:23 bioinstrumentation bash[28075]: Compiled successfully!
Jun 29 19:34:23 bioinstrumentation bash[28075]: You can now view my-app in the browser.
Jun 29 19:34:23 bioinstrumentation bash[28075]: Local: http://localhost:3000
Jun 29 19:34:23 bioinstrumentation bash[28075]: On Your Network: http://172.21.50.183:3000
Jun 29 19:34:23 bioinstrumentation bash[28075]: Note that the development build is not optimized.
Jun 29 19:34:23 bioinstrumentation bash[28075]: To create a production build, use npm run build.
Jun 29 19:34:23 bioinstrumentation bash[28075]: webpack compiled successfully

```

Fig 45. Servicio systemctl para el frontend

## F. Prueba de usabilidad

La evaluación de la aplicación se realizó mediante una prueba de usabilidad que constaba de once preguntas de selección múltiple y dos de respuesta abierta. Las preguntas de selección fueron evaluadas con una escala de 1 a 5, en donde, 1 es totalmente en desacuerdo y 5 es totalmente de acuerdo.

Para realizar esta evaluación se encuestaron ocho participantes, en donde inicialmente se entregó una serie de indicaciones para que el usuario hiciera las pruebas necesarias para dar solución al cuestionario. Las indicaciones entregadas fueron las siguientes:

## TEST USABILIDAD

ElectroEG es una aplicación especializada para el procesamiento y análisis de calidad de datos de electroencefalografía.

A continuación, se describe la información y pasos que deberá seguir:

1. La aplicación en estos momentos se encuentra diseñada únicamente para un administrador, por lo tanto, no es necesario loguearse. Sin embargo, hay notificaciones que le permite saber al usuario si se encuentra dentro del sistema o no, ya que algunas pestañas son únicamente de acceso autorizado.
2. Base de datos a cargar, se encuentra en el escritorio en una carpeta llamada "CARPETA\_PRUEBA\_BASES\_DATOS". Cargue la base de datos que se le indique de forma descomprimida en la pestaña de 'Carga de datos'.
3. Ingrese los campos solicitados y guarde la información una vez completada. Esta información solicitada requiere del conocimiento del estándar BIDS. Por lo tanto, se proporciona las indicaciones para subir de acuerdo con la base de datos y al estándar. Aquí podrá ver las indicaciones: [https://udeaeducomy.sharepoint.com/:w:/g/personal/luisa\\_zapatas\\_udea\\_edu\\_co/EQ319pohV1BMjn1fwLS1gwEBeaPE8NnbeH2e625o-H332g?e=6fldtF](https://udeaeducomy.sharepoint.com/:w:/g/personal/luisa_zapatas_udea_edu_co/EQ319pohV1BMjn1fwLS1gwEBeaPE8NnbeH2e625o-H332g?e=6fldtF).
4. Elimine la base de datos, vuelva y repita los pasos 2 y 3.
5. Una vez cargada la base de datos e ingresado la información solicitada, podrá procesar los datos. Procese únicamente la base de datos cargada (Este proceso suele tardar dependiendo de la cantidad de datos y la duración de la señal). Una vez sea notificado que fue procesada por medio de correo electrónico, elimine la base de datos.
6. Visualice la información de las demás bases de datos previamente procesada en la pestaña de control de calidad de datos. Navegue por cada una de las secciones e interactúe con las herramientas que le proporciona cada una de las interfaces.

En la Tabla XIV, se encuentra los resultados obtenidos de la prueba, en donde cada pregunta de acuerdo con la valoración resultante de todos los participantes se obtuvo un porcentaje. La



mayor puntuación por cada una de las preguntas se resaltó con un formato en negrita para identificar aspectos a mejorar.

