

Análisis de Redes de Conectividad Cerebral en Pacientes Epilépticos a través de su  
Representación en Espacios de Baja Dimensionalidad

Steven Fernando Rico Aparicio

Tesis presentada para optar al título de Magíster en Física

Director  
PhD. Mario Chavez  
Director de Investigación, CNRS

Co-Director  
PhD. Luis Alberto Núñez  
UIS

Universidad Industrial de Santander  
Facultad de Ciencias Básicas  
Escuela de Física  
Maestría en Física  
Bucaramanga  
2025

## Dedicatoria

La totalidad de este trabajo y mi camino como académico, como persona, siempre ha sido y será dedicado a mi madre, Subsy, y a mi hermano, Walter. Mis pilares en este mundo.

A William, quien con su amor, compañía, fortaleza y paciencia me ha acompañado de la mano y me ha motivado a seguir adelante.

Este paso no es solo mío, también de ustedes.

# Agradecimientos

He tenido la fortuna en mi vida de contar con gran apoyo incondicional.

Agradezco a mis abuelos, Rosa y Vicente. A Dayana, James, Ludy y Milena. Familia que nunca ha apagado la llama del cariño y alegría.

A mis amigos: Luigui, María Fernanda y Rafael; y aquellos que están lejos pero siempre muy cerca: Danna, Laura, Lucía, Rolando y Juan Manuel.

A Mario, Luis y Johann. Más que profesores, mentores y guías.

## Tabla de Contenido

<b>Introducción</b>	<b>9</b>
<b>1. Redes complejas en Neurociencia</b>	<b>12</b>
1.1. El cerebro como un sistema complejo	12
1.2. Geometría latente de las redes cerebrales	13
1.3. El caso de la epilepsia	14
1.4. Electroencefalografía	15
1.4.1. Electroencefalografía Intracranial	16
<b>2. Descripción de pacientes y datos iEEG</b>	<b>18</b>
2.1. Características clínicas	18
2.1.1. Desbalance entre clases	20
2.2. Conectividad funcional en señales iEEG	21
2.2.1. Método <i>Phase Locking Value</i>	22
2.2.2. Método <i>Phase Locking Value</i>	22
<b>3. Construcción, tratamiento y proyección de las redes de conectividad cerebral</b>	<b>26</b>
3.1. Construcción de las redes de conectividad	26
3.2. Filtrado de las redes de conectividad	27
3.2.1. Filtrado por Árbol Recubridor Mínimo	27
3.2.2. Filtrado de Grado Medio $\langle k \rangle$	28
3.3. Proyección de las redes de conectividad en espacio Euclídeo	29
3.3.1. Método de reducción de dimensionalidad por Mapas de Difusión	30
3.4. Alineación de las redes proyectadas en $\mathbb{R}^2$	33
3.4.1. Método de Análisis de Procrustes Generalizado	33
<b>4. Discriminación estadística entre estados Interictales y Preictales</b>	<b>36</b>
4.1. Discriminación de distribuciones locales	36
4.2. Identificación de nodos relevantes para la discriminación de estados	39
4.3. Biomarcador preictal $\mathcal{B}$	40
<b>5. Detección y predicción de estados preictales</b>	<b>45</b>
5.1. Resultados de clasificación de estados	45
5.2. Resultados de predicción diaria de riesgo de crisis	50
<b>6. Conclusiones y discusión</b>	<b>54</b>
<b>Referencias Bibliográficas</b>	<b>57</b>

**Lista de Figuras**

Figura 1.1.	<i>Representación en grafos de la dinámica cerebral.</i>	13
Figura 1.2.	<i>Sistema típico de colocación de electrodos en EEG</i>	16
Figura 1.3.	<i>Trazado de señales EEG de un episodio epiléptico</i>	17
Figura 2.1.	<i>Matrices de conectividad funcional</i>	25
Figura 3.1.	<i>Visualización del proceso de filtrado de las redes de conectividad</i>	29
Figura 3.2.	<i>Proyección y alineación de las redes de conectividad funcional en <math>\mathbb{R}^2</math></i>	35
Figura 4.1.	<i>Casos de configuración nodal en el espacio proyectado</i>	38
Figura 4.2.	<i>Histograma polar de la prevalencia de nodos representativos</i>	41
Figura 4.3.	<i>Histograma de valor del biomarcador <math>\mathcal{B}</math></i>	44
Figura 5.1.	<i>Resultado de clasificación</i>	46
Figura 5.2.	<i>Matrices de confusión</i>	47
Figura 5.3.	<i>Resultados de predicción pseudo-prospectiva por paciente</i>	51

**Lista de Tablas**

<i>Tabla 2.1. Características clínicas de los pacientes incluidos en el estudio</i>	19
<i>Tabla 2.2. Desbalance de clases por paciente y a nivel global</i>	21
<i>Tabla 5.1. Métricas globales de discriminación entre estados interictales y preictales</i>	49
<i>Tabla 5.2. Comparación de métricas globales entre diferentes enfoques de clasificación.</i>	50
<i>Tabla 5.3. Métricas globales de predicción pseudo-prospectiva de estados preictales</i>	53
<i>Tabla 5.4. Comparación de métricas globales entre diferentes enfoques de predicción.</i>	53

# Resumen

**Título:** Análisis de Redes de Conectividad Cerebral en Pacientes Epilépticos a través de su Representación en Espacios de Baja Dimensión

**Autor:** Steven Fernando Rico Aparicio

**Palabras Clave:** Epilepsia, *Embedding* Euclideo, Análisis de Datos, Neurociencia

**Descripción:** La identificación de estados preictales, periodos en los cuales la probabilidad de ocurrencia de una crisis epiléptica aumenta, sigue siendo un desafío central en la neurociencia computacional clínica. En este trabajo, se propone un marco metodológico novedoso que consiste en proyectar redes de conectividad cerebral funcional, obtenidas de señales registradas de electroencefalografía intracraneal (iEEG), en un espacio euclidiano de baja dimensión. Esta representación compacta permite capturar características topológicas esenciales de la dinámica cerebral y facilita la detección de cambios sutiles en la conectividad que preceden a una crisis. A partir de técnicas estándar de aprendizaje automático, se define un biomarcador adimensional,  $\mathcal{B}$ , capaz de discriminar entre estados interictales (libres de crisis) y preictales (dentro de las 24 horas previas a una crisis). El método desarrollado se centra en los patrones de conectividad de un subconjunto de electrodos iEEG informativos, lo cual permite una clasificación robusta de los estados cerebrales a lo largo del tiempo. Cuantificamos nuestro enfoque mediante un esquema de *leave-one-out cross-validation* y una estrategia de predicción pseudo-prospectiva, evaluando el desempeño mediante métricas como el *F1-score* y la precisión balanceada. Los resultados muestran que las proyecciones euclidianas de baja dimensión aplicadas a la conectividad iEEG producen marcadores interpretables y predictivos de la actividad preictal, lo que ofrece implicaciones prometedoras para la predicción en tiempo real de crisis epilépticas y para el diseño de intervenciones terapéuticas personalizadas.

# Abstract

**Title:** Analysis of brain connectivity networks in epileptic patients through their representation in low- dimensional spaces

**Author:** Steven Fernando Rico Aparicio

**Keywords:** Euclidean manifold embedding, Epilepsy, Data analysis, Neuroscience.

**Description:** Identifying preictal states, periods during which seizures are more likely to occur, remains a central challenge in clinical computational neuroscience. In this study, we introduce a novel framework that embeds functional brain connectivity networks, derived from intracranial EEG (iEEG) recordings, into a low-dimensional Euclidean space. This compact representation captures essential topological features of brain dynamics and facilitates the detection of subtle connectivity changes preceding seizures. Using standard machine learning techniques, we define a dimensionless biomarker,  $\mathcal{B}$ , that discriminates between interictal (seizure-free) and preictal (within 24 hours of seizure) network states. Our method focuses on connectivity patterns among a subset of informative iEEG electrodes, enabling robust classification of brain states across time. We validate our approach using a leave-one-out cross-validation scheme and a pseudo-prospective forecasting strategy, assessing performance with metrics such as F1-score and balanced accuracy. Results show that low-dimensional Euclidean embeddings of iEEG connectivity yield interpretable and predictive markers of preictal activity, offering promising implications for real-time seizure forecasting and individualized therapeutic interventions.

# Introducción

El presente trabajo se enmarca en un enfoque interdisciplinario que interseca la física, la neurociencia clínica y la ciencia de datos. La física encuentra aquí una aplicación directa en la caracterización matemática de la actividad cerebral, proponiendo herramientas cuantitativas capaces de revelar patrones latentes en la conectividad neuronal a través del análisis de señales eléctricas cerebrales, la cuantificación de su sincronización y conceptos fundamentales como la física estadística, la dinámica no lineal de sistemas complejos, la teoría espectral, entre otros.

A lo largo de la historia de la física, uno de los desafíos más persistentes ha sido la descripción de sistemas compuestos por múltiples cuerpos en interacción. Desde la mecánica clásica de los sistemas de  $N$  cuerpos hasta la mecánica estadística de gases y sólidos se han desarrollado herramientas teóricas potentes, pero muchas veces limitadas por la complejidad combinatoria y el carácter emergente de ciertos fenómenos colectivos. Frente a esta limitación, ha emergido un nuevo paradigma dentro de la física moderna: el estudio de los sistemas complejos (Vemuri, 1978), en el cual las interacciones entre elementos no se modelan únicamente por ecuaciones diferenciales locales. El campo ha evolucionado orgánicamente apoyado en el formalismo de la teoría de grafos y en herramientas de física estadística, donde la llamada teoría de redes complejas ha permitido modelar y entender sistemas formados por miles de unidades interconectadas, revelando propiedades universales como la robustez estructural, la emergencia de jerarquías y la aparición de fenómenos críticos (Strogatz, 2001). Este enfoque ha cobrado tal relevancia que ha sido reconocido por la comunidad científica con el Premio Nobel de Física 2021 “por sus contribuciones pioneras a nuestra comprensión de los sistemas físicos complejos” otorgado en parte a Giorgio Parisi “por el descubrimiento de la interacción entre el desorden y las fluctuaciones en los sistemas físicos, desde la escala atómica hasta la planetaria” (Nobel Prize Outreach, 2021).

Se encuentran muchas aplicaciones de este enfoque en campos sociales, biológicos, tecnológicos y físicos. En el contexto de la neurociencia presenta un avance prometedor en la caracterización de patrones de la conectividad cerebral, pues ha revelado topologías intrincadas que requieren marcos analíticos sofisticados capaces de capturar su estructura y dinámica subyacentes (Boccaletti et al., 2006). El análisis de redes construidas a partir de señales de electroencefalografía intracraneal (iEEG) ofrece una vía poderosa para explorar la organización funcional del cerebro, facilitando la identificación de patrones emergentes vinculados a estados fisiológicos y patológicos.

En particular, el estudio de la epilepsia ha servido como un modelo clínico relevante donde se han aplicado métricas de conectividad basadas en coherencia, información mutua, sincronización de fase y entropía de transferencia para cuantificar interacciones neuronales y detectar reorganizaciones asociadas a eventos ictales (Van Mierlo et al., 2014; Duncan et al., 2013a; Cousyn et al., 2023). Su naturaleza altamente sincronizada y recurrente permite explorar cómo ciertos estados funcionales del cerebro se reconfiguran antes de un evento patológico, convirtiéndola en un modelo apropiado para el análisis de reorganización en redes cerebrales (Kwan et al., 2011).

El uso de representaciones geométricas latentes ha demostrado ser una herramienta eficaz para analizar estas redes cerebrales. En el caso de pacientes con epilepsia fármaco-resistente, dichas proyecciones han permitido evaluar el impacto de intervenciones como la cirugía sobre regiones específicas del cerebro (Longhena et al., 2024; Guillemaud et al., 2024), lo que refuerza la relevancia de aplicar herramientas de la física de redes complejas al estudio de dinámicas cerebrales alteradas. Ahora bien, en la misma cohorte de pacientes usadas en este trabajo, un estudio reciente exploró la representación de redes funcionales en espacios hiperbólicos combinando *embeddings* geométricos y modelos estadísticos para discriminar entre días interictales y preictales (Guillemaud et al., 2025). Al integrar información de múltiples bandas de frecuencia, dicho enfoque alcanzó un *F1-score* promedio de 0.79 y una precisión del 85 %, lo que pone de manifiesto el potencial de geometrías no euclidianas para capturar firmas funcionales relevantes.

Estudios recientes han debatido si las inmersiones euclidianas o no euclidianas (*embeddings*) ofrecen una representación geométrica más adecuada para las redes funcionales cerebrales (Allard and Serrano, 2020; Boguna et al., 2021). Si bien algunos *embeddings* no euclidianos (como los hiperbólicos) se han considerado superiores para capturar estructuras de conectividad no triviales (jerárquicas o de múltiples escalas) (Allard and Serrano, 2020), su optimización presenta dificultades numéricas significativas (Moreira et al., 2024). Por el contrario, los *embeddings* euclidianos ofrecen una ventaja fundamental: proporcionan una representación geométrica intuitiva con una métrica de distancia bien definida en el espacio  $\mathbb{R}^d$ , lo que facilita un análisis más interpretable de los datos proyectados, incluyendo el cálculo de distancias, baricentros y distribuciones de probabilidad.

En este marco, el presente trabajo propone evaluar si las representaciones euclidianas de redes funcionales cerebrales, construidas a partir de registros intracraneales (iEEG) en estado de reposo, permiten identificar configuraciones topológicas asociadas a cambios en el estado funcional del cerebro. Se adopta como caso de estudio la epilepsia, específicamente la transición entre estados interictales y preictales, con el objetivo de determinar si tales diferencias pueden capturarse mediante métricas geométricas en espacios de baja dimensión. A partir de estas representaciones se plantea un biomarcador

adimensional del riesgo, útil para alimentar modelos de predicción diaria calibrados de forma personalizada. Esta contribución se inscribe en el campo de las neurociencias computacionales, y ofrece un marco geométrico novedoso para representar y clasificar redes funcionales, con potencial aplicación en la identificación de factores con alto valor discriminativo.

La estructura de este documento es la siguiente: en el **Capítulo 1**, se presenta una contextualización de la dinámica cerebral y el registro de señales cerebrales; en el **Capítulo 2**, se describe el conjunto de datos y las etapas de preprocesamiento aplicadas a los registros iEEG; el **Capítulo 3** introduce las herramientas matemáticas utilizadas para construir las redes funcionales y sus proyecciones en espacios euclídeos mediante técnicas de *embedding* y alineamiento geométrico; en el **Capítulo 4**, se propone el biomarcador y se detalla la metodología empleada para su evaluación; el **Capítulo 5** expone los resultados obtenidos en las fases de clasificación y predicción, y finalmente, en el **Capítulo 6**, se discuten las principales conclusiones del estudio y se delinear posibles direcciones para trabajos futuros en el marco de las neurociencias computacionales y la física de redes complejas.

# CAPÍTULO 1

## Redes complejas en Neurociencia

### 1.1. El cerebro como un sistema complejo

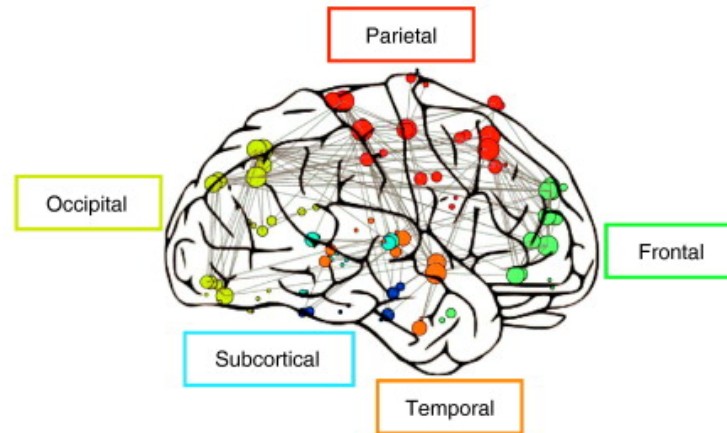
El cerebro humano constituye uno de los sistemas físicos más complejos conocidos en la naturaleza. Su arquitectura se basa en una intrincada red de interconexiones estructurales que involucra cerca de  $10^9$  neuronas y hasta  $10^{12}$  sinapsis (Mountcastle, 1997). Esta organización sostiene capacidades cognitivas, perceptivas y motoras que emergen de la interacción entre unidades neuronales a múltiples escalas temporales y espaciales. Desde la física, el cerebro puede abordarse como un sistema dinámico de muchos cuerpos en interacción, cuya actividad colectiva genera patrones emergentes que no pueden inferirse directamente del comportamiento de sus componentes individuales.

Este tipo de sistemas ha sido estudiado mediante el formalismo de la teoría de sistemas complejos, un marco interdisciplinario que permite describir y analizar fenómenos emergentes a partir de reglas locales de interacción (Strogatz, 2001). En el caso del cerebro, esta aproximación se realiza a través de la construcción de redes funcionales o estructurales donde las unidades del sistema (neuronas, grupos neuronales, regiones corticales, etc.) son representadas como nodos, y las interacciones entre ellas como enlaces o aristas. Estos enlaces pueden definirse a partir de relaciones estadísticas significativas entre señales neurofisiológicas, como correlación, coherencia, sincronización de fase o transferencia de información (Bullmore and Sporns, 2009; Boccaletti et al., 2006).