Tabla XIV. Resultados de la prueba de usabilidad

<b>PREGUNTAS DE SELECCIÓN MÚLTIPLE</b>					
<b>Preguntas/Puntuación</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
La legibilidad de la información me parece adecuada.	0%	0%	12,5%	<b>50%</b>	37,5%
Las secciones de las que consta la página, menú, cabecera, etcétera, están claramente definidas de modo que queda claro a qué sección pertenece cada elemento de la página.	0%	25%	12,5%	<b>37,5%</b>	25%
Al navegar por la página se diferencia claramente la profundidad del árbol de navegación y la información a la que se va a poder acceder.	0%	0%	0%	<b>62,5%</b>	37,5%
Existen ayudas a la navegación y éstas son útiles y suficientes.	0%	12,5%	12,5%	<b>62,5%</b>	12,5%
Existen elementos en la página que dificultan la navegación o el sistema de navegación es difícil de utilizar.	25%	<b>50%</b>	0%	12,5%	12,5%
El número de pasos requeridos para acceder a la información es alto, por lo que la navegación es tediosa.	25%	<b>50%</b>	25%	0%	0%
Hay elementos que me distraen de la navegación.	<b>87,5%</b>	12,5%	0%	0%	0%
En general, considero que la usabilidad de esta página es buena.	0%	0%	12,5%	12,5%	<b>75%</b>
Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema.	37,5%	<b>50%</b>	12,5%	0%	0%
Encontré el sistema muy complejo de usar.	<b>62,5%</b>	37,5%	0%	0%	0%
Creo que necesito aprender muchas cosas antes de poder empezar con este sistema.	37,5%	<b>50%</b>	12,5%	0%	0%
<b>PREGUNTAS ABIERTAS</b>					
De forma general, ¿Cómo se sintió utilizando la aplicación?					
Recomendaciones o sugerencias.					

Los resultados promedios de las puntuaciones son representados a partir de un diagrama de barras mostrado en la Fig 46, en donde cada una de estas, representan cada una de las preguntas presentadas en la Tabla XIV, respectivamente. Las respuestas de las preguntas abiertas se encuentran en la sección de los anexos.

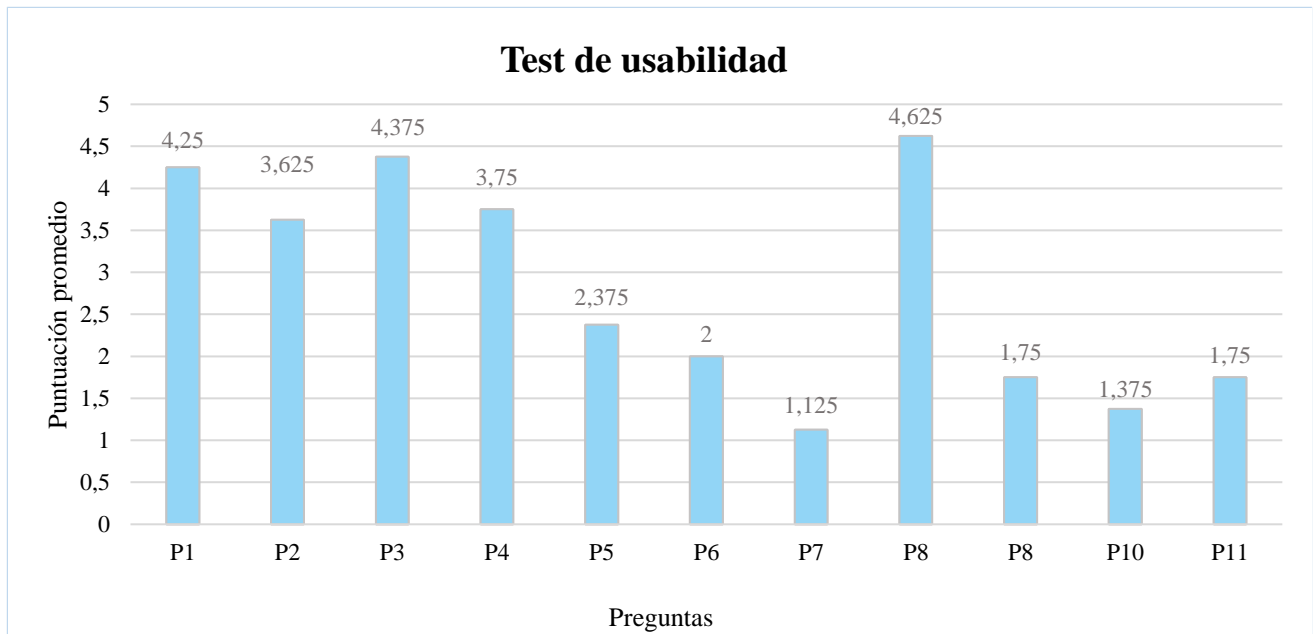


Fig 46. Resultados promedio de la prueba de usabilidad. El eje 'X' representa la puntuación promedio de los participantes encuestados, y el eje 'Y' hace referencia a cada una de las preguntas de selección múltiple.