Gracias a técnicas modernas de neuroimagen y electrofisiología, como la resonancia magnética funcional (fMRI) o la electroencefalografía intracraneal (iEEG), es posible inferir estas relaciones y construir representaciones gráficas que codifican la organización funcional del cerebro como una red compleja (Bassett and Bullmore, 2009; Mubarak, 2020). Estas redes (Figura 1.1) pueden analizarse con herramientas de la teoría de grafos para cuantificar propiedades topológicas como modularidad, eficiencia global, transitividad, longitud de camino medio o centralidad nodal. Tales métricas permiten comparar configuraciones cerebrales bajo distintos estados fisiológicos o patológicos y explorar fenómenos críticos como transiciones súbitas en el patrón de actividad cerebral.

**Figura 1.1.**

*Representación en grafos de la dinámica cerebral.*



*Nota.* Los nodos representan las regiones cerebrales, diferente color para cada región. Los enlaces representan la relación (comunicación) con otras partes del cerebro dentro o fuera de una misma región. *Modificado de Bassett and Gazzaniga (2011).*

## 1.2. Geometría latente de las redes cerebrales

Más allá de su estructura topológica explícita, las redes cerebrales pueden proyectarse en espacios latentes que preservan relaciones funcionales o estructurales de forma geométrica. Este proceso, conocido como *embedding*, permite representar la red en un espacio vectorial de baja dimensión donde los nodos mantienen relaciones de proximidad significativas según su conectividad original (Guillemaud et al., 2024).

Este tipo de representación habilita una descripción métrica del cerebro como un espacio funcional emergente. Al proyectar las redes en espacios como  $\mathbb{R}^q$ , se obtiene una estructura donde es posible calcular distancias, baricentros y distribuciones nodales, facilitando la definición de métricas derivadas que cuantifican el estado global o local de la red. Estas herramientas permiten estudiar cómo varía la organización funcional del cerebro a lo largo del tiempo, o entre condiciones experimentales, mediante desplazamientos geométricos en el espacio proyectado.

Sin embargo, se ha observado que muchas redes funcionales exhiben jerarquías o propiedades auto-similares que no siempre se reflejan fielmente en espacios euclidianos. En respuesta a esta limitación, se han propuesto espacios no euclidianos, como los espacios hiperbólicos, que permiten representar eficientemente estructuras con crecimiento exponencial de conectividad y niveles jerárquicos explícitos (Allard and Serrano, 2020; Boguna et al., 2021). A pesar de las ventajas teóricas de estos espacios,

su implementación numérica suele requerir optimización compleja y computacionalmente costosa ([Moreira et al., 2024](#)).

Por ello, las inmersiones euclidianas conservan un valor práctico considerable. Su métrica definida, interpretación geométrica clara y menor costo computacional las hacen particularmente atractivas para aplicaciones clínicas o exploratorias. Su simplicidad permite, además, incorporar fácilmente estos modelos en esquemas de predicción o clasificación, como se abordará en capítulos posteriores, en el caso específico de la discriminación de estados epilépticos.

### 1.3. El caso de la epilepsia

La epilepsia es un trastorno cerebral caracterizado principalmente por interrupciones recurrentes e impredecibles de la actividad cerebral normal. Estas interrupciones, conocidas como crisis epilépticas, pueden manifestarse de formas muy variadas. Las crisis pueden afectar funciones sensoriales, motoras y autonómicas, así como alterar el estado de conciencia, las emociones, la memoria, la cognición o el comportamiento ([Fisher et al., 2005](#)).

Esta diversidad de manifestaciones clínicas responde a la topografía funcional del cerebro, lo que implica que la presentación de una crisis depende críticamente de su localización inicial. Cuando la descarga epiléptica se inicia y se contiene en una región específica del cerebro, se habla de crisis focal; mientras que, si la actividad anómala se propaga desde el inicio a una región amplia del cerebro y afecta múltiples funciones al mismo tiempo, se conoce como crisis generalizada, donde su síntoma principal son los conocidos episodios convulsivos.

Se estima que hasta un 70% de las personas con epilepsia podrían alcanzar un estado libre de crisis si fueran diagnosticadas y tratadas adecuadamente ([World Health Organization et al., 2019](#)). Sin embargo, también se estima que un 22.5% de los pacientes poseen resistencia a los fármacos medicados y presentan mayor riesgo de muerte prematura, lesiones, disfunción psicosocial y calidad de vida ([Kwan et al., 2011](#)). En estos casos, pueden considerarse intervenciones más invasivas, como la resección quirúrgica de la región cerebral identificada como origen de las crisis. Dada la heterogeneidad clínica de la epilepsia y la individualidad anatómica y funcional de cada paciente, su estudio exige la implementación de nuevas herramientas que integren de forma conjunta la dinámica espacial y temporal de la actividad neuronal.

En este contexto, el estudio de la epilepsia a través de redes de conectividad

cerebral ha cobrado relevancia como una herramienta poderosa para caracterizar la dinámica funcional subyacente a las crisis. La construcción de grafos a partir de registros electrofisiológicos permite identificar reconfiguraciones topológicas que anteceden a eventos ictales y evaluar el impacto de intervenciones terapéuticas sobre la organización de la red cerebral. Trabajos recientes han demostrado que representaciones geométricas latentes de estas redes permiten detectar señales tempranas de desorganización funcional e incluso anticipar la ocurrencia de crisis (Guillemaud et al., 2025). De manera complementaria, se ha mostrado que estas herramientas también permiten cuantificar el efecto de procedimientos quirúrgicos sobre regiones específicas del cerebro en pacientes con epilepsia farmacorresistente, revelando alteraciones topológicas consistentes con la mejoría clínica postoperatoria (Longhena et al., 2024).

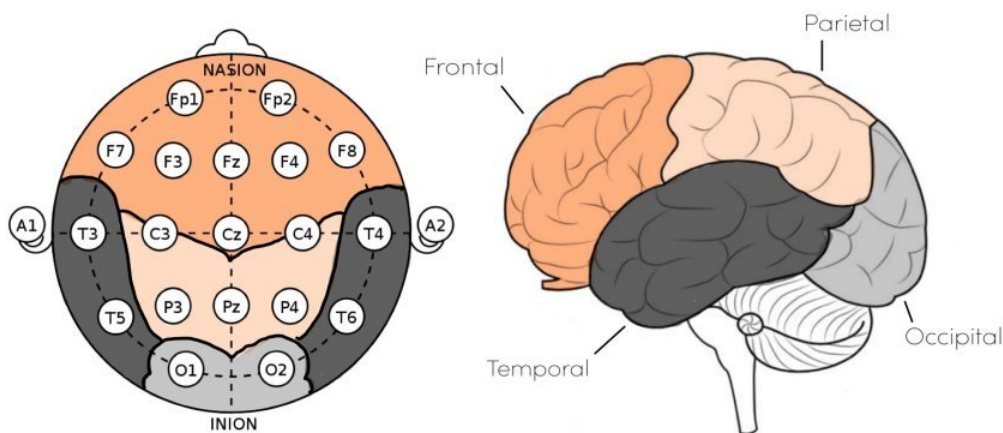
## 1.4. Electroencefalografía

Uno de los métodos más usados para medir las señales eléctricas en el cerebro y su respuesta a estímulos en la epilepsia es la electroencefalografía (EEG). Cuando muchas neuronas corticales se activan de manera sincronizada y con una orientación espacial coherente, se generan corrientes eléctricas; estas crean dipolos que producen campos eléctricos detectables con amplitudes típicas que oscilan entre 10 y 100  $\mu\text{V}$ . La electroencefalografía (EEG) es una técnica no invasiva que permite registrar estos campos eléctricos a través de sensores denominados electrodos al posicionarlos directamente sobre el cuero cabelludo. El posicionamiento de estos electrodos sigue el sistema internacional 10–20, un estándar ampliamente utilizado que garantiza una distribución uniforme y reproducible entre individuos [Figura 1.2](#).

La señal EEG obtenida por los electrodos es una superposición de oscilaciones eléctricas generadas por poblaciones neuronales que interactúan en distintas escalas temporales. La señal se amplifica y filtra de manera analógica para eliminar ruidos de base (señales de parpadeos, micro-movimientos, entre otros) y se digitaliza. Posteriormente se descompone en distintas bandas de frecuencia mediante filtros digitales o por transformada rápida de Fourier (FFT). Las bandas de frecuencia clásicas del EEG incluyen:  $\delta$  (0.5–4 Hz),  $\theta$  (4–7 Hz),  $\alpha$  (8–13 Hz),  $\beta$  (13–30 Hz) y  $\gamma$  (>30 Hz) (Kandel et al., 2000). Estas bandas reflejan distintos ritmos oscilatorios asociados a funciones cognitivas, sensoriales o de estados de conciencia, y constituyen una base para el análisis funcional. Como resultado se aprecian lecturas EEG como la presente en la [Figura 1.3](#) (lectura típica de episodio epiléptico).

**Figura 1.2.**

*Sistema típico de colocación de electrodos en EEG*



*Nota.* Sistema de mapeo 10-20. Los puntos de colocación de los electrodos se nombran con letras que indican la región cerebral subyacente (*F* para frontal, *T* para temporal, *C* para central, *P* para parietal, *O* para occipital), acompañadas de números impares para el hemisferio izquierdo, pares para el derecho, y la letra *z* para los puntos centrales. El nombre 10-20 hace referencia a las distancias relativas entre electrodos, que corresponden al 10% y al 20% de la longitud total entre nasion a inion y de oreja a oreja. Tomado de <https://www.tmsi.artinis.com/blog/the-10-20-system-for-eeeg>.

### 1.4.1. Electroencefalografía Intracraneal

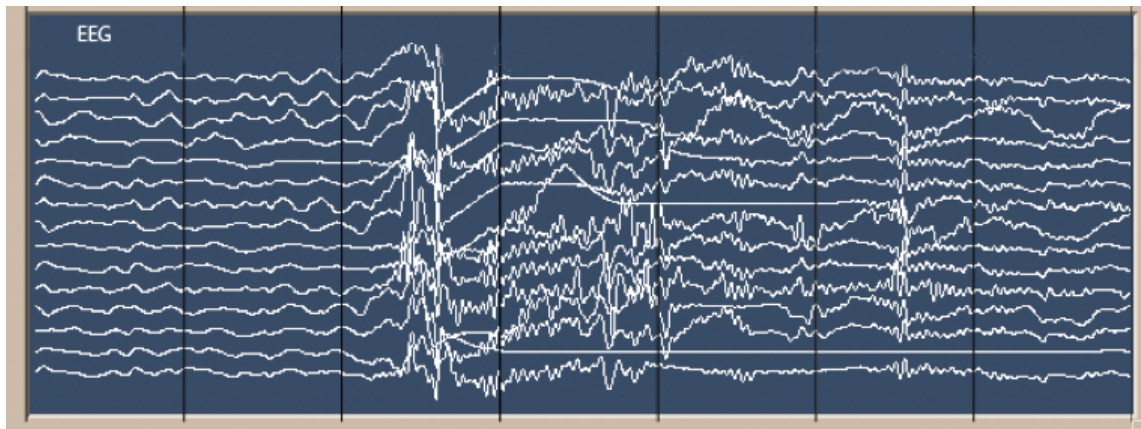
En el presente estudio se trabajó con registros obtenidos mediante una variante de la electroencefalografía conocida como electroencefalografía intracraneal (iEEG). Su funcionamiento, en términos de los principios físicos que rigen la medición de campos eléctricos generados por la actividad sináptica, es el mismo que en el EEG convencional. Sin embargo, a diferencia de este último, la iEEG es una técnica quirúrgica invasiva en la que los electrodos se implantan directamente sobre la superficie cortical, lo que permite una captación más directa y precisa de las señales neuronales.

Gracias a la proximidad de los electrodos al tejido cerebral, la señal iEEG presenta una resolución espacial y temporal significativamente mayor, con menor interferencia por artefactos extracraneales o atenuación provocada por el cráneo y el cuero cabelludo (Mercier et al., 2022). Esta mayor fidelidad convierte a la iEEG en una herramienta fundamental para el estudio detallado de la dinámica neuronal, siendo la técnica de referencia para el análisis clínico y neurofisiológico de la epilepsia (Jobst et al., 2020).

A diferencia del EEG convencional, donde la colocación de electrodos sigue un

**Figura 1.3.**

*Trazado de señales EEG de un episodio epiléptico*



*Nota.* Las barras verticales negras indican una escala temporal de un segundo. Fase de baja amplitud en los primeros 2 segundos; en el segundo 3 Se observa una activación generalizada en múltiples canales, característica de la propagación del evento eléctrico. *Modificado de Ilmoniemi and Sarvas (2019).*

sistema estandarizado como el 10–20, la iEEG emplea configuraciones personalizadas según el objetivo clínico, orientadas al mapeo funcional o a la localización precisa del foco epiléptico. Como ejemplificación, el paciente 0 de este estudio presentó epilepsia focal del lóbulo temporal, zona en la que se implantaron siete (7) electrodos directamente en la región identificada como foco epileptógeno (SOZ), treinta (30) electrodos en una zona circundante irritativa (EIZ) y catorce (14) electrodos de control en otras áreas, siguiendo un protocolo de evaluación individualizada (Tabla 2.1) .

## CAPÍTULO 2

# Descripción de pacientes y datos iEEG

El conjunto de datos utilizado en este estudio fue recolectado entre enero de 2019 y julio de 2021 en la Unidad de Epilepsia del Hospital Universitario Pitié-Salpêtrière (París, Francia), y se encuentra descrito en detalle en [Cousyn et al. \(2023\)](#). El protocolo fue aprobado por un comité institucional de ética (proyectos C11-16 y C19-55), bajo la supervisión del Instituto Nacional de Salud e Investigación Médica de Francia (INSERM), y desarrollado en conformidad con la Declaración de Helsinki.

### 2.1. Características clínicas

Los datos analizados corresponden a una cohorte de 10 pacientes (6 mujeres, 4 hombres; edad entre 18 y 47 años, media = 30.7) con epilepsia focal farmacorresistente. La mayoría presentó focos epilépticos en regiones temporales, particularmente en el lóbulo temporal medial o posterior, con un predominio en el hemisferio izquierdo. El tipo de crisis más común fue la crisis focal con alteración de la conciencia (FIAS), ya sea de forma aislada o combinada con otros tipos. La cantidad total de registros iEEG tomados en días consecutivos por paciente varió entre 4 y 17 días (media  $\approx$  11), mientras que la cantidad de electrodos implantados varía entre 10 y 62 contactos por paciente de acuerdo con las necesidades clínicas particulares de localización del foco epileptógeno. Los detalles demográficos y técnicos específicos de cada paciente pueden consultarse en la [Tabla 2.1](#).

Los registros intracraneales fueron realizados bajo condiciones de monitoreo audiovisual continuo y personal especializado disponible para la atención inmediata de cualquier evento neurológico. Durante cada día de seguimiento, se obtuvo de cada paciente un registro diario de iEEG de 10 minutos en estado de reposo. Cada uno de estos fragmentos fue etiquetado como “**interictal**” si no se presentó ninguna crisis epiléptica en las 24 horas posteriores, o como “**preictal**” si al menos una crisis tuvo lugar dentro de ese mismo intervalo. Como el presente trabajo se enfoca en la clasificación y diferenciación geométrica y estadística de estos dos estados, estos términos clínicos serán referenciados de forma recurrente a lo largo de este escrito.

**Tabla 2.1**  
*Características clínicas de los pacientes incluidos en el estudio*

Paciente	0	1	2	3	4	5	6	7	8	9
<b>Género</b>	F	F	F	F	M	F	M	M	M	F
<b>Edad (años)</b>	47	30	18	25	27	27	39	31	25	38
<b>Foco epileptico</b>	Temporal posterior izquierda	Multifocal (MT derecha y TPO)	Giro frontal superior derecha	MT izquierda	Precentral izquierda	Temporal posterior izquierda	Temporo-polar izquierda	MT izquierda	MT derecha y temporo-polar	MT derecha y temporo-polar
<b>Tipo de crisis</b>	FIAS	FIAS	FIAS	FBTCS	FAS y FIAS	FAS y FIAS	FAS y FIAS	FIAS y FBTCS	FAS y FIAS	FIAS
<b>Posición de electrodos</b>	Bitemporal	Temporo-parieto-occipital derecha	Fronto-temporal derecha	Temporal fronto-izquierda y insular	Pericentral izquierdo	Temporo-parieto-occipital izquierdo	Fronto-temporal izquierdo y frontal derecho	Bitemporal	Temporo-parietal derecho	Temporal derecho
<b>Número de contactos</b>	4	5	4	7	3	4	2	4	10	8
<b>EIZ</b>	30	58	15	13	22	7	13	23	24	6
<b>Otras áreas</b>	14	5	22	12	9	10	25	5	7	17
<b>Número de RS (días)</b>	2	10	3	2	2	2	11	1	3	2
<b>Intercitales</b>	6	1	8	10	2	4	3	16	9	10

*Nota.* Datos clínicos de los pacientes sometidos al estudio. **Género:** femenino (F), masculino (M). **Foco epileptico:** medial temporal (MT), temporo-parieto-occipital (TPO). **Tipo de crisis:** crisis focal consciente (FAS focal aware seizure), crisis focal con alteración de la conciencia (FIAS focal impaired awareness seizure), crisis foca a crisis tónico-clónica bilateral (FBTCS focal to bilateral tonic-clonic seizure). **Número de contactos:** Zona de inicio de crisis (SOZ seizure onset zone), Zona irritativa (EIZ epileptogenic irritative zone). **Número de RS:** Estados de reposo (RS resting-state). Modificado de (Cousyn et al., 2023).

### 2.1.1. Desbalance entre clases

Como el objetivo central es discriminar correctamente los estados preictales a partir de información funcional cerebral y dado que la ocurrencia de estos estados es menos frecuente que la de los estados interictales, es fundamental cuantificar el grado de desbalance en la distribución de clases, ya que esto condiciona tanto el análisis como la interpretación posterior de los resultados. Para caracterizar este desbalance, se utilizan las siguientes medidas:

- **Porcentaje de muestras preictales:**

$$\%Preictal = \frac{N_{preictal}}{N_{total}} \times 100$$

- **Índice de desbalance** o *Imbalance Ratio (IR)*:

$$IR = \frac{\max(N_{preictal}, N_{interictal})}{\min(N_{preictal}, N_{interictal})}$$

La Tabla 2.2 resume estas métricas aplicadas al conjunto de datos de este estudio, tanto de forma individual por paciente como a nivel global. Se observa que el grado de desbalance varía considerablemente entre pacientes: algunos presentan proporciones relativamente equilibradas (por ejemplo, P4 con 50% de muestras preictales), mientras que otros muestran una fuerte asimetría, como P7, donde sólo el 5.9% de las muestras corresponden a estados preictales. A nivel general, el conjunto completo contiene un total de 107 muestras, de las cuales 38 son preictales (35.5%) y 69 interictales (64.5%), lo que equivale a un *Imbalance Ratio* global de 1.82.

Si bien la proporción global es aproximadamente 2:1, se observa que el grado de desbalance varía ampliamente entre individuos, desde casos casi balanceados ( $IR \approx 1.0$ ) hasta otros con fuerte asimetría ( $IR > 5$ ), lo que refuerza la necesidad de analizar cada caso de forma independiente como estrategia que permite que el modelo presentado se adapte a las características específicas de cada sujeto, aislando las diferencias en la proporción de clases entre pacientes.

Aunque en algunos pacientes el número de muestras preictales es muy reducido, lo cual puede limitar la capacidad del modelo para generalizar, el uso de métricas apropiadas y robustas a este tipo de situaciones, como el *F1-score* y el *Balanced accuracy* (presentado en secciones posteriores), contribuyen a mantener la validez del análisis. Como alternativa frente a este desbalance se exploró inicialmente la estrategia de sobre-muestreo sintético, *SMOTE*, con el fin de equilibrar las clases. Sin embargo, esta técnica no mostró mejoras sustanciales en las métricas de desempeño y fueron descartadas en los análisis finales.

**Tabla 2.2***Desbalance de clases por paciente y a nivel global*

Paciente	Preictal	Interictal	Total	% Preictal	IR local
P0	2	6	8	25.0%	3.00
P1	10	1	11	90.9%	0.10
P2	3	8	11	27.3%	2.67
P3	2	10	12	16.7%	5.00
P4	2	2	4	50.0%	1.00
P5	2	4	6	33.3%	2.00
P6	11	3	14	78.6%	0.27
P7	1	16	17	5.9%	16.00
P8	3	9	12	25.0%	3.00
P9	2	10	12	16.7%	5.00
<b>Total</b>	<b>38</b>	<b>69</b>	<b>107</b>	<b>35.5%</b>	<b>1.82</b>

*Nota.* Distribución de muestras preictales e interictales por paciente, incluyendo el porcentaje de muestras preictales y el *Imbalance Ratio* (IR) local. En la fila final se presenta el resumen global del conjunto de datos.

## 2.2. Conectividad funcional en señales iEEG

Existen diversos métodos para la estimación y el análisis de la conectividad cerebral a partir de las señales eléctricas registradas mediante iEEG (Figura 1.3). Estas señales se representan como series de tiempo, cuya información relevante suele explorarse inicialmente a través del análisis espectral, identificando componentes oscilatorios característicos en distintas bandas de frecuencia. No obstante, múltiples estudios han demostrado que los potenciales de campo oscilatorios captados por electrodos implantados en diferentes regiones del cerebro están relacionados de manera significativa con la actividad neuronal subyacente (Eckhorn, 1994). Esta evidencia ha motivado el desarrollo de enfoques alternativos basados en medidas de acoplamiento de fase, capaces de reflejar interacciones funcionales entre áreas corticales. En este marco, el análisis de sincronización se ha consolidado como una herramienta poderosa para caracterizar la conectividad funcional a partir de la dinámica temporal de las señales cerebrales.

### 2.2.1. Método *Phase Locking Value*

Dentro de las distintas métricas desarrolladas para cuantificar la sincronización en registros iEEG, destaca una en particular por su simplicidad computacional y solidez interpretativa: el *Phase Locking Value* (PLV). La fase instantánea de la señal registrada por cada canal iEEG (electrodo), estimada a partir de la transformada de Hilbert, se representa como un punto sobre el círculo unitario en el plano complejo (Gabor, 1946). Al analizar dos señales, se calcula la diferencia de fase entre ellas en cada instante y se proyecta esta diferencia como un vector unitario (Chávez et al., 2003).

### 2.2.2. Método *Phase Locking Value*

Dentro de las distintas métricas desarrolladas para cuantificar la sincronización en registros iEEG, destaca una en particular por su simplicidad computacional y solidez interpretativa: el *Phase Locking Value* (PLV). La elección del PLV como métrica para estimar la conectividad funcional se fundamenta en varias consideraciones metodológicas y fisiológicas. En primer lugar, el PLV permite evaluar la consistencia de la relación de fase entre señales sin depender de la amplitud, lo que lo hace especialmente robusto frente a artefactos de potencia y variabilidad inter-electrodo, condiciones frecuentes en registros iEEG. Además, su formulación se basa en un promedio vectorial complejo que captura directamente la dispersión angular de las fases instantáneas, lo que lo convierte en una medida sensible a interacciones neuronales genuinas de tipo sincrónico. A nivel computacional, el PLV presenta una implementación eficiente y estable, adecuada para el análisis de grandes volúmenes de datos segmentados. Por otro lado, su validez neurofisiológica ha sido ampliamente respaldada en estudios previos, donde se ha demostrado que la sincronización de fase es un mecanismo clave en la comunicación interregional del cerebro, particularmente en contextos de reorganización funcional como los que preceden a eventos epilépticos (Lachaux et al., 1999; Fries, 2005). Así, el PLV proporciona una estimación cuantitativa del acoplamiento funcional que es tanto interpretable como comparativamente sólida dentro del conjunto de métricas disponibles para el análisis de conectividad en señales intracraneales.

La fase instantánea de la señal registrada por cada canal iEEG (electrodo), estimada a partir de la transformada de Hilbert, se representa como un punto sobre el círculo unitario en el plano complejo (Gabor, 1946). Al analizar dos señales, se calcula la diferencia de fase entre ellas en cada instante y se proyecta esta diferencia como un vector unitario (Chávez et al., 2003). Para una señal arbitraria  $x(t)$ , su correspondiente señal analítica  $\zeta_x(t)$  se define como una función compleja del tiempo:

$$\zeta_x(t) = x(t) + i\tilde{x}(t) = A_x(t)e^{i\theta_x(t)},$$

donde  $\tilde{x}(t)$  representa la transformada de Hilbert de  $x(t)$ , dada por:

$$\tilde{x}(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau,$$

siendo *p.v.* el valor principal de Cauchy, que indica que la integral debe interpretarse en ese sentido. A partir de esta formulación, se obtiene la amplitud instantánea  $A_x(t)$  y la fase instantánea  $\theta_x(t)$  de la señal  $x(t)$ .

Ahora, dos señales  $x(t)$  y  $y(t)$  se consideran sincronizadas en fase bajo una relación  $n : m$  si la diferencia  $\theta_{x,y}(t) = n\theta_x(t) - m\theta_y(t)$  permanece acotada en el tiempo para enteros  $n$  y  $m$ . Sin embargo, dado que esta relación no siempre es conocida de antemano, en la práctica se asume generalmente una sincronía 1:1 para fines de cuantificación. Así, el grado de acoplamiento de fase puede evaluarse utilizando el valor medio de la diferencia de fase compleja:

$$\gamma_{x,y} = \left\langle e^{i\theta_{x,y}(t)} \right\rangle,$$

donde el operador  $\langle \cdot \rangle$  indica el promedio temporal. En ausencia de sincronización, las diferencias de fase se distribuyen de forma uniforme sobre el círculo unitario y  $\gamma_{x,y} \approx 0$ . En contraste, si las señales están perfectamente acopladas en fase, la diferencia se mantiene constante y  $\gamma_{x,y} \approx 1$  (Chávez et al., 2003).

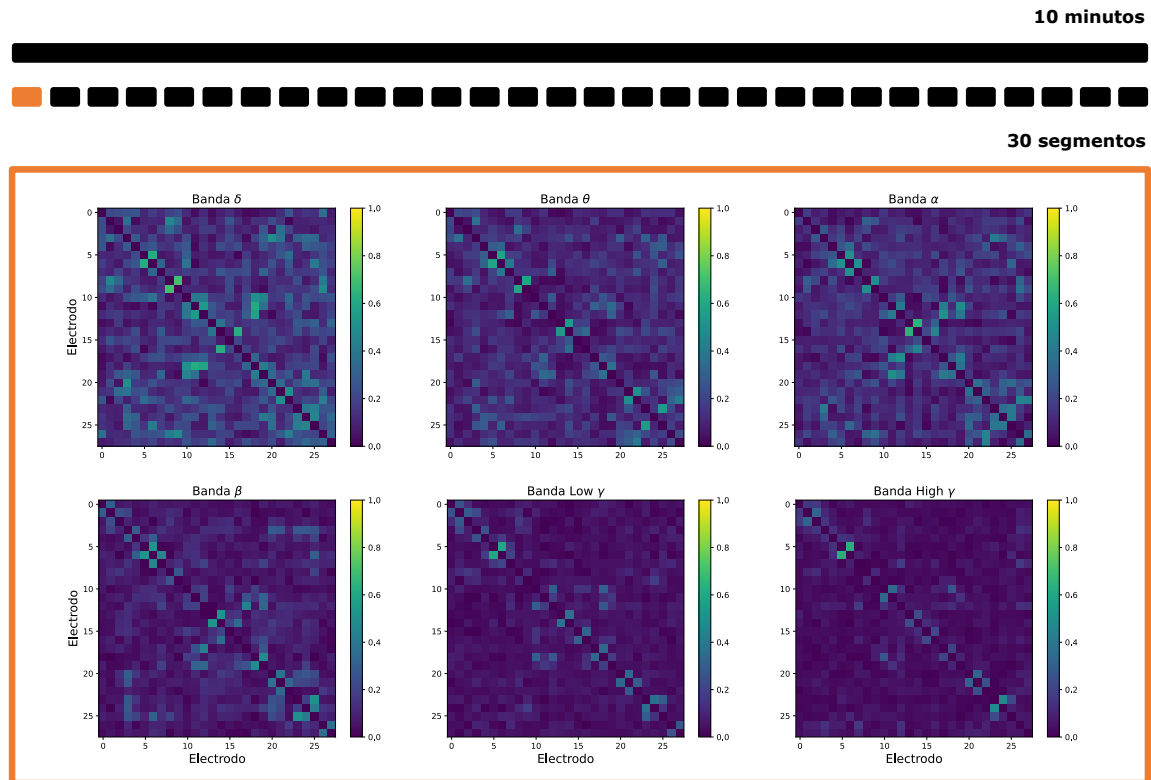
El PLV entonces es capaz de capturar la estabilidad relativa de la fase entre señales sin depender de su amplitud ni de su distribución espectral (Aydore et al., 2013). Pues, si dos electrodos implantados en regiones corticales distintas registran señales con un PLV alto en una banda de frecuencia específica, esto resalta la existencia de una interacción funcional sostenida entre dichas regiones dentro de ese régimen dinámico. Al extender este cálculo a todos los pares de electrodos de interés (típicamente la totalidad de electrodos SOZ y EIZ, pero pueden ser un subconjunto de estos según criterio clínico), se obtiene una matriz de conectividad funcional, en la que cada entrada representa el grado de sincronización de fase entre dos regiones. Esta matriz constituye un mapeo de las interacciones cerebrales durante el intervalo analizado y permite identificar patrones de sincronización que pueden asociarse con distintos estados fisiológicos o patológicos.

Con el fin de maximizar la resolución temporal de los 10 minutos de registro iEEG sin comprometer la estabilidad de las estimaciones de conectividad, cada uno de estos registros diarios de 10 minutos fue dividido en 30 ventanas consecutivas (segmentos) no solapadas de 20 segundos. Esta partición permite capturar la variabilidad dinámica de los patrones de sincronización cerebral en escalas de tiempo más finas, lo cual es especialmente relevante en estudios de estados cerebrales, ya que los cambios funcionales pueden ser graduales y transitorios. Además, trabajar con ventanas más cortas mejora

la estabilidad numérica del cálculo del PLV, al reducir la probabilidad de que una sola oscilación dominante sesgue la medida global de sincronización en ventanas demasiado largas.

La [Figura 2.1](#) ilustra el conjunto inicial de datos funcionales empleados en este estudio. Para cada segmento (por día) de un paciente, se calculan seis matrices de conectividad correspondientes a las bandas de frecuencia ( $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$ , Low  $\gamma$ , High  $\gamma$ ), cada una vinculada a diferentes procesos neurofisiológicos. De esta manera, se obtiene un seguimiento temporal de la evolución de la actividad iEEG mediante un total de 30 segmentos por día de registro, por paciente. Todas las matrices generadas, en todas las bandas y segmentos, son clasificadas según su estado funcional como interictales o preictales, siguiendo la categorización descrita en la [Tabla 2.1](#).

**Figura 2.1.**  
*Matrices de conectividad funcional*



*Nota.* División de una toma diaria de iEEG de 10 minutos en 30 segmentos consecutivos, para el paciente 0 con 28 electrodos de interés. Se ejemplifica el primer segmento (en color naranja), para el cual se construyen seis matrices de conectividad funcional simétrica  $m \times m$  ( $m$  = número de electrodos de interes) correspondientes a las bandas de frecuencia  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$ , Low  $\gamma$  y High  $\gamma$ , utilizando el método PLV. Cada uno de los segmentos restantes (en negro) cuenta con su propio conjunto de matrices de conectividad.

## CAPÍTULO 3

# Construcción, tratamiento y proyección de las redes de conectividad cerebral

En este capítulo se aborda el procesamiento de las matrices de conectividad funcional obtenidas previamente, con el objetivo de transformarlas en representaciones adecuadas para su análisis geométrico. El tratamiento de estas matrices sigue una secuencia metodológica que comienza con su interpretación como redes, donde los nodos corresponden a electrodos de iEEG y los enlaces reflejan relaciones funcionales cuantificadas por medidas de acoplamiento de fase. Estas redes son filtradas y posteriormente son proyectadas en espacios de menor dimensión mediante técnicas de reducción espectral, con el fin de preservar su estructura interna y facilitar la exploración de patrones dinámicos.

### 3.1. Construcción de las redes de conectividad

Como ya fue mencionado, las matrices de conectividad funcional cuantifican el grado de sincronización entre todos los pares de electrodos implantados en un paciente, y se calculan de forma independiente para cada banda de frecuencia. Una estrategia eficaz para analizar esta información consiste en representarla mediante grafos. En este marco, los electrodos se interpretan como nodos, y los valores de sincronización se traducen en enlaces ponderados que conectan dichos nodos.

Cada red cerebral funcional queda entonces definida como  $G = (\Omega, W)$ , donde  $\Omega$  es el conjunto de nodos/electrodos específicos para cada paciente y  $W = \{w_{ij}\}_{i,j \in \Omega}$  es la matriz simétrica de pesos de conectividad con elementos  $w_{ij} \in [0, 1]$ , que representan el valor de PLV entre los electrodos  $i$  y  $j$ . Esta construcción permite representar la dinámica funcional del cerebro en forma de red, sin imponer restricciones previas sobre su topología. Esta representación mediante grafos facilita la visualización de la arquitectura funcional cerebral, [Figura 3.1\(a\)](#), y proporciona una base estructurada para el análisis posterior de sus propiedades dinámicas y topológicas.

## 3.2. Filtrado de las redes de conectividad

En esta etapa del procesamiento, las redes obtenidas corresponden a grafos densos y se vuelve necesario aplicar filtrados de redes que conservan un porcentaje significativo de los enlaces más fuertes para facilitar su interpretación. Aunque esta estrategia conlleva una inevitable pérdida de información, el filtrado atenúa la incertidumbre asociada a los enlaces más débiles, reduce la presencia de falsos positivos y clarifica la topología funcional inferida a partir de los datos (Vidal et al., 2011; De Vico Fallani et al., 2014).