Los resultados obtenidos mediante esta prueba demuestran que en general la aplicación es fácil de navegar y resulta útil para la utilización de estos sistemas, sin embargo, hay información que requiere de correcciones como lo es identificación del árbol de navegación, aspecto de las gráficas como rejilla, tamaños, y demás. En general, los comentarios y resultados de la encuesta apuntan a que la aplicación es buena para ser un primer producto mínimo viable.

---

## CONCLUSIONES

La aplicación fue diseñada buscando estrategias que permitieran compatibilidad e interoperabilidad. Para abordar este desafío se utilizó el estándar BIDS que permite la optimización en los procesos de organización y el manejo de grandes volúmenes de datos, facilitando la centralización de la información, en donde cada base de datos tiene tanto la información de los archivos crudos como la de los datos procesados siguiendo la nomenclatura y reglas de este estándar. BIDS es una herramienta que tiende a ser muy útil para el manejo e investigación de volúmenes multimodales.

Además, la creación de servicios para la ejecución en paralelo de las tareas permite mejorar el uso de los recursos del sistema, logrando recibir solicitudes durante la ejecución del preprocesamiento, el cual es una etapa que tiene una duración de ejecución muy extensa. Esta metodología permite que cualquier usuario al arrancar un servidor pueda utilizar la aplicación, por tanto, realizar gestión de servicios es una forma de realizar pruebas de despliegues de la aplicación para verificar que los procesos corran efectivamente.

Por otra parte, la prueba de usabilidad permitió conocer diferentes puntos de vista, para realizar procesos de mejora en el desarrollo de la aplicación, en donde la mayoría de las recomendaciones apuntan a mejorar la información entregada para el usuario y a mejorar el contenido de algunas interfaces; en síntesis, este tipo de pruebas son importantes para verificar que las aplicaciones sean fáciles de manejar y útiles, de forma que se encuentre toda la información necesaria del lado del usuario cumpliendo los criterios de la prueba SUS.

En conclusión, el desarrollo de la aplicación permite centralizar la información, procesar los datos de forma organizada y realizar diversos procesos de análisis de calidad que son importantes antes de realizar cualquier tipo de estudio. El reporte de calidad de datos indica por medio de la evaluación de las métricas si los datos analizados presentan alta cantidad de datos atípicos.

Para futuros trabajos se propone el desarrollo de estrategias que permitan optimizar el tiempo de generación de gráficos bajo demanda de manera más rápida y accesible, teniendo en cuenta los aspectos que ya se implementaron en el desarrollo de este proyecto. Además de buscar metodologías y opciones para el despliegue de la aplicación teniendo en cuenta el tamaño de entrada de los datos y el tamaño de salida de estos. El desafío que implicaría tener un servicio de usuarios que genere demandas al mismo tiempo.

## REFERENCIAS

- [1] L. Yaneth *et al.*, “TECNOLOGÍAS DE LA INFORMACIÓN Y LA COMUNICACIÓN (TICS) EN EL SECTOR SALUD,” 2013.
- [2] E. Menasalvas and C. Gonzalo Alejandro Rodríguez-González, “BIG DATA EN SALUD: RETOS Y OPORTUNIDADES.”
- [3] M. Beier, T. Penzel, and D. Krefting, “A Performant Web-Based Visualization, Assessment, and Collaboration Tool for Multidimensional Biosignals,” *Front. Neuroinform.*, vol. 13, p. 65, Sep. 2019, doi: 10.3389/FNINF.2019.00065/BIBTEX.
- [4] A. Estiven and A. R. Escudero, “DESARROLLO Y DESPLIEGUE DE UNA PLATAFORMA WEB PARA EL ALMACENAMIENTO Y ADMINISTRACIÓN DE BIOSEÑALES.”