En este contexto, la estrategia de doble filtrado aplicada aquí, compuesta por un árbol recubridor mínimo (*MST*) seguido de una expansión controlada del grado medio, ofrece una alternativa más fundamentada frente al uso convencional de umbrales de conectividad fija. Este último enfoque, si bien ampliamente utilizado, presenta limitaciones importantes: la elección del umbral suele carecer de una base teórica clara y depende fuertemente de la distribución específica de los valores de conectividad en cada conjunto de datos, lo que introduce sesgos inter-paciente o inter-banda y afecta la comparabilidad entre redes. Además, el thresholding puede inducir fragmentación en la red o eliminar enlaces funcionalmente relevantes si estos no superan el umbral arbitrario definido. Por el contrario, la construcción de un *MST* garantiza conectividad global y mínima redundancia y ha demostrado ser especialmente útil en neurociencia para obtener estructuras funcionales estables y robustas frente a variaciones locales en la conectividad, con aplicaciones ampliamente aceptadas en estudios de iEEG y MEG (Stam et al., 2014); mientras que la posterior expansión controlada permite enriquecer la topología de manera uniforme, manteniendo un balance entre simplicidad estructural y riqueza informativa. Esta estrategia combinada resulta particularmente adecuada en contextos clínicos donde se requiere robustez analítica sin comprometer la interpretabilidad de las redes obtenidas.

### 3.2.1. Filtrado por Árbol Recubridor Mínimo

Primero, se construye el árbol recubridor mínimo (*Minimum Spanning Tree*, *MST*), el cual es un subgrafo del grafo original  $G = (\Omega, W)$  que cumple con tres condiciones fundamentales: i) conserva todos los nodos del grafo original, ii) no contiene ciclos, y iii) posee exactamente  $n - 1$  enlaces, donde  $n$  es el número total de nodos. Para su construcción se emplea el algoritmo de Kruskal, que opera sobre una matriz de costos definida como  $C = 1 - W$ , en la cual cada entrada  $c_{ij} = 1 - w_{ij}$  representa el costo asociado al enlace entre los nodos  $i$  y  $j$  (Kruskal, 1956). Esta transformación garantiza que las conexiones con mayor sincronización (es decir, mayores valores de  $w_{ij}$  se traduzcan en menores costos, favoreciendo su inclusión en el árbol resultante.

El algoritmo ordena todos los enlaces posibles  $E$  según  $c_{ij}$ , de menor a mayor costo, y selecciona secuencialmente aquellos que minimizan el costo total sin formar ciclos.

Matemáticamente, el árbol resultante se expresa como:

$$T_{\text{MST}} = \arg \min_{T \subseteq E} \sum_{(i,j) \in T} c_{ij}, \quad \text{tal que } T \text{ es conexo y acíclico, } |T| = n - 1$$

Este procedimiento garantiza que la red resultante retenga las conexiones funcionales más relevantes desde el punto de vista de sincronización, al tiempo que reduce la complejidad estructural del grafo completo. Además de mitigar el impacto del ruido en las señales, esta representación facilita el análisis e interpretación de los patrones dominantes de interacción neuronal en el cerebro.

### 3.2.2. Filtrado de Grado Medio $\langle k \rangle$

El MST reduce el ruido y conserva la conectividad más fuerte del grafo; sin embargo, con solo  $n - 1$  enlaces para  $n$  nodos produce una red extremadamente simple que puede omitir interacciones funcionales importantes. Por lo tanto, es fundamental realizar una expansión controlada de enlaces con el objetivo de enriquecer la estructura sin comprometer su dispersión. Esta expansión se fundamenta en mantener el grado medio de la red en un valor predefinido.

El grado ( $k_i$ ) de un nodo  $i$  en un grafo  $G$  se define como el número de enlaces que inciden sobre dicho nodo. Esta propiedad permite caracterizar topológicamente la red a través de su distribución de grados  $P(k)$ , que representa la probabilidad de que un nodo elegido al azar tenga  $k$  conexiones. Una medida global derivada de esta distribución es el grado medio  $\langle k \rangle$ , correspondiente al primer momento de  $P(k)$  (Boccaletti et al., 2006), y calculado como:

$$\langle k \rangle = \sum_k k P(k).$$

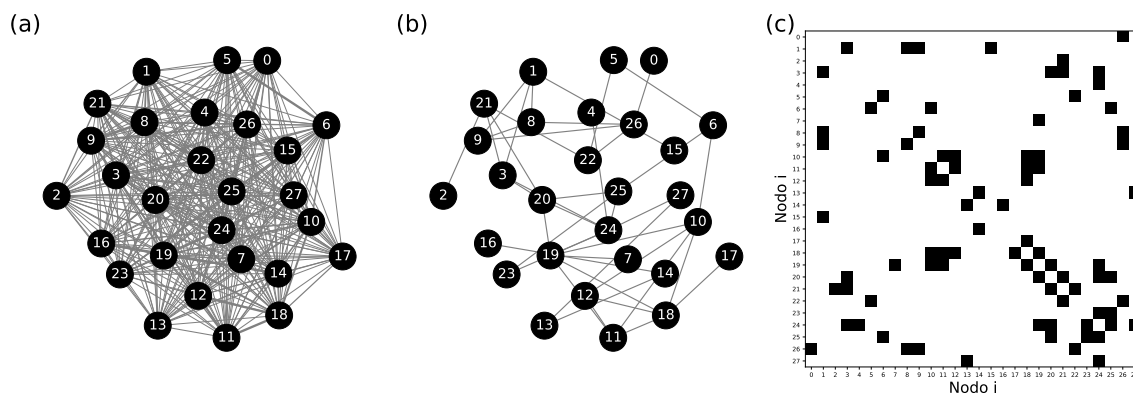
Este valor indica, en promedio, cuántas conexiones tiene un nodo en la red y proporciona una estimación de su densidad estructural: un grado medio bajo indica una red dispersa con pocas conexiones por nodo, mientras que un grado medio alto implica una red densa. Estudios sobre diferentes redes cerebrales, específicamente en redes cerebrales construidas con iEEG, determinan que el grado medio óptimo para lograr un equilibrio entre fidelidad estructural, claridad analítica y preservación de los vínculos más informativos es de  $\langle k \rangle = 3$  (De Vico Fallani et al., 2017).

Expandimos entonces, de forma controlada, el árbol obtenido mediante el MST con la adición progresiva de enlaces seleccionados entre aquellos con mayor valor de

sincronización que no fueron incluidos inicialmente, con el objetivo de alcanzar un grado medio de 3 conexiones por nodo, reforzando la estructura de la red filtrada sin comprometer su carácter disperso ni introducir ruido estructural excesivo. A esta red resultante, que contiene las componentes más significativas de las interacciones funcionales, se le asigna un peso unitario a todos sus enlaces. La visualización del proceso de filtrado puede observarse en la [Figura 3.1\(b\)](#). Como resultado del filtrado obtenemos la red  $G(\Omega, W)$  con  $W \in \{0, 1\}^{n \times n}$ , donde  $w_{ij} = 1$  indica la existencia de un enlace entre los nodos  $i$  y  $j$ , y  $w_{ij} = 0$  su ausencia, como se muestra en la [Figura 3.1\(c\)](#).

**Figura 3.1.**

*Visualización del proceso de filtrado de las redes de conectividad*



*Nota.* Redes de conectividad para el paciente 0 con 28 electrodos/nodos. (a) Grafo funcional completo derivado de la matriz de conectividad. (b) Grafo filtrado mediante MST y expansión controlada manteniendo  $\langle k \rangle = 3$ . (c) Matriz de conectividad binaria asociada al grafo filtrado.

### 3.3. Proyección de las redes de conectividad en espacio Euclídeo

Si bien las redes de conectividad filtradas ya capturan las interacciones más fuertes de cada uno de los estados y pueden ser descritas individualmente a través de las métricas de la teoría de grafos, su comparación directa plantea dos dificultades fundamentales. En primer lugar, estas redes corresponden a estructuras de alta dimensionalidad  $O(n^2)$ , lo que incrementa significativamente el costo computacional. En segundo lugar, comparar redes en su forma original requiere definir métricas sobre el espacio de grafos, lo cual conlleva ambigüedades conceptuales y limita la interpretabilidad de los resultados.

Ante estas limitaciones, resulta necesario reducir la dimensionalidad de las redes

de conectividad preservando al mismo tiempo su estructura funcional esencial. En este trabajo, se proyectan las redes en un espacio euclídeo, aprovechando la presencia de una métrica explícita que permite realizar análisis cuantitativos clásicos y facilitar la interpretación geométrica de las interacciones funcionales. Estudios previos han demostrado la utilidad de este enfoque en redes funcionales derivadas de señales magnetoencefalográficas (MEG), permitiendo discriminar entre sujetos sanos y pacientes con epilepsia (Chavez et al., 2010), así como identificar desplazamientos en las proyecciones asociados a la transición hacia estados preictales en registros iEEG (Duncan et al., 2013b).

### 3.3.1. Método de reducción de dimensionalidad por Mapas de Difusión

Para lograr una reducción de dimensionalidad que preserve la geometría intrínseca de los datos, se adoptó el algoritmo de mapas de difusión (*Diffusion Maps*), un método no lineal capaz de capturar de manera efectiva la estructura de variedades subyacentes en datos de alta dimensión (Coifman and Lafon, 2006; Lafon and Lee, 2006). *Diffusion Maps* construye un espacio en  $\mathbb{R}^q$  donde la proximidad entre los puntos/nodos proyectados está gobernada no solo por las conexiones directas entre dos nodos, sino también por sus trayectorias de difusión. Estas trayectorias representan la probabilidad acumulada de transitar de un nodo a otro a través de caminos intermedios; es decir, la proyección de dos nodos en el espacio creado resulta cercana si estos se encuentran conectados por varios caminos en la red.

El *Diffusion maps* requiere de dos condiciones fundamentales que la matriz de conectividad  $W$  de la red filtrada  $G(\Omega, W)$  posee: i) simetría  $W = W^T$ , lo cual asegura que la conectividad entre nodos no depende del orden de observación; y ii) positividad punto a punto:  $w(i, j) \geq 0$  para todo  $i, j \in \Omega$ . El método estima la matriz de probabilidad de transición  $P$  de una cadena de Markov con entradas

$$p_{ij} = \frac{w_{ij}}{d_i}$$

donde el grado de cada nodo  $d_i = \sum_{k \in \Omega} w_{ik}$ . Cada entrada  $p_{ij}$  codifica la probabilidad de transitar del nodo  $i$  al nodo  $j$  mediante un *random walk* de longitud 1. Este *random walk* induce una difusión geométrica con una distancia asociada entre los nodos  $i$  y  $j$  que sigue la forma de distancia euclidiana clásica  $L^2$  y se define como (Lafon and Lee, 2006):

$$d_{ij}^2 = \sum_{k \geq 0} \frac{(p_{ik} - p_{jk})^2}{\mu_k^*},$$

donde  $\mu_k^* = \frac{d_k}{\sum_{l \in \Omega} d_l}$  denota la distribución estacionaria única de la cadena de Markov  $P$ . Por construcción, esta distancia de difusión está fuertemente influenciada por la conectividad de la red: toma valores pequeños si los nodos están conectados a través de múltiples caminos. Al considerar la representación espectral de  $P$ , se obtienen autovalores  $|\lambda_0| \geq |\lambda_1| \geq \dots \geq |\lambda_{n-1}|$  y autovectores  $\phi_k$  y  $\psi_k$  tales que

$$\phi_k^T P = \lambda_k \phi_k^T \quad \text{y} \quad P \psi_k = \lambda_k \psi_k,$$

donde  $\lambda_0 = 1$  es el mayor autovalor de  $P$ , asociado al estado estacionario del proceso de difusión, y  $\psi_0 \equiv 1$  representa el autovector derecho constante, correspondiente a una caminata aleatoria en equilibrio. Esta descomposición espectral permite interpretar  $P$  como un operador compacto cuya dinámica está determinada por la tasa de decaimiento de los autovalores  $\lambda_k$ , y cuyos autovectores forman la base del espacio proyectado.

Entonces, la distancia de difusión se puede reescribir en términos de sus componentes espectrales. Para un tiempo de difusión  $t = 1$ , dicha distancia entre dos nodos  $i$  y  $j$  se expresa como:

$$d_{ij}^2 = \sum_{k \geq 1} \lambda_k^2 (\psi_k(i) - \psi_k(j))^2,$$

donde  $\psi_k(i)$  representa la componente  $i$ -ésima del  $k$ -ésimo autovector derecho de la matriz de transición  $P$ . El término  $k = 0$  no contribuye a la distancia, ya que  $\psi_0 \equiv 1$  es constante en todos los nodos. Esta distancia puede aproximarse de forma eficiente reteniendo únicamente los primeros  $q$  autovectores no triviales, lo que permite reducir la dimensionalidad preservando la estructura difusiva global:

$$d_{ij}^2 \simeq \sum_{k=1}^q \lambda_k^2 (\psi_k(i) - \psi_k(j))^2,$$

siendo esta la interpretación directa de la distancia euclídea en  $\mathbb{R}^q$ . A partir de esta representación, se construye el mapa de difusión como:

$$\Psi : x \mapsto \begin{pmatrix} \lambda_1 \psi_1(x) \\ \lambda_2 \psi_2(x) \\ \vdots \\ \lambda_q \psi_q(x) \end{pmatrix}.$$

Este mapeo  $\Psi : \Omega \rightarrow \mathbb{R}^q$  proyecta el conjunto de nodos  $\Omega$  como una nube de puntos en

un espacio euclídeo de dimensión reducida, donde cada coordenada está determinada por los autovectores derechos de la matriz de transición  $P$ , ponderados por sus respectivos autovalores. En esta nueva representación, las distancias euclídeas entre puntos reflejan con fidelidad la conectividad funcional de largo alcance presente en la red original (Lafon and Lee, 2006).

En el caso del paciente 5 (Tabla 2.1), cuya red funcional cuenta con  $n = 11$  nodos, la matriz de transición  $P \in \mathbb{R}^{11 \times 11}$  genera, por segmento, un conjunto de 11 autovalores  $\lambda_k$  y 11 autovectores derechos  $\psi_k$ , cada uno de dimensión  $11 \times 1$ . La Figura 3.2(a) ilustra el comportamiento promediado de los valores absolutos de los autovalores, ordenados de forma descendente, a lo largo de todos los días y tanto en estado interictal como preictal. Este patrón fue consistente en todos los pacientes tras aplicar *Diffusion Maps* con paso de transición  $t = 1$ .

El análisis del espectro de autovalores no reveló la presencia de un salto espectral claro que sugiriera, de manera natural, una dimensión óptima para la proyección. No obstante, los dos primeros autovalores no triviales presentaron magnitudes significativamente superiores en comparación con los siguientes, lo que indica que capturan una proporción sustancial de la estructura global de la red. Por esta razón, se adoptó de forma consistente la dimensión  $q = 2$  a lo largo de todo el estudio. La Figura 3.2(b) muestra el mapa de puntos proyectados  $\Psi(i) = (\lambda_1 \psi_1(i), \lambda_2 \psi_2(i)) \in \mathbb{R}^2$  para tres segmentos consecutivos del estado interictal para el paciente 5.

Si bien no se realizó una evaluación sistemática del efecto de diferentes valores de  $q$ , se estimó la fracción acumulada de información espectral representada por los dos primeros autovalores, promediada a lo largo de todos los pacientes y bandas de frecuencia:

$$\frac{1}{60} \sum_{p=1}^{10} \sum_{b=1}^6 \left( \frac{\sum_{k=1}^2 |\lambda_k^{(p,b)}|}{\sum_{k=1}^{n-1} |\lambda_k^{(p,b)}|} \right) \approx 0.5,$$

lo que indica que, en promedio, aproximadamente el 50% de la energía espectral no trivial es capturada en la proyección bidimensional. Cabe señalar que esta elección implica una pérdida considerable de información estructural, por lo que podría considerarse el uso de un umbral de captura más elevado (por ejemplo, entre 70% y 80%) con el fin de preservar un mayor contenido funcional. Esta decisión conllevaría el uso de un espacio euclidiano de mayor dimensión, lo cual introduce desafíos adicionales tanto en la visualización como en la interpretación geométrica de los resultados. Como dirección metodológica para trabajos futuros, se propone evaluar de forma sistemática la estabilidad del *embedding* y su capacidad discriminativa ante diferentes restricciones dimensionales, con el objetivo de determinar si la inclusión de componentes espectrales adicionales aporta significativamente al análisis comparativo entre estados interictal y preictal.

### 3.4. Alineación de las redes proyectadas en $\mathbb{R}^2$

La descomposición espectral utilizada en el método de *Diffusion Maps* puede dar lugar a inversiones de signo en los autovectores o rotaciones del espacio proyectado. Estas variaciones se deben a propiedades algebraicas inherentes: la dirección de un autovector no es única y pequeñas perturbaciones pueden rotar el *embedding*. Aunque estas transformaciones no alteran la estructura interna del espacio proyectado, redes de conectividad similares pueden generar coordenadas de proyección sustancialmente distintas, [Figura 3.2\(b\)](#), dificultando así la comparación directa entre estados e incluso entre segmentos de un mismo día ([Richiardi et al., 2013](#); [Gürsoy et al., 2023](#)). Este efecto puede modificar artificialmente las distancias entre nodos correspondientes, sin reflejar cambios reales en la conectividad funcional. Para evitarlo, se aplicó un procedimiento de alineación de nodos sobre las redes proyectadas en  $\mathbb{R}^2$ , con el fin de garantizar la comparabilidad entre muestras.

#### 3.4.1. Método de Análisis de Procrustes Generalizado

Se implementó el Análisis de Procrustes Generalizado (GPA, por sus siglas en inglés), un método estadístico multivariado diseñado para comparar y alinear múltiples configuraciones de puntos que viven en un espacio euclídeo común ([Dijksterhuis and Gower, 1991](#)). Su objetivo es eliminar las variaciones debidas a transformaciones geométricas como traslaciones, rotaciones, reflexiones y escalamiento, para permitir la comparación directa entre configuraciones preservando su estructura relativa. ([Gower, 1975](#)).

Supóngase un conjunto de  $K$  matrices  $\{\mathbf{X}^{(k)}\}_{k=1}^K$ , donde cada  $\mathbf{X}^{(k)} \in \mathbb{R}^{n \times q}$  representa una configuración de  $n$  puntos en un espacio euclídeo de dimensión  $q$ . En nuestro caso, estas configuraciones corresponden a las redes funcionales proyectadas  $\Psi^{(k)}(i) : \mathbf{x}_i^{(k)} = \left( \lambda_1^{(k)} \psi_1^{(k)}(i), \lambda_2^{(k)} \psi_2^{(k)}(i) \right) \in \mathbb{R}^2$ , de modo que cada fila de la matriz  $\mathbf{X}^{(k)}$  contiene las coordenadas proyectadas del nodo  $i$  en el segmento  $k$ . El objetivo del GPA es encontrar, para cada matriz  $\mathbf{X}^{(k)}$ , una transformación de la forma:

$$\mathbf{X}^{(k)} \mapsto \mathbf{Y}^{(k)} = b^{(k)} \mathbf{X}^{(k)} \mathbf{T}^{(k)} + \mathbf{C}^{(k)},$$

donde  $b^{(k)}$  es un factor de escalamiento,  $\mathbf{T}^{(k)} \in \mathbb{R}^{2 \times 2}$  es una matriz ortogonal que representa rotaciones y/o reflexiones, con  $\mathbf{T}^{(k)'} \mathbf{T}^{(k)} = \mathbf{I}_2$ , y  $\mathbf{C}^{(k)} \in \mathbb{R}^{1 \times 2}$  es un vector de traslación. El procedimiento se realiza en los siguientes pasos:

1. **Centrado:** Cada configuración proyectada  $\mathbf{X}^{(k)} \in \mathbb{R}^{n \times 2}$ , se traslada a su centroide.

Se calcula el baricentro  $\bar{\mathbf{x}}^{(k)} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^{(k)}$ , y se resta a cada punto:

$$\tilde{\mathbf{X}}^{(k)} = \mathbf{X}^{(k)} - \mathbf{1}\bar{\mathbf{x}}^{(k)}.$$

2. **Selección de referencia:** Se selecciona una de las configuraciones centradas como referencia  $\mathbf{M}^{(0)}$  para la alineación de las demás matrices. En nuestro caso, la elección de la matriz de referencia fue aleatoria dentro del conjunto de segmentos interictales de cada paciente.
3. **Alineamiento:** Para cada configuración  $\tilde{\mathbf{X}}^{(k)}$ , se determina la matriz ortogonal  $\mathbf{T}^{(k)} \in \mathbb{R}^{2 \times 2}$  que minimiza la distancia de Frobenius respecto a la matriz de referencia  $\mathbf{M}^{(0)}$ . Esta operación equivale a resolver el siguiente problema de optimización:

$$\text{mín tr} \left[ \left( \tilde{\mathbf{X}}^{(k)} \mathbf{T}^{(k)} - \mathbf{M}^{(0)} \right)' \left( \tilde{\mathbf{X}}^{(k)} \mathbf{T}^{(k)} - \mathbf{M}^{(0)} \right) \right],$$

Este problema se resuelve mediante la descomposición en valores singulares (*Singular Value Decomposition*, SVD) de la matriz producto:

$$\mathbf{M}^{(0)'} \tilde{\mathbf{X}}^{(k)} = \mathbf{U} \mathbf{S} \mathbf{V}',$$

y la solución óptima se obtiene como:

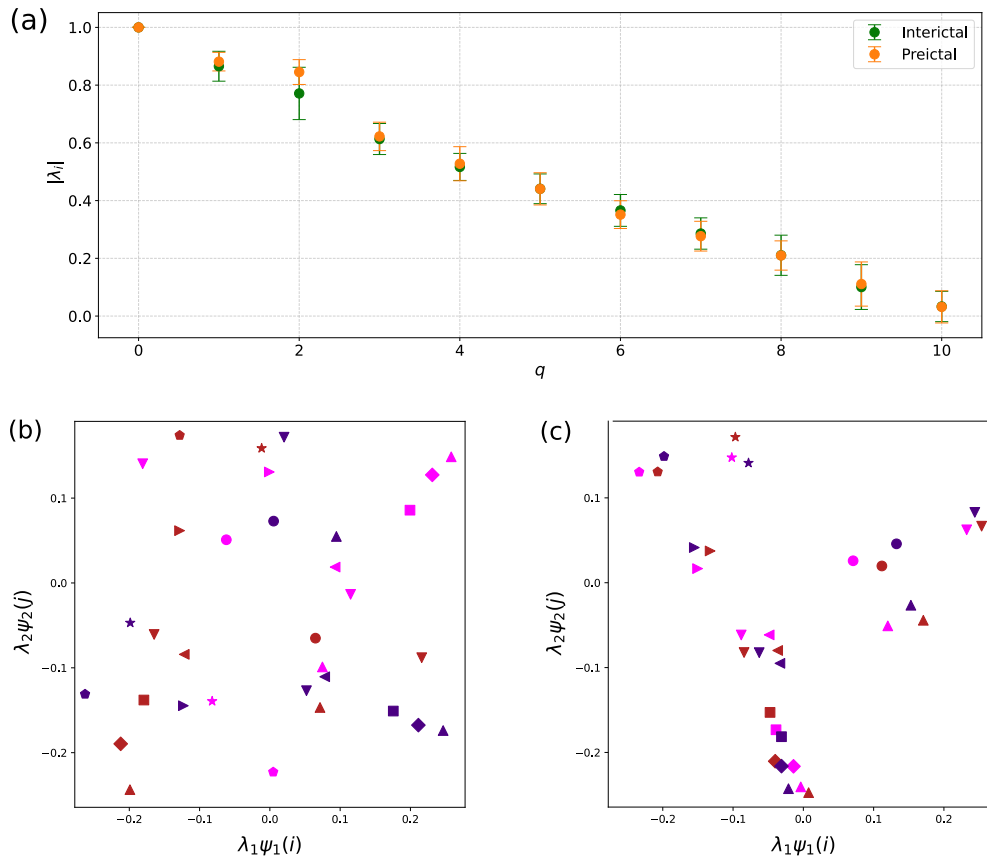
$$\mathbf{T}^{(k)} = \mathbf{V} \mathbf{U}'.$$

4. **Escalamiento:** Se calcula el escalar óptimo por mínimos cuadrados como:

$$b^{(k)} = \frac{\text{tr}(\mathbf{M}' \mathbf{X}^{(k)} \mathbf{T}^{(k)})}{\text{tr}(\mathbf{X}^{(k)'} \mathbf{X}^{(k)})},$$

sin embargo, en este trabajo se fijó  $b^{(k)} = 1$  para evitar distorsiones de escala.

El resultado final de la transformación es un conjunto de configuraciones  $\{\mathbf{Y}^{(k)}\}$  alineadas en un mismo sistema de coordenadas, conservando la geometría relativa original de cada una y permitiendo comparaciones directas nodo a nodo. Esta alineación se ilustra en la [Figura 3.2\(c\)](#) para tres segmentos consecutivos del estado interictal del paciente 5. Dado que el presente trabajo busca realizar comparaciones locales entre estados fisiológicos, es relevante destacar que, tras el alineamiento, los nodos correspondientes (por ejemplo, el nodo 1 en cada segmento) tienden a ubicarse en regiones próximas del espacio proyectado, lo que facilita la interpretación estructural y la identificación de variaciones funcionales específicas por nodo.

**Figura 3.2.***Proyección y alineación de las redes de conectividad funcional en  $\mathbb{R}^2$* 

*Nota.* Muestra: Paciente 5,  $n = 11$  nodos. (a) Organización de los autovalores de la descomposición espectral, promediados a lo largo de todos los días interictales y preictales del paciente. (b) Proyección espectral en el plano  $\mathbb{R}^2$  de tres segmentos consecutivos del estado interictal ( $k = 1, 2, 3$ ), codificados por color (rojo, azul y Rosa). Cada nodo  $i$  se representa mediante una figura geométrica distinta. (c) Proyección espectral tras la aplicación del alineamiento por Procrustes generalizado.

## CAPÍTULO 4

# Discriminación estadística entre estados Interictales y Preictales

Este capítulo describe el procedimiento para cuantificar las diferencias entre las distribuciones espaciales interictales y preictales por nodo. Posteriormente, se detalla el proceso de identificación y selección de los nodos con mayor capacidad discriminativa entre los estados. Finalmente, se introduce un biomarcador adimensional que, a partir de esta información, permite asignar una probabilidad de pertenencia de una proyección desconocida a alguno de los dos estados, permitiendo así su clasificación.

Dado que estudios previos han sugerido que la distribución espacial de los nodos proyectados en el espacio euclídeo varía significativamente antes de la aparición de una crisis epiléptica (Duncan et al., 2013a), en nuestras proyecciones en  $\mathbb{R}^2$  esta diferencia resulta evidente. Específicamente, se observa que las proyecciones de un nodo  $i$  a lo largo de varios segmentos consecutivos de un mismo estado (por ejemplo, el interictal) tienden a agruparse en una región espacial coherente [Figura 3.2](#). No obstante, debido a la evolución temporal natural del registro iEEG, al aumentar el número de segmentos analizados se hace perceptible un desplazamiento progresivo en las coordenadas proyectadas de cada nodo. En este trabajo, interpretamos dicho desplazamiento como una trayectoria que describe una distribución nodal, construida a partir de  $30 \times d$  proyecciones, correspondientes a los 30 segmentos diarios a lo largo de  $d$  días de registro por paciente, la cual encapsula la variabilidad local de la actividad funcional en cada nodo  $i$ . De forma consistente, se observó que estas distribuciones presentan diferencias entre los estados interictal y preictal, lo que sugiere su utilidad para discriminar entre ambos.

### 4.1. Discriminación de distribuciones locales

La [Figura 4.1](#) muestra el alineamiento de las proyecciones en  $\mathbb{R}^2$  para el paciente 5, considerando todos los días en estados interictales y preictales. A nivel global, las configuraciones proyectadas tienden a concentrarse en tres regiones principales del plano, lo cual sugiere la existencia de hubs funcionales. Aunque las proyecciones de ambos estados parecen superponerse en su conjunto, lo que sugeriría una respuesta similar, el

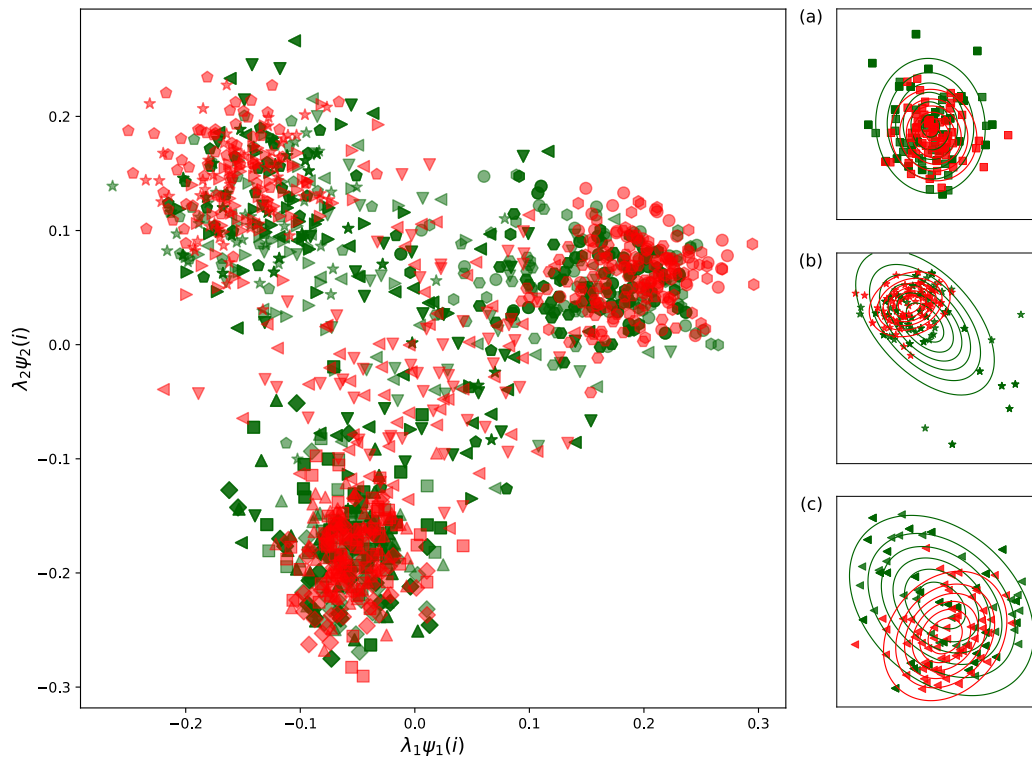
análisis local por nodo revela patrones diferenciables. Cabe señalar que esta geometría no es uniforme entre pacientes, ya que depende de factores como la localización anatómica y cantidad de electrodos implantados, las diferencias individuales en la organización funcional cerebral y la variabilidad clínica propia de cada caso. Sin embargo, las diferencias locales de respuesta entre los dos estados sí se mantienen. A lo largo de los pacientes, se observaron tres tipos generales de organización local:

- **Caso 1** [Figura 4.1\(a\)](#): los nodos  $i$  de todos los segmentos muestran una organización completamente superpuesta entre ambos estados, indicando una respuesta funcional indistinguible entre las fases interictal y preictal.
- **Caso 2** [Figura 4.1\(b\)](#): se observa separabilidad directa entre estados. Por ejemplo, la distribución interictal se mantiene confinada en una región específica del espacio, mientras que la preictal presenta una dispersión más amplia. También se encuentran casos donde este comportamiento se presenta de forma invertida.
- **Caso 3** [Figura 4.1\(c\)](#): ambos estados presentan alta dispersión espacial, pero las proyecciones se organizan de forma diferente a lo largo de los ejes principales del espacio embebido, lo que sugiere una diferencia estructural más sutil en la conectividad del nodo.

Estas características permiten establecer que, a partir de las distribuciones locales por nodo, es posible implementar métricas capaces de discriminar entre estados. Para facilitar esta tarea y aprovechar ventajas computacionales y teóricas, se decidió aproximar estas distribuciones mediante modelos gaussianos bivariados. No obstante, se reconoce que futuras investigaciones podrían explorar con mayor profundidad otras familias de distribuciones que se ajusten de manera más fiel a la estructura intrínseca y variable de las proyecciones locales o modelos más flexibles, como las mezclas de gaussianas, que permiten capturar posibles comportamientos multimodales dentro de estas configuraciones.

Para cuantificar el grado de separación entre las distribuciones locales de un mismo nodo  $i$  en los estados interictal y preictal, se utilizó la distancia de Bhattacharyya  $D_b(i)$ , una métrica ampliamente empleada en análisis estadístico y teoría de clasificación por su capacidad para medir la similitud entre distribuciones de probabilidad ([Fukunaga, 2013](#)). En nuestro contexto, cada nodo genera una nube de puntos  $\mathbf{x}_i^s \in \mathbb{R}^2$  que representa su proyección a lo largo de todos los segmentos en un estado  $s \in \{\text{inter, pre}\}$ . Supondremos que estas distribuciones pueden aproximarse mediante modelos gaussianos bivariados, lo cual permite aplicar una expresión cerrada y eficiente para  $D_b(i)$ .

La distancia de Bhattacharyya se define como:

**Figura 4.1.***Casos de configuración nodal en el espacio proyectado*

*Nota.* Organización espacial global en  $\mathbb{R}^2$  para el paciente 5. Cada figura geométrica representa un nodo  $i$  distinto. Estados interictales en verde y preictales en rojo: (a) sobreposición completa entre estados; (b) aglomeración en un estado y dispersión en el otro; (c) diferencias en la orientación direccional de las distribuciones.

$$D_b(i) = \frac{1}{8} (\mu_i^{\text{pre}} - \mu_i^{\text{inter}})^T \left[ \frac{\Sigma_i^{\text{inter}} + \Sigma_i^{\text{pre}}}{2} \right]^{-1} (\mu_i^{\text{pre}} - \mu_i^{\text{inter}}) + \frac{1}{2} \ln \frac{\left| \frac{\Sigma_i^{\text{inter}} + \Sigma_i^{\text{pre}}}{2} \right|}{|\Sigma_i^{\text{inter}}|^{1/2} |\Sigma_i^{\text{pre}}|^{1/2}}$$

donde  $\mu_i^s$  y  $\Sigma_i^s$  representan, respectivamente, el vector de media y la matriz de covarianza de las proyecciones del nodo  $i$  en el estado  $s \in \{\text{inter}, \text{pre}\}$ . Implementar esta métrica trae varias ventajas como: La distancia es simétrica, es decir,  $D_b(i; \text{inter}, \text{pre}) = D_b(i; \text{pre}, \text{inter})$ . A diferencia de métricas centradas únicamente en la distancia entre medias,  $D_b(i)$  también incorpora diferencias en forma, tamaño y orientación de las distribuciones a través de sus matrices de covarianza. Al estar definida sobre distribuciones en  $\mathbb{R}^2$ , su interpretación es directa: un mayor valor de  $D_b(i)$  implica una mayor diferencia

entre la organización espacial del nodo  $i$  en los dos estados, facilitando su interpretación geométrica.