- [5] L. Dong *et al.*, “WeBrain: A web-based brainformatics platform of computational ecosystem for EEG big data analysis,” *Neuroimage*, vol. 245, p. 118713, Dec. 2021, doi: 10.1016/J.NEUROIMAGE.2021.118713.
- [6] S.-P. Kim, “Preprocessing of EEG,” pp. 15–33, 2018, doi: 10.1007/978-981-13-0908-3\_2.
- [7] N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. M. Su, and K. A. Robbins, “The PREP pipeline: Standardized preprocessing for large-scale EEG analysis,” *Front. Neuroinform.*, vol. 9, no. JUNE, pp. 1–19, Jun. 2015, doi: 10.3389/FNINF.2015.00016/BIBTEX.
- [8] B. Abdi-Sargezeh, R. Foodeh, V. Shalchyan, and M. R. Daliri, “EEG artifact rejection by extracting spatial and spatio-spectral common components,” *J. Neurosci. Methods*, vol. 358, p. 109182, Jul. 2021, doi: 10.1016/J.JNEUMETH.2021.109182.
- [9] C. R. Pernet *et al.*, “EEG-BIDS, an extension to the brain imaging data structure for electroencephalography,” *Sci. Data* 2019 61, vol. 6, no. 1, pp. 1–5, Jun. 2019, doi: 10.1038/s41597-019-0104-8.
- [10] A. Biasiucci, B. Franceschiello, and M. M. Murray, “Electroencephalography,” *Curr. Biol.*, vol. 29, no. 3, pp. R80–R85, Feb. 2019, doi: 10.1016/J.CUB.2018.11.052.
- [11] S. J. M. Smith and S. Smith, “EEG in the diagnosis, classification, and management of patients with epilepsy,” *J. Neurol. Neurosurg. Psychiatry*, vol. 76, no. suppl 2, pp. ii2–ii7, Jun. 2005, doi: 10.1136/JNNP.2005.069245.
- [12] R. W. Homan, J. Herman, and P. Purdy, “Cerebral location of international 10-20 system

- electrode placement,” *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, 1987, doi: 10.1016/0013-4694(87)90206-9.
- [13] D. Yao, Y. Qin, S. Hu, L. Dong, M. L. Bringas Vega, and P. A. Valdés Sosa, “Which Reference Should We Use for EEG and ERP practice?,” *Brain Topogr.*, vol. 32, no. 4, pp. 530–549, Jul. 2019, doi: 10.1007/S10548-019-00707-X/FIGURES/12.
- [14] G. Qi, S. Zhao, A. (Avi) Ceder, W. Guan, and X. Yan, “Wielding and evaluating the removal composition of common artefacts in EEG signals for driving behaviour analysis,” *Accid. Anal. Prev.*, vol. 159, p. 106223, Sep. 2021, doi: 10.1016/J.AAP.2021.106223.
- [15] R. Pérez Raba, “Análisis de la electroencefalografía cuantitativa y sus aplicaciones en el ámbito médico,” no. ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN, 2018.
- [16] C. Babiloni *et al.*, “International Federation of Clinical Neurophysiology (IFCN) – EEG research workgroup: Recommendations on frequency and topographic analysis of resting state EEG rhythms. Part 1: Applications in clinical research studies,” *Clin. Neurophysiol.*, vol. 131, no. 1, pp. 285–307, Jan. 2020, doi: 10.1016/J.CLINPH.2019.06.234.
- [17] C. S. Nayak and A. C. Anilkumar, “EEG Normal Waveforms,” *StatPearls*, pp. 1–6, May 2022, Accessed: Jun. 06, 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK539805/>.
- [18] F. Li *et al.*, “Differentiation of Schizophrenia by Combining the Spatial EEG Brain Network Patterns of Rest and Task P300,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 594–602, Apr. 2019, doi: 10.1109/TNSRE.2019.2900725.
- [19] S. Enriquez-Geppert, R. J. Huster, and C. S. Herrmann, “EEG-neurofeedback as a tool to modulate cognition and behavior: A review tutorial,” *Front. Hum. Neurosci.*, vol. 11, p. 51, Feb. 2017, doi: 10.3389/FNHUM.2017.00051/BIBTEX.