## 4.2. Identificación de nodos relevantes para la discriminación de estados

Si bien la métrica de Bhattacharyya permite cuantificar de forma efectiva la diferencia entre las distribuciones nodales asociadas a cada estado, enfrentamos una dificultad metodológica relevante. Como se ha mencionado previamente, el proceso de alineamiento es un paso necesario para poder comparar las proyecciones interictales y preictales dentro del espacio  $\mathbb{R}^2$ . No obstante, cada paciente cuenta con un conjunto amplio de redes interictales, específicamente  $30 \times d^{inter}$  matrices correspondientes a  $d$  días de registro interictal, lo cual implica que existen múltiples opciones válidas para seleccionar la red de referencia utilizada en dicho alineamiento.

Esta elección no es trivial: al variar la red interictal de referencia, se introducen cambios estructurales en las coordenadas del *embedding* alineado. Como consecuencia, se observan variaciones significativas en las mediciones de distancia de Bhattacharyya y, por ende, en la identificación de los nodos con mayor capacidad discriminativa entre estados. Este fenómeno revela una alta sensibilidad del procedimiento respecto a la elección de la red de referencia, lo que requiere de un proceso adicional para reducir el impacto en la selección de nodos relevantes. Se implementó entonces una estrategia basada en análisis de permutación (*shuffle test*). Este tipo de pruebas no paramétricas se fundamenta en la construcción de distribuciones nulas empíricas mediante la aleatorización de etiquetas de clase, y permiten evaluar si las separaciones observadas entre estados pueden atribuirse al azar. Este enfoque ha sido ampliamente adoptado en neurociencia computacional, especialmente en estudios con electroencefalografía (EEG), donde la variabilidad intra e intersujeto puede dificultar la aplicación de pruebas paramétricas tradicionales (Maris and Oostenveld, 2007).

Para cada paciente, se seleccionaron aleatoriamente 21 segmentos interictales que sirvieron como referencias independientes de alineamiento de Procrustes. A partir de cada uno de estos alineamientos, se calculó la correspondiente distancia de Bhattacharyya  $D_B(i)$  para cada nodo  $i$ , y estos valores fueron tomados como observaciones empíricas de referencia para su posterior comparación frente a los resultados obtenidos bajo permutaciones. A continuación, para cada una de las 21 alineaciones obtenidas, se construyeron 80 permutaciones independientes, siguiendo recomendaciones previas en análisis no paramétrico de EEG (Maris and Oostenveld, 2007), que sugieren usar al menos entre 50 y 100 permutaciones para construir distribuciones nulas empíricas robustas. En

cada permutación, los segmentos interictales y preictales del paciente fueron mezclados aleatoriamente, manteniendo la transformación de alineamiento previamente aplicada.

Este procedimiento simuló un escenario sin distinción real entre clases, generando un conjunto de distribuciones bajo la hipótesis nula. En cada una de las 80 permutaciones generadas por alineamiento, se recalculó la distancia de Bhattacharyya  $D_B^{\text{shuf}}(i)$  por nodo. Esto dio lugar a un conjunto de 80 valores de prueba (por nodo), los cuales conforman una distribución nula empírica bajo la hipótesis de no distinción entre estados. Para cada nodo  $i$ , se ajustó una distribución gaussiana a este conjunto de valores, estimando su media  $\mu_i$  y desviación estándar  $\sigma_i$ .

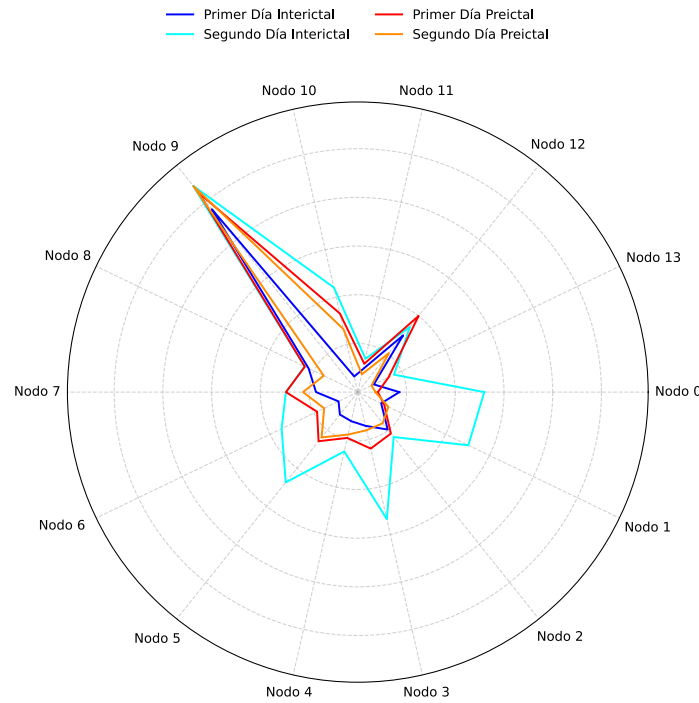
Una vez construida la distribución nula para cada nodo, se procedió a tipificar la distancia observada  $D_B(i)$  obtenida con las etiquetas verdaderas mediante el cálculo del *Z-score* :

$$Z(i) = \frac{D_B(i) - \mu_i}{\sigma_i},$$

este valor  $Z(i)$  representa el grado de desviación, en unidades de dispersión relativa estándar, de la distancia observada respecto a lo esperado bajo la hipótesis nula. Para este trabajo, valores elevados de  $Z(i)$  indican que las configuraciones espaciales de las proyecciones de ese nodo distan significativamente entre estados, y por tanto, ese nodo presenta un comportamiento diferenciador. En general, el carácter iterativo del método, basado en la combinación de múltiples alineamientos de referencia y permutaciones controladas, permitió evaluar de forma robusta la estabilidad del patrón de separación detectado. Se observó que los nodos con mayores valores de  $Z(i)$  emergen de forma consistente a través de las distintas iteraciones, como se ejemplifica en la [Figura 4.2](#). Esta emergencia de nodos representativos se mantiene desde la escala de segmentos hasta la escala de días, lo cual respalda la solidez estadística y la reproducibilidad de su capacidad discriminativa. Se observó que cada paciente presenta su propio conjunto de nodos representativos y que estos pueden variar a través de la respuesta en las diferentes bandas de frecuencia.

### 4.3. Biomarcador preictal $\mathcal{B}$

Paralelamente, en este trabajo definimos el parámetro  $\mathcal{B}$  como etapa final para la discriminación del estado preictal. Su objetivo es asignar una probabilidad relativa de pertenencia a uno u otro estado (interictal o preictal) con base en la morfología espacial de un nodo en el espacio proyectado. Para ello, se considera una red de conectividad asociada a un estado desconocido, la cual es proyectada y alineada en  $\mathbb{R}^2$ . El análisis se centra en el

**Figura 4.2.***Histograma polar de la prevalencia de nodos representativos*

*Nota.* Representación en gráfico polar del valor promedio del  $z$ -score de la distancia de Bhattacharyya  $D_B$  para los 14 nodos cerebrales del paciente 6. Se muestran los valores correspondientes a dos días consecutivos de estado interictal y dos días consecutivos de estado preictal.

nodo previamente identificado como el más discriminativo  $i^*$ , cuyo punto proyectado  $\mathbf{x}_{i^*}$  es evaluado frente a las distribuciones empíricas de referencia obtenidas para ese nodo en ambos estados.

La clasificación se realiza mediante la evaluación de las funciones de densidad de probabilidad (PDF) multivariadas, ajustadas durante el entrenamiento con base en la distribución de las proyecciones del nodo  $i^*$  en estados interictales y preictales. La expresión de la PDF bajo un estado  $s \in \{\text{inter, pre}\}$  para un punto  $\mathbf{x}_{i^*} \in \mathbb{R}^2$  es:

$$PDF_s(\mathbf{x}_{i^*}) = \frac{1}{(2\pi)^{|\Sigma_{i^*}^s|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_{i^*} - \mu_{i^*}^s)^T (\Sigma_{i^*}^s)^{-1} (\mathbf{x}_{i^*} - \mu_{i^*}^s)\right)}$$

Donde  $\mu_{i^*}^s$  y  $\Sigma_{i^*}^s$  corresponden, respectivamente, al vector de media y la matriz de covarianza que describen la distribución de las proyecciones del nodo  $i^*$  bajo el estado  $s$ .

A partir de esta formulación, definimos el biomarcador  $\mathcal{B}$  como una razón de verosimilitudes, adimensional y computacionalmente eficiente, que permite determinar si la morfología espacial de una nueva proyección se asemeja más al patrón nodal característico del estado preictal o interictal. Este enfoque permite decidir si la morfología de la red observada sigue la dinámica espacial típica de un estado preictal o interictal, utilizando como criterio una medida probabilística directa y fundamentada en la geometría nodal proyectada.

#### Definición del biomarcador adimensional $\mathcal{B}$

El biomarcador  $\mathcal{B}$  clasifica el estado de una red desconocida mediante la evaluación del nodo más discriminativo  $i^*$ , comparando la verosimilitud de su proyección  $\mathbf{x}_{i^*}$  bajo ambas distribuciones de referencia:

$$\mathcal{B}(\mathbf{x}_{i^*}) = \frac{PDF_{\text{inter}}(\mathbf{x}_{i^*})}{PDF_{\text{pre}}(\mathbf{x}_{i^*})}$$

La clasificación se define por:

$$s(\mathbf{x}_{i^*}) = \begin{cases} \text{preictal,} & \text{si } \mathcal{B}(\mathbf{x}_{i^*}) \leq 1 \\ \text{interictal,} & \text{si } \mathcal{B}(\mathbf{x}_{i^*}) > 1 \end{cases}$$

El cálculo del biomarcador  $\mathcal{B}$  se realiza de manera independiente para cada nodo discriminativo y por cada red proyectada correspondiente a un segmento temporal. Con el objetivo de mejorar la robustez del procedimiento, en este estudio se seleccionaron tres nodos representativos por paciente, definidos previamente a partir de los valores más altos de  $Z(i)$ . No obstante, este enfoque reveló una alta sensibilidad a la elección de la red de referencia utilizada en el alineamiento, de nuevo. En particular, se observó que múltiples segmentos presentaban valores promedio de  $\mathcal{B}$  cercanos al umbral de decisión  $\mathcal{B} = 1$ , generando ambigüedad en la clasificación binaria entre estados y dificultando una interpretación robusta a nivel individual.

Para estudiar de manera sistemática esta variabilidad, se construyó la distribución empírica de valores individuales del biomarcador  $\mathcal{B}$ . Este conjunto incluyó los resultados obtenidos para cada nodo representativo, en todos los segmentos, días, pacientes y bandas cerebrales consideradas. La [Figura 4.3](#) muestra la densidad de estos valores, destacando una concentración pronunciada en torno al valor crítico  $\mathcal{B} = 1$ , así como una cola extendida hacia valores grandes.

Aunque comúnmente se eliminan los valores atípicos para evitar sesgos estadísticos, en este trabajo se adoptó un criterio opuesto: se decidió conservar los valores extremos del biomarcador, pues estos reflejan configuraciones altamente discriminativas entre los estados. Para ello, se aplicó un análisis de rango intercuartílico (IQR) sobre el

conjunto completo de valores, expresados en escala logarítmica:

$$x_i = \log(\mathcal{B}_i), \quad \text{para cada nodo representativo y cada segmento.}$$

Se calcularon los percentiles 25 y 75, denotados por  $Q_1$  y  $Q_3$ , respectivamente, y el IQR se definió como:

$$\text{IQR} = Q_3 - Q_1.$$

Los límites para la detección de valores extremos se determinaron según el criterio

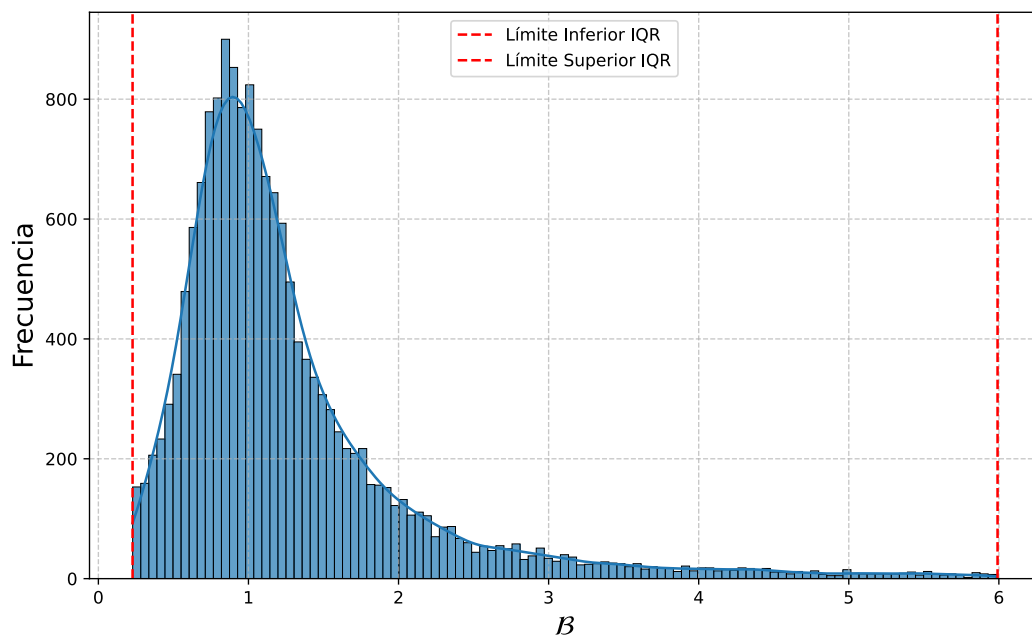
$$x_i \notin [Q_1 - 1.5 \cdot \text{IQR}, Q_3 + 1.5 \cdot \text{IQR}].$$

A diferencia de los enfoques tradicionales, en este estudio no se descartaron los valores fuera de ese rango, sino que se conservaron aquellos puntos ubicados fuera del intervalo intercuartílico extendido, es decir, los más alejados de la región ambigua del umbral. Estos valores extremos fueron transformados de nuevo a escala lineal mediante la función exponencial, obteniéndose los siguientes límites de conservación:

$$\mathcal{B} < 0.23 \quad \text{o} \quad \mathcal{B} > 6.00.$$

De esta manera, el conjunto retenido incluyó únicamente aquellas observaciones que reflejan una separación más pronunciada entre las densidades interictales y preictales. Esta estrategia permitió excluir regiones cercanas al valor  $\mathcal{B} = 1$ , altamente sensibles al ruido o a pequeñas variaciones geométricas, y enfocarse en configuraciones morfológicas con mayor poder discriminativo. Como resultado, el análisis de clasificación binaria posterior se realizó sobre una base más confiable y representativa de las diferencias funcionales entre estados.

**Figura 4.3.**  
*Histograma de valor del biomarcador  $\mathcal{B}$*



*Nota.* distribución empírica de valores individuales del biomarcador  $\mathcal{B}$  en todos los segmentos de todos los pacientes y todas las bandas de frecuencia. Acotación en Q1 y Q3.

# CAPÍTULO 5

## Detección y predicción de estados preictales

Este capítulo presenta los resultados obtenidos tras la aplicación de la metodología propuesta al conjunto de datos clínicos. Se reporta la discriminación entre estados interictales y preictales a nivel de segmentos diarios, así como el desempeño del clasificador a lo largo de las distintas bandas de frecuencia. Adicionalmente, se describe un análisis de predicción pseudo-prospectiva que emplea la misma metodología para anticipar el estado ictal de un día únicamente a partir de los registros previos, lo cual representa un paso hacia su implementación en esquemas reales de monitoreo y alerta.

### 5.1. Resultados de clasificación de estados

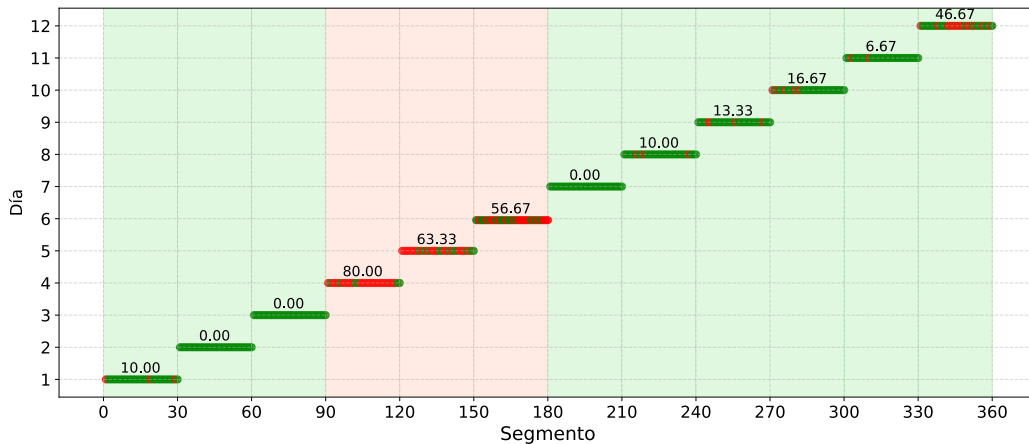
La clasificación de estados se llevó a cabo mediante el esquema de validación cruzada *Leave-One-Out* aplicado a nivel de segmentos. En este enfoque, cada uno de los segmentos se considera, de forma iterativa, como dato desconocido y se excluye temporalmente del conjunto de testeo. Dicho segmento es proyectado en el espacio euclídeo y alineado con respecto a un segmento de referencia interictal del mismo paciente. Se extraen las coordenadas correspondientes a los tres nodos más discriminativos del paciente, seleccionados con anterioridad según sus valores elevados de  $Z(i)$ . Para cada uno de estos nodos, se calcula el valor del biomarcador  $\mathcal{B}$  al comparar su posición en el espacio proyectado frente a las distribuciones nodales de referencia construidas con los segmentos restantes.