- [20] X. Zhao, C. Liu, Z. Xu, L. Zhang, and R. Zhang, “SSVEP Stimulus Layout Effect on Accuracy of Brain-Computer Interfaces in Augmented Reality Glasses,” *IEEE Access*, vol. 8, pp. 5990–5998, 2020, doi: 10.1109/ACCESS.2019.2963442.
- [21] K. J. Gorgolewski *et al.*, “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Sci. Data* 2016 31, vol. 3, no. 1, pp. 1–9, Jun. 2016, doi: 10.1038/sdata.2016.44.
- [22] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and

- stewardship,” *Sci. Data* 2016 31, vol. 3, no. 1, pp. 1–9, Mar. 2016, doi: 10.1038/sdata.2016.18.
- [23] S.-P. Kim, “Preprocessing of EEG,” 2018, doi: 10.1007/978-981-13-0908-3\_2.
- [24] P. Ablin, J. F. Cardoso, and A. Gramfort, “Faster independent component analysis by preconditioning with hessian approximations,” *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4040–4049, Aug. 2018, doi: 10.1109/TSP.2018.2844203.
- [25] K. Yasoda, R. S. Ponmagal, K. S. Bhuvaneshwari, and K. Venkatachalam, “Automatic detection and classification of EEG artifacts using fuzzy kernel SVM and wavelet ICA (WICA),” *Soft Comput.*, vol. 24, no. 21, pp. 16011–16019, Nov. 2020, doi: 10.1007/S00500-020-04920-W/TABLES/3.
- [26] R. Swarnkar and K. P. Miyapuram, “Temporal EEG Neural Activity Predicts Visuo-Spatial Motor Sequence Learning,” *Commun. Comput. Inf. Sci.*, vol. 1333, pp. 204–211, Nov. 2020, doi: 10.1007/978-3-030-63823-8\_25.
- [27] N. Bigdely-Shamlo, J. Touryan, A. Ojeda, C. Kothe, T. Mullen, and K. Robbins, “Automated EEG mega-analysis I: Spectral and amplitude characteristics across studies,” *Neuroimage*, vol. 207, p. 116361, Feb. 2020, doi: 10.1016/J.NEUROIMAGE.2019.116361.
- [28] P. A. Valdes-Sosa *et al.*, “The Cuban Human Brain Mapping Project, a young and middle age population-based EEG, MRI, and cognition dataset,” *Sci. Data* 2021 81, vol. 8, no. 1, pp. 1–12, Feb. 2021, doi: 10.1038/s41597-021-00829-7.
- [29] F. Taverna *et al.*, “BIOMEX: an interactive workflow for (single cell) omics data interpretation and visualization,” *Nucleic Acids Res.*, vol. 48, no. W1, pp. W385–W394, Jul. 2020, doi: 10.1093/NAR/GKAA332.
- [30] S. Ramírez, “tiangolo/fastapi: FastAPI framework, high performance, easy to learn, fast to code, ready for production.” <https://github.com/tiangolo/fastapi> (accessed Jul. 13, 2022).
- [31] Néstor Eduardo Canchigña Catagña, “Desarrollo de un prototipo de aplicación web para el monitoreo de dispositivos en un WISP,” Escuela politécnica nacional, Quito, 2021.
- [32] Y. Li, Z. Yang, X. Chen, H. Yuan, and W. Liu, “A stacking model using URL and HTML features for phishing webpage detection,” *Futur. Gener. Comput. Syst.*, vol. 94, pp. 27–39, May 2019, doi: 10.1016/J.FUTURE.2018.11.004.
- [33] J. Eguíluz Pérez, “Introducción a JavaScript,” Accessed: Jul. 13, 2022. [Online]. Available: [www.librosweb.es](http://www.librosweb.es).

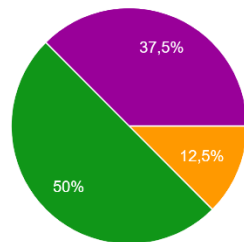
- 
- [34] P. Kirkbride, “systemd,” *Basic Linux Termin. Tips Tricks*, pp. 221–234, 2020, doi: 10.1007/978-1-4842-6035-7\_11.
- [35] G. W. Sasmito, L. O. M. Zulfiqar, and M. Nishom, “Usability Testing based on System Usability Scale and Net Promoter Score,” *2019 2nd Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2019*, pp. 540–545, Dec. 2019, doi: 10.1109/ISRITI48646.2019.9034666.
- [36] F. J. García-Pretelt, J. X. Suárez-Relevo, D. F. Aguillon-Niño, F. J. Lopera-Restrepo, J. F. Ochoa-Gómez, and C. A. Tobón-Quintero, “Automatic Classification of Subjects of the PSEN1-E280A Family at Risk of Developing Alzheimer’s Disease Using Machine Learning and Resting State Electroencephalography,” *J. Alzheimers. Dis.*, vol. 87, no. 2, pp. 817–832, May 2022, doi: 10.3233/JAD-210148.
- [37] J. Bosch-Bayard, L. Galan, E. Aubert Vazquez, T. Virues Alba, and P. A. Valdes-Sosa, “Resting State Healthy EEG: The First Wave of the Cuban Normative Database,” *Front. Neurosci.*, vol. 14, p. 1241, Dec. 2020, doi: 10.3389/FNINS.2020.555119/BIBTEX.
- [38] T. W. Rygvold, C. Hatlestad-Hall, T. Elvsåshagen, T. Moberget, and S. Andersson, “Do visual and auditory stimulus-specific response modulation reflect different mechanisms of neocortical plasticity?,” *Eur. J. Neurosci.*, vol. 53, no. 4, pp. 1072–1085, Feb. 2021, doi: 10.1111/EJN.14964.



## ANEXOS

La legibilidad de la información me parece adecuada

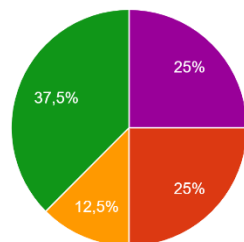
8 respuestas



- 1: Totalmente en desacuerdo
- 2: En desacuerdo
- 3: Neutral
- 4: De acuerdo
- 5: Totalmente de acuerdo

Las secciones de las que consta la página, menú, cabecera, etcétera, están claramente definidas de modo que queda claro a qué sección pertenece cada elemento de la página

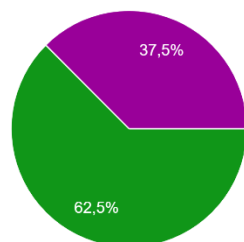
8 respuestas



- 1: Totalmente en desacuerdo
- 2: En desacuerdo
- 3: Neutral
- 4: De acuerdo
- 5: Totalmente de acuerdo

Al navegar por la página se diferencia claramente la profundidad del árbol de navegación y la información a la que se va a poder acceder.

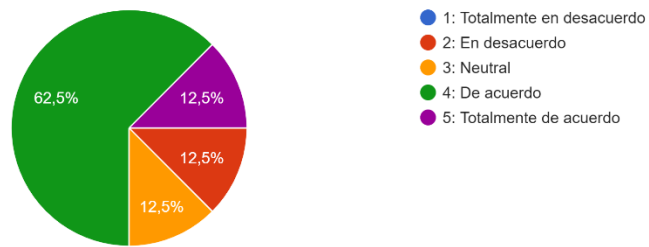
8 respuestas



- 1: Totalmente en desacuerdo
- 2: En desacuerdo
- 3: Neutral
- 4: De acuerdo
- 5: Totalmente de acuerdo

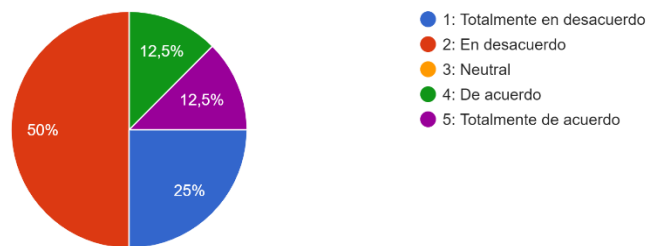
Existen ayudas a la navegación y éstas son útiles y suficientes.