El valor de  $\mathcal{B}$  obtenido para cada nodo indica la afinidad del nodo proyectado hacia uno u otro estado. Para decidir la clase del segmento, se toma el promedio categórico de los tres nodos: si al menos dos de ellos son clasificados como preictales (es decir, presentan  $\mathcal{B} \leq 1$  (superando el umbral), el segmento se etiqueta como preictal; en caso contrario, se clasifica como interictal. Este procedimiento se repite iterativamente para todos los segmentos del paciente y de todas las bandas de frecuencia. Es importante destacar que, una vez clasificado un segmento, este se incorpora al conjunto de segmentos conocidos y puede ser utilizado como referencia para clasificaciones posteriores.

Finalmente, la probabilidad de clasificación de un día completo se calcula como la proporción de sus segmentos clasificados como preictales. Si al menos el 50% de los segmentos de ese día cumplen esta condición, el día se considera preictal; en caso contrario, se clasifica como interictal. Este procedimiento se ejemplifica en la [Figura 5.1](#), que muestra el caso del paciente 8 en la banda Low  $\gamma$ , con un registro de doce días, tres de los cuales corresponden a estados preictales y nueve a estados interictales. Todos clasificados correctamente.

### Figura 5.1.

*Resultado de clasificación*



*Nota.* Resultados de clasificación diaria para el paciente 8 en la banda Low  $\gamma$ . Se indica la proporción de segmentos diarios predichos como preictales mediante el esquema de validación cruzada *Leave-One-Out*. El fondo indica el estado conocido a priori, verde (interictal) y rojo (preictal).

Esta representación permite visualizar con mayor granularidad la distribución temporal de las predicciones preictales e interictales a lo largo de cada día. Si bien cada día posee una etiqueta global conocida (preictal o interictal), no todos los segmentos dentro de dicho día son clasificados de forma homogénea. Este patrón refleja la variabilidad intrínseca de la señal iEEG, influida por la dinámica fisiológica cerebral, la evolución temporal del estado clínico y la posible existencia de períodos intermedios o transicionales difíciles de asignar a una sola categoría.

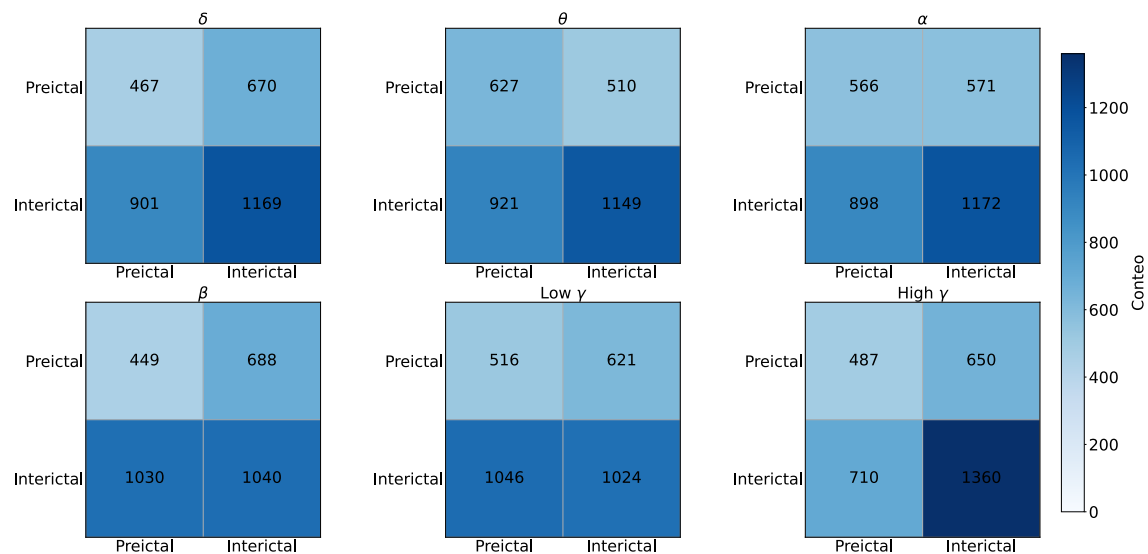
El paciente 8 presentó una clasificación perfecta del 100 % en todos sus días registrados cuando se utilizó la banda Low  $\gamma$ , lo que indica una capacidad discriminativa sobresaliente del modelo en esa frecuencia específica para este individuo. Ahora, este rendimiento no se replica en otras bandas del mismo paciente, donde la proporción de días correctamente clasificados puede disminuir considerablemente, llegando incluso a valores cercanos al azar. Esta variabilidad entre bandas puede atribuirse a múltiples factores: la individualidad de los patrones electrofisiológicos del paciente, la naturaleza específica de

su tipo de epilepsia, y la posible sensibilidad diferencial de cada banda de frecuencia frente a los cambios en la conectividad cerebral asociados al estado preictal. Es bien conocido que distintas bandas reflejan distintos procesos neurofisiológicos, como actividad tónica, oscilaciones relacionadas con la atención, o sincronización cortical, por lo que es esperable que su utilidad para la predicción de crisis varíe sustancialmente entre sujetos y condiciones.

Para evaluar de forma general la efectividad del modelo, se analizaron los resultados combinados de todos los pacientes en todas las bandas de frecuencia. Para ello, se generaron matrices de confusión agregadas, considerando las clases 0 (estado preictal) y 1 (estado interictal) [Figura 5.2](#). En estas matrices, cada elemento representa la cantidad de predicciones realizadas por el modelo frente a las verdaderas etiquetas. Los verdaderos positivos (VP) corresponden a predicciones correctas del estado preictal, mientras que los verdaderos negativos (VN) indican predicciones correctas del estado interictal. Los falsos positivos (FP) representan casos en los que se predijo erróneamente un estado preictal cuando en realidad era interictal, y los falsos negativos (FN) corresponden al caso inverso.

**Figura 5.2.**

*Matrices de confusión*



*Nota.* Matrices de confusión agregadas por banda de frecuencia para una iteración. Cada subgráfico muestra el desempeño del modelo en la clasificación entre estados preictales e interictales para una banda específica. Estados totales clasificados por banda 3.207

Estas matrices son una herramienta fundamental para evaluar el rendimiento del modelo de clasificación binaria. Al organizar los conteos de predicciones correctas e incorrectas en función de las clases verdaderas y predichas, permiten no solo visualizar el tipo de errores cometidos, sino también calcular múltiples métricas de desempeño que

ofrecen una caracterización más completa de la capacidad del modelo para distinguir entre los estados preictal e interictal. Dado que el objetivo principal del modelo es detectar los estados preictales, la sensibilidad (también conocida como *recall*) cobra particular importancia.

Esta métrica refleja la proporción de verdaderos positivos sobre el total de eventos verdaderamente preictales. Una baja sensibilidad implica que se estarían pasando por alto eventos críticos, lo cual resulta clínicamente indeseable. Además, debido al desbalance inherente en la cantidad de datos entre estados preictales e interictales, es necesario utilizar métricas que no se vean afectadas por dicho desequilibrio. Entre ellas se encuentran el *F1-score*, y la precisión balanceada (*balanced accuracy*).

El *F1-score* es la media armónica entre la precisión y la sensibilidad, y se expresa mediante la siguiente fórmula:

$$F1\text{-score} = 2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN}$$

esta métrica combina de forma equilibrada la capacidad del modelo para evitar falsos positivos (precisión) y para detectar los verdaderos eventos positivos (sensibilidad). En el contexto de este estudio, el *F1-score* permite evaluar qué tan eficaz es el modelo para identificar correctamente los días preictales, penalizando simultáneamente los errores de omisión (falsos negativos) y de falsa alarma (falsos positivos). Su rango está acotado en el intervalo  $[0, 1]$ , donde 1 representa una clasificación perfecta, y 0 indica que el modelo no detecta correctamente ningún evento preictal. Si el modelo predice correctamente algunos días preictales (altos VP), pero al mismo tiempo comete muchos errores (altos FP o FN), el *F1-score* disminuirá drásticamente, reflejando que el rendimiento no es confiable. Esto es particularmente importante en escenarios desbalanceados como el presente, donde los días interictales son dos (2) veces más frecuentes.

Por su parte, la precisión balanceada se define como el promedio entre la sensibilidad y la especificidad (o tasa de verdaderos negativos):

$$\text{Precisión balanceada} = \frac{1}{2} \left( \frac{VP}{VP + FN} + \frac{VN}{VN + FP} \right) \quad (5.1)$$

esta métrica es útil para cuantificar el rendimiento del modelo de forma equitativa entre ambas clases, lo que permite una evaluación justa incluso cuando una clase (como la preictal) está subrepresentada ya que la precisión balanceada otorga el mismo peso al desempeño del modelo en ambas clases, independientemente de cuántas muestras tenga cada una.

Su valor también se encuentra en el intervalo  $[0, 1]$ , siendo 1 el caso ideal de

clasificación perfecta y 0.5 el valor esperado en un clasificador aleatorio que predice sin aprender nada (es decir, con sensibilidad y especificidad del 50%). En nuestro caso, si el modelo predice correctamente casi todos los días interictales (alta especificidad), pero falla en identificar los días preictales (baja sensibilidad), la precisión balanceada revelará esta deficiencia al reflejar un promedio bajo.

El desempeño global del modelo, promediado entre todos los pacientes y todas las bandas de frecuencia, se muestra en la [Tabla 5.1](#). Estos resultados permiten valorar la estabilidad del modelo en poblaciones clínicas heterogéneas y la sensibilidad diferencial según el contenido espectral de las señales. Ambas métricas toman valores entre 0 y 1.

**Tabla 5.1**

*Métricas globales de discriminación entre estados interictales y preictales*

<b>Banda</b>	<b>F1-score</b>	<b>Precisión balanceada</b>
$\delta$	$0.52 \pm 0.27$	$0.53 \pm 0.14$
$\theta$	$0.52 \pm 0.25$	$0.55 \pm 0.19$
$\alpha$	$0.52 \pm 0.23$	$0.52 \pm 0.16$
$\beta$	$0.49 \pm 0.26$	$0.50 \pm 0.19$
<i>Low <math>\gamma</math></i>	$0.51 \pm 0.20$	$0.51 \pm 0.16$
<i>High <math>\gamma</math></i>	$0.60 \pm 0.23$	$0.57 \pm 0.12$
<b>Mejor banda por paciente</b>	$0.70 \pm 0.23$	$0.68 \pm 0.12$

*Nota.* Desempeño general del modelo para la clasificación entre estados interictales y preictales. Se reportan los valores promedio del *F1-score* y la precisión balanceada para cada banda de frecuencia y el desempeño específico con la mejor banda de todos los pacientes.

Al promediar las métricas de clasificación binaria a lo largo de todos los pacientes para cada banda de frecuencia, se identificaron bandas específicas en las que el modelo alcanzó un mayor poder discriminativo. En particular, la banda *High  $\gamma$*  mostró el mejor desempeño global, con un *F1-score* promedio de  $0.60 \pm 0.23$  y una precisión balanceada de  $0.57 \pm 0.12$ , superando al resto del espectro. No obstante, en términos generales, tanto los valores promedio como la dispersión de las métricas para las demás bandas se mantuvieron cercanos a 0.5, lo cual sugiere un comportamiento clasificatorio equivalente al esperado por azar.

Esta aparente limitación en el desempeño global contrasta con los resultados obtenidos al considerar únicamente la banda de frecuencia más discriminativa para cada paciente. Esta selección individualizada elevó el rendimiento promedio a un *F1-score* de  $0.70 \pm 0.23$  y una precisión balanceada de  $0.68 \pm 0.12$ , lo que evidencia que la metodología propuesta logra capturar información significativa cuando se identifican adecuadamente

las bandas más informativas en cada caso. Para el caso global de elección de la mejor banda, el modelo planteado presenta una **sensibilidad** (o recall) de 0.69.

Comparando nuestros resultados con trabajos previos realizados sobre los mismos pacientes, se observa que el desempeño de nuestro modelo se encuentra dentro de un rango competitivo frente a otros enfoques, específicamente los llevados a cabo con los registros iEEG de nuestros pacientes proyectados sobre un espacio hiperbólico (Guillemaud et al., 2025) y a través de algoritmos de aprendizaje automático (*support vector machine*) (Cousyn et al., 2023). A continuación, se presenta una tabla comparativa que resume las métricas clave reportadas en dichos estudios y las obtenidas en el presente trabajo.

**Tabla 5.2**

*Comparación de métricas globales entre diferentes enfoques de clasificación.*

Método	F1-score	Precisión balanceada	Precisión
Euclídeo	$0.70 \pm 0.23$	$0.68 \pm 0.12$	–
Hiperbólico	$0.78 \pm 0.18$	–	$0.77 \pm 0.13$
Support Vector Machine	$0.79 [0.62-1.00]$	–	$0.85 [0.62-1.00]$

*Nota.* Se reportan los valores promedio  $\pm$  desviación estándar o, cuando corresponde, los rangos [mín–máx]. Los métodos comparados corresponden a un análisis en espacios hiperbólicos y a un clasificador SVM entrenado con valores de PLV en la zona de inicio de crisis (SOZ) (Guillemaud et al., 2025; Cousyn et al., 2023).

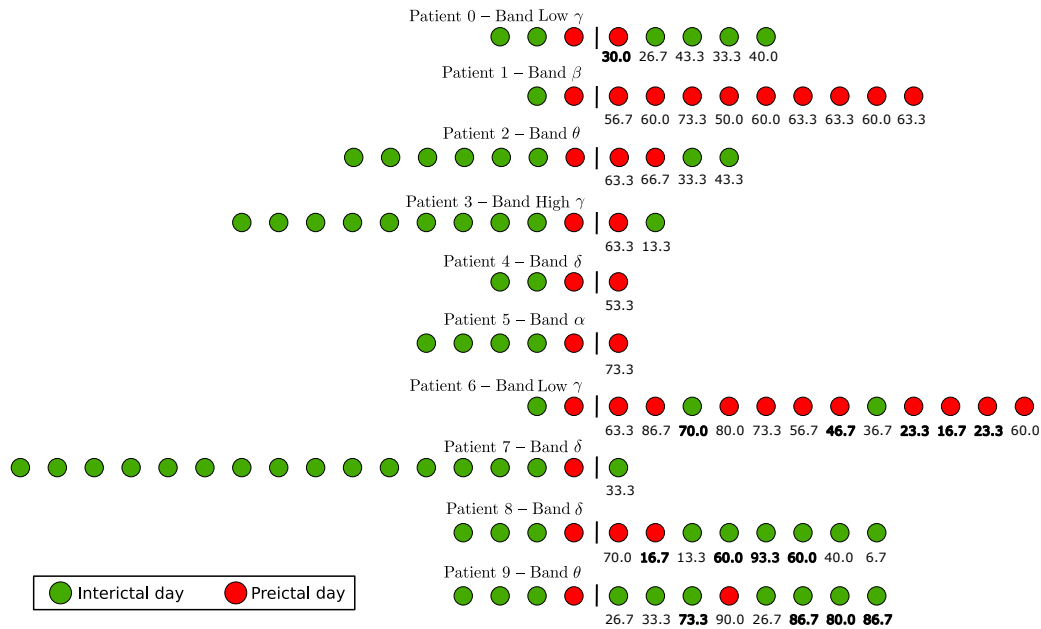
## 5.2. Resultados de predicción diaria de riesgo de crisis

Con el fin de evaluar la utilidad clínica del biomarcador  $\mathcal{B}$  para la predicción anticipada de crisis epilépticas, se implementó un esquema de predicción pseudo-prospectiva. En este enfoque, se utilizó la misma metodología planteada a lo largo del trabajo de investigación, pero se simuló un escenario de pronóstico realista, en el que para cada día por predecir, se entrenó el modelo únicamente con los días anteriores, asegurando que el conjunto de entrenamiento incluyera al menos un día interictal y uno preictal. Esta condición garantiza una base mínima para construir las distribuciones nodales de referencia necesarias para el cálculo de  $\mathcal{B}$ . Para cada iteración de predicción, se identificaron los tres nodos más discriminativos a partir de los datos de entrenamiento, y se calculó el biomarcador  $\mathcal{B}$  en cada uno de los 30 segmentos del día siguiente. La probabilidad de que un día fuera clasificado como preictal se definió como la proporción de segmentos etiquetados como tal; si esta proporción era mayor o igual al 50%, el día completo se consideraba preictal, en caso contrario, interictal. Una vez clasificado, el día

predicho se incorpora al conjunto de entrenamiento con su etiqueta verdadera conocida, lo cual permite actualizar iterativamente la base de conocimiento del modelo y extender el proceso de predicción hacia los días posteriores. Los resultados por paciente se resumen en la [Figura 5.3](#).