8 respuestas



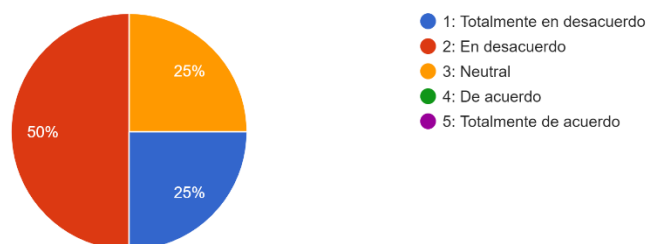
Existen elementos en la página que dificultan la navegación o el sistema de navegación es difícil de utilizar

8 respuestas



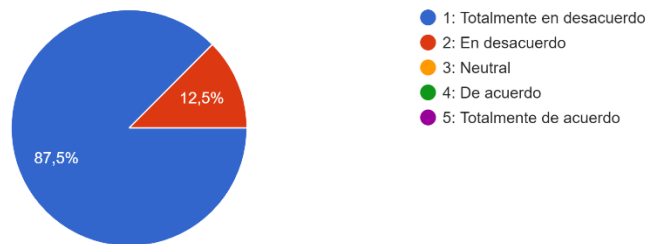
El número de pasos requeridos para acceder a la información es alto, por lo que la navegación es tediosa

8 respuestas



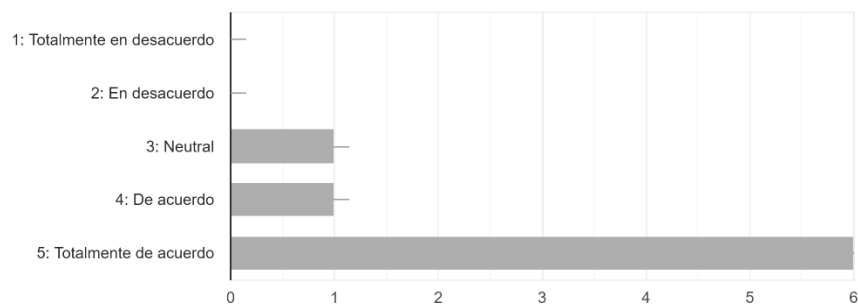
Hay elementos que me distraen de la navegación.

8 respuestas



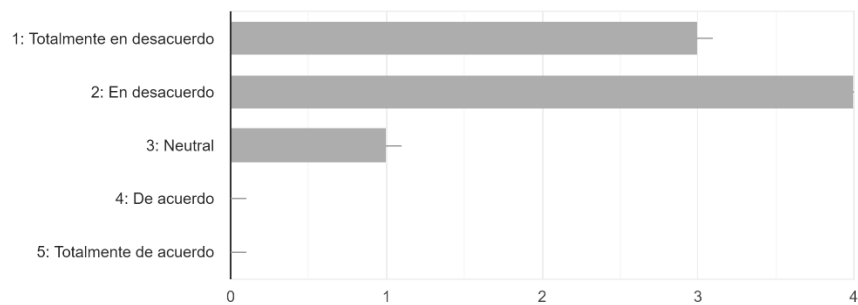
En general, considero que la usabilidad de esta página es buena.

8 respuestas



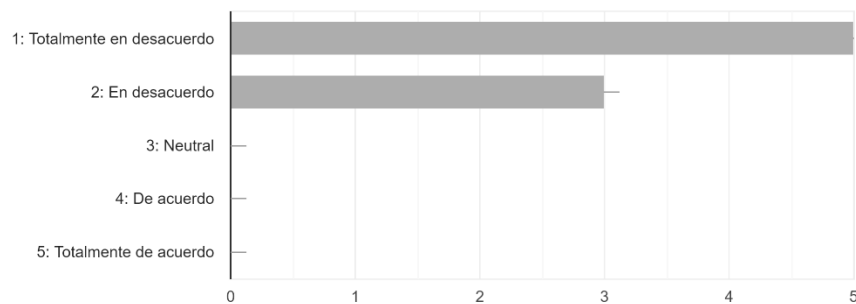
Creo que necesitaría el apoyo de un técnico para poder utilizar este sistema

8 respuestas



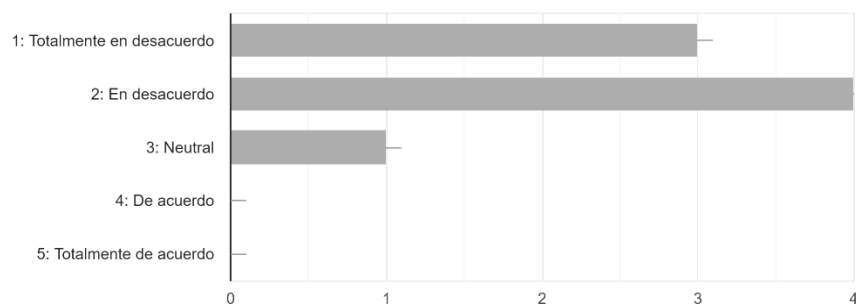
Encontré el sistema muy complejo de usar

8 respuestas



Creo que necesito aprender muchas cosas antes de poder empezar con este sistema.

8 respuestas



### De forma general, ¿Cómo se sintió utilizando la aplicación?

- Es una aplicación intuitiva fácil de manejar. Requiere de conocimientos previos sobre EEG más avanzados a los vistos en cursos de pregrado, sin embargo, en cuanto a la página web como tal se entiende su navegación y organización. Los 3 posibles usos que tiene la aplicación están bastante claros y bien divididos
- La aplicación es bastante intuitiva desde el punto de vista del software. Es interesante poder seguir navegando por la página mientras el procesamiento de otras bases de datos se está generando.
- Muy bien, muy útil es intuitiva
- Muy bien. Creo que la aplicación es lo suficientemente clara para una persona con conocimientos de EEG
- Me sentí muy cómoda, la aplicación es muy amigable
- Muy cómoda, es amigable y el diseño me gusta
- Es una aplicación muy buena en cuanto a usabilidad. Sin embargo, hay temas de diseño que necesitan mejorarse para que sea más fácil de navegar en el sitio web.
- Me parece que tiene una interfaz muy interactiva y fácil de manejar. Es clara y concisa.

## Recomendaciones o sugerencias

- Usar leyendas y labels en español. Usar labels en las gráficas aclarando que representan los colores. Usar elementos visuales que le permitan saber al usuario que se está cargando la información.
- Establecer un tamaño estándar para las gráficas, seleccionar colores diferentes cuando se gráfica varias cosas en la misma gráfica, también mencionar una pequeña descripción de cada parámetro y como requiere de altos tiempos de procesamiento mostrar de alguna manera que ese proceso se está realizando y cuando este finalice. Mostrar el el menú de navegación en que sección de la página se encuentra el usuario.
- Mensajes de espera a la hora de cargar la página.
- Notas de que cada métrica o grafico que se saca y como comprenderlo mirar que manera es mas intuitiva y adecuada.
- Mejor delineación de secciones principales(visual)
- Quitarle el hue cuando solo se tienen datos preprocesados.
- Centrar algunas gráficas y algunas cajas o violín dentro del plot para que estén bien centrados.
- Colores de la página todos parecidos
- Textos de ayuda.
- Grids en plots.
- Alguna recomendación en la mejora de la estética de algunas imágenes
- Tamaño de títulos, corrección de ortografía, añadir párrafos introductorios en las gráficas, espacios entre gráficas, colores en títulos, indicador de estado en página, añadir logos en páginas secundarias, añadir título de la aplicación en el mensaje que envía al correo
- Espacio visual, tamaño de los textos, colores, franjas, añadir descripciones, distribución del logo, transiciones, diagramación. Tipografía (definir qué textos van en bold y cuáles en light). Mejor orden en las tablas. Mejor distribución del menú.
- Distribución de espacios, tamaño de letras, explicación breve de interpretación de gráficos, más imágenes, íconos en bases de datos y resultados, utilizar más el logo, reconocimiento de marca en correo electrónico, asunto en el correo, indicación de paso a seguir al procesamiento de datos, signos de admiración.