**Figura 5.3.**

*Resultados de predicción pseudo-prospectiva por paciente*



*Nota.* Desempeño del modelo en la predicción pseudo-prospectiva, donde todos los días previos fueron utilizados como entrenamiento para predecir el día siguiente. Se presenta el mejor resultado obtenido para cada paciente, utilizando su banda de frecuencia más discriminativa. Las probabilidades que no coincidieron con la clase real del día se muestran en negrilla.

Al examinar el desempeño individual, se observa que 6 de los 10 pacientes (pacientes 1, 2, 3, 4, 5 y 7) fueron clasificados correctamente en todos sus días registrados, lo cual confirma la capacidad del modelo para detectar señales preictales específicas en determinados sujetos. A nivel global, se clasificaron correctamente 22 de los 28 días preictales (sensibilidad del 78.57%) y 15 de los 23 días interictales (especificidad del 65.22%). Estos resultados consolidan la evidencia de que el modelo propuesto, cuando se adapta a las características espectrales individuales, permite una predicción anticipada de crisis epilépticas con alta efectividad.

Para cuantificar el desempeño del modelo en este escenario predictivo, se utilizaron dos métricas complementarias: la precisión (*accuracy*) y el *Brier score*. La precisión se define como la proporción de días correctamente clasificados entre el total, y permite

evaluar de forma directa la tasa de aciertos del modelo:

$$\text{Precisión (accuracy)} = \frac{VP + VN}{VP + VN + FP + FN},$$

esta métrica toma valores entre 0 y 1, siendo 1 el valor óptimo. Por su parte, el *Brier score* mide la calidad de la probabilidad asignada a la predicción y penaliza especialmente los errores cometidos con alta confianza. Se define como:

$$\text{Brier score} = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2,$$

donde  $f_i$  es la probabilidad predicha para el evento (dado por la proporción de segmentos preictales),  $o_i$  es la etiqueta real binaria del día (1 si es preictal, 0 si es interictal), y  $N$  es el número total de días evaluados. El valor del *Brier score* también varía entre 0 y 1, siendo 0 el valor ideal que indica una predicción perfectamente calibrada.

Los resultados promedio de estas métricas, calculados sobre todos los pacientes para todas las bandas de frecuencia, se presentan en la [Tabla 5.3](#). Al igual que en la tarea de clasificación por validación cruzada, las métricas globales se aproximan a los niveles por azar ( $\sim 0.5$ ), lo que refleja una limitada capacidad predictiva al considerar uniformemente todas las bandas. No obstante, este resultado cambia de forma significativa cuando se selecciona, para cada paciente, su banda de frecuencia más discriminativa. En este caso, la precisión promedio alcanza un valor de  $0.65 \pm 0.14$  y el *Brier score* disminuye a  $0.36 \pm 0.14$ , evidenciando un aumento considerable tanto en la exactitud como en la calibración probabilística del modelo.

Comparando directamente las métricas de desempeño predictivo con los trabajos similares previamente mencionados, se construyó la siguiente tabla con el fin de contextualizar los resultados obtenidos en este estudio. En ella se resumen los valores globales de precisión (*accuracy*) y del *Brier score*, permitiendo evaluar no solo la tasa de aciertos del modelo, sino también la calidad de las probabilidades asignadas en sus predicciones [Tabla 5.4](#).

**Tabla 5.3***Métricas globales de predicción pseudo-prospectiva de estados preictales*

<b>Banda</b>	<b>Precisión</b>	<b>Brier score</b>
$\delta$	$0.55 \pm 0.17$	$0.45 \pm 0.17$
$\theta$	$0.55 \pm 0.09$	$0.45 \pm 0.09$
$\alpha$	$0.49 \pm 0.15$	$0.51 \pm 0.15$
$\beta$	$0.50 \pm 0.14$	$0.50 \pm 0.14$
<i>Low <math>\gamma</math></i>	$0.51 \pm 0.12$	$0.49 \pm 0.12$
<i>High <math>\gamma</math></i>	$0.49 \pm 0.13$	$0.51 \pm 0.13$
<b>Mejor banda por paciente</b>	$0.65 \pm 0.14$	$0.36 \pm 0.14$

*Nota.* Resultados de predicción pseudo-prospectiva entre estados interictales y preictales. Se reportan los valores promedio de precisión (*accuracy*) y *Brier score* para cada banda de frecuencia, así como el desempeño al seleccionar la mejor banda individual por paciente.

**Tabla 5.4***Comparación de métricas globales entre diferentes enfoques de predicción.*

<b>Método</b>	<b>Precisión (<i>accuracy</i>)</b>	<b>Brier Score</b>
Euclídeo	$0.65 \pm 0.14$	$0.36 \pm 0.14$
Hiperbólico	$0.87 \pm 0.17$	$0.12 \pm 0.12$
Support Vector Machine	–	0.13 [0.00–0.35]

*Nota.* Se presentan métricas globales de desempeño para los distintos enfoques aplicados sobre el mismo conjunto de pacientes. Los valores corresponden a la media  $\pm$  desviación estándar o, cuando no está disponible, al rango [mín–máx]. Los modelos de comparación incluyen el análisis en espacios hiperbólicos y un clasificador SVM ([Guillemaud et al., 2025](#); [Cousyn et al., 2023](#)).

## CAPÍTULO 6

### Conclusiones y discusión

En este capítulo final se presentan cinco conclusiones principales derivadas del trabajo de investigación, las cuales se discuten a la luz de la metodología desarrollada y de los resultados obtenidos. Estas conclusiones sintetizan los aportes más relevantes del estudio y destacan su importancia en el contexto de la predicción personalizada del riesgo de crisis epilépticas. Como primera conclusión:

- Las representaciones de redes de conectividad cerebral proyectadas en espacios euclídeos de baja dimensión, derivadas de registros iEEG diarios en reposo, permiten discriminar eficazmente entre días interictales (sin crisis) y preictales (con crisis en las siguientes 24 horas).

El uso de proyecciones euclídeas permite capturar de forma robusta los cambios en la conectividad funcional asociados al estado preictal. Gracias a la estructura métrica bien definida del espacio euclídeo, este enfoque facilita un análisis interpretable de la dinámica cerebral, posibilitando la visualización de configuraciones espaciales en la red de conectividad que anteceden a la aparición de una crisis. Estos hallazgos respaldan trabajos previos que demuestran que las proyecciones en espacios de baja dimensión son suficientes para anticipar crisis epilépticas ([Duncan et al., 2013b](#)), sin necesidad de recurrir a inmersiones más complejas. No obstante, es importante seguir explorando en qué contextos podrían resultar ventajosos otros espacios geométricos, en especial aquellos que modelan estructuras jerárquicas ([Guillemaud et al., 2025](#)).

A diferencia de enfoques que requieren monitoreo EEG continuo de largo plazo, este estudio reafirma que sesiones breves diarias en reposo, bajo vigilancia, son suficientes para detectar dinámicas preictales relevantes ([Cousyn et al., 2023](#)). Esto abre la puerta a estrategias de monitoreo más prácticas y menos invasivas. Además, al proyectar la conectividad cerebral, la metodología permite identificar nodos cuya conectividad funcional difiere de forma consistente entre los estados interictal y preictal. Esto conduce a la segunda conclusión:

- El modelo permite identificar nodos discriminativos específicos que se mantienen estables a lo largo del tiempo.

Esta propiedad reviste relevancia clínica, al permitir estrategias de monitoreo e intervención personalizadas basadas en electrodos clave, posiblemente asociados a regiones críticas en la generación o propagación de actividad epileptica. De forma consistente, se observó que los nodos más discriminativos tienden a conservarse entre días de registro para un mismo paciente, lo que sugiere su implicación estable en la susceptibilidad a crisis.

El rendimiento del modelo puede verse influido por la elección del segmento de referencia interictal, que define la distribución base para el cálculo del biomarcador. Esta dependencia introduce una fuente de variabilidad que podría mitigarse mediante estrategias de selección adaptativa. Futuros trabajos deberían explorar métodos más robustos para definir dicha referencia y evaluar si los nodos identificados conservan su relevancia en contextos clínicos complejos, como durante reducción de medicación o intervención quirúrgica.

Un componente central del proceso desarrollado es el biomarcador  $\mathcal{B}$ , lo que conduce a la tercera conclusión:

- El biomarcador  $\mathcal{B}$  propuesto constituye una herramienta prometedora para la clasificación y predicción del riesgo de crisis epilépticas.

Este biomarcador interpretable permite discriminar de forma robusta entre estados interictales y preictales a partir de registros breves diarios, detectando señales confiables de crisis inminentes y ofreciendo una ventana temporal más amplia para la intervención clínica o el monitoreo personalizado. En cuanto al desempeño del modelo, se destaca lo siguiente como cuarta conclusión:

- El modelo basado en inmersiones euclídeas alcanza niveles de desempeño competitivos en tareas de clasificación y predicción, especialmente cuando se emplean bandas de frecuencia personalizadas por paciente.

Aunque las métricas globales presentan valores cercanos al azar, el análisis individual revela una mejora notable al seleccionar la mejor banda por paciente. En clasificación, se alcanzó un *F1-score* promedio de  $0.70 \pm 0.23$  y una precisión balanceada de  $0.68 \pm 0.12$ . En predicción pseudo-prospectiva, la precisión fue de  $0.65 \pm 0.14$  con un *Brier score* de  $0.36 \pm 0.14$ , demostrando la capacidad del modelo para anticipar estados preictales. Si bien proyecciones más complejas, como aquellas en los espacios hiperbólicos, pueden ofrecer mejoras adicionales (Guillemaud et al., 2025), nuestros resultados respaldan la eficacia de las proyecciones euclídeas. El rendimiento alcanzado es comparable al de modelos como el *support vector machine* (SVM), entrenado con regiones seleccionadas manualmente dentro de la zona de inicio de crisis (Cousyn et al., 2023). De los resultados de clasificación y predicción también se desprende una quinta conclusión:

- Las bandas  $\delta$ ,  $\theta$  y *Low*  $\gamma$  concentran la mayor parte de la información discriminativa, aunque su efectividad varía entre pacientes.

La sensibilidad del modelo a la banda empleada resalta la naturaleza individualizada de la dinámica epiléptica. Estas tres bandas concentraron la mayor parte de la información útil. La banda  $\theta$  mostró los mejores desempeños entre pacientes, en línea con estudios que destacan su papel en la actividad neural relacionada con crisis (Cousyn et al., 2023). La banda *Low*  $\gamma$  también mostró alta capacidad discriminativa, reforzando observaciones sobre su rol en la configuración del estado preictal (Guillemaud et al., 2025).

Sin embargo, no existe una banda universalmente óptima. La elección más adecuada depende de las características individuales de cada paciente, lo que subraya la necesidad de análisis espectrales personalizados para mejorar la sensibilidad del modelo en la práctica clínica.

Finalmente, más allá de la clasificación y predicción, esta metodología permite comprender mejor cómo cambia la conectividad cerebral antes de una crisis. Las reconfiguraciones funcionales en espacios de baja dimensión ofrecen una representación compacta y dinámica de la evolución espacio-temporal de la red neuronal.

Este trabajo fue presentado como contribución oral en la escuela internacional LA-CoNGA physics 2023 y constituye la base de un artículo actualmente enviado a una revista arbitrada para su posible publicación. El artículo está disponible en <https://arxiv.org/abs/2505.00856>.

---

## Referencias Bibliográficas

- Allard, A. and Serrano, M. Á. (2020). Navigable maps of structural brain networks across species. *PLoS Computational Biology*, 16(2):e1007584.
- Aydore, S., Pantazis, D., and Leahy, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage*, 74:231–244.
- Bassett, D. S. and Bullmore, E. T. (2009). Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340–347.
- Bassett, D. S. and Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308.
- Boguna, M., Bonamassa, I., De Domenico, M., Havlin, S., Krioukov, D., and Serrano, M. Á. (2021). Network geometry. *Nature Reviews Physics*, 3(2):114–135.
- Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198.
- Chávez, M., Le Van Quyen, M., Navarro, V., Baulac, M., and Martinerie, J. (2003). Spatio-temporal dynamics prior to neocortical seizures: amplitude versus phase couplings. *IEEE Transactions on Biomedical Engineering*, 50(5):571–583.
- Chavez, M., Valencia, M., Navarro, V., Latora, V., and Martinerie, J. (2010). Functional modularity of background activities in normal and epileptic brain networks. *Physical Review Letters*, 104.
- Coifman, R. R. and Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30.
- Cousyn, L., Messaoud, R. B., Lehongre, K., Frazzini, V., Lambrecq, V., Adam, C., Mathon, B., Navarro, V., and Chavez, M. (2023). Daily resting-state intracranial eeg connectivity for seizure risk forecasts. *Epilepsia*, 64(2):e23–e29.
- De Vico Fallani, F., Latora, V., and Chavez, M. (2017). A topological criterion for filtering information in complex brain networks. *PLoS computational biology*, 13(1):e1005305.

- De Vico Fallani, F., Richiardi, J., Chavez, M., and Achard, S. (2014). Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1653):20130521.
- Dijksterhuis, G. B. and Gower, J. C. (1991). The interpretation of generalized procrustes analysis and allied methods. *Food quality and preference*, 3(2):67–87.
- Duncan, D., Duckrow, R. B., Pincus, S. M., Goncharova, I., Hirsch, L. J., Spencer, D. D., Coifman, R. R., and Zaveri, H. P. (2013a). Intracranial eeg evaluation of relationship within a resting state network. *Clinical Neurophysiology*, 124(10):1943–1951.
- Duncan, D., Talmon, R., Zaveri, H. P., and Coifman, R. R. (2013b). Identifying pre seizure state in intracranial eeg data using diffusion kernels. *Mathematical Biosciences & Engineering*, 10(3):579–590.
- Eckhorn, R. (1994). Oscillatory and non-oscillatory synchronizations in the visual cortex and their possible roles in associations of visual features. *Progress in brain research*, 102:405–426.
- Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., and Engel Jr, J. (2005). Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia*, 46(4):470–472.
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in cognitive sciences*, 9(10):474–480.
- Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Elsevier.
- Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-part III: radio and communication engineering*, 93(26):429–441.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.
- Guillemaud, M., Cousyn, L., Navarro, V., and Chavez, M. (2025). Hyperbolic embedding of brain networks as a tool for epileptic seizures forecasting. *Physical Review Research*. *In press*.
- Guillemaud, M., Longhena, A., Cousyn, L., Frazzini, V., Mathon, B., Navarro, V., and Chavez, M. (2024). Hyperbolic embedding of brain networks can predict the surgery outcome in temporal lobe epilepsy. *npj Systems Bioogy and Applications*. *Articulo aceptado y en proceso de publicación*.

- Gürsoy, F., Haddad, M., and Bothorel, C. (2023). Alignment and stability of embeddings: measurement and inference improvement. *Neurocomputing*, 553:126517.
- Ilmoniemi, R. J. and Sarvas, J. (2019). *Brain signals: physics and mathematics of MEG and EEG*. Mit Press.
- Jobst, B. C., Bartolomei, F., Diehl, B., Frauscher, B., Kahane, P., Minotti, L., Sharan, A., Tardy, N., Worrell, G., and Gotman, J. (2020). Intracranial eeg in the 21st century. *Epilepsy currents*, 20(4):180–188.
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., Mack, S., et al. (2000). *Principles of neural science*, volume 4. McGraw-hill New York.
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1):48–50.
- Kwan, P., Schachter, S. C., and Brodie, M. J. (2011). Drug-resistant epilepsy. *New England Journal of Medicine*, 365(10):919–926.
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208.
- Lafon, S. and Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1393–1403.
- Longhena, A., Guillemaud, M., and Chavez, M. (2024). Detecting local perturbations of networks in a latent hyperbolic embedding space. *Chaos*, 34:63117.
- Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of neuroscience methods*, 164(1):177–190.
- Mercier, M. R., Dubarry, A.-S., Tadel, F., Avanzini, P., Axmacher, N., Cellier, D., Del Vecchio, M., Hamilton, L. S., Hermes, D., Kahana, M. J., et al. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. *Neuroimage*, 260:119438.
- Moreira, G., Marques, M., Costeira, J. P., and Hauptmann, A. (2024). Hyperbolic vs euclidean embeddings in few-shot learning: Two sides of the same coin. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2082–2090.
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain: a journal of neurology*, 120(4):701–722.

- Mubaraki, A. A. H. (2020). *Predicting epileptic seizures from electroencephalography*. PhD thesis, University of Southampton.
- Nobel Prize Outreach (2021). The nobel prize in physics 2021. Accessed: 2025-05-08.
- Richiardi, J., Achard, S., Bunke, H., and Van De Ville, D. (2013). Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Processing Magazine*, 30(3):58–70.
- Stam, C., Tewarie, P., Van Dellen, E., Van Straaten, E., Hillebrand, A., and Van Mieghem, P. (2014). The trees and the forest: characterization of complex brain networks with minimum spanning trees. *International Journal of Psychophysiology*, 92(3):129–138.
- Strogatz, S. H. (2001). Exploring complex networks. *nature*, 410(6825):268–276.
- Van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., and Marinazzo, D. (2014). Functional brain connectivity from eeg in epilepsy: Seizure prediction and epileptogenic focus localization. *Progress in Neurobiology*, 121:19–35.
- Vemuri, V. R. (1978). *Modeling of Complex Systems: An Introduction*. Academic Press, New York.
- Vidal, M., Cusick, M. E., and Barabási, A.-L. (2011). Interactome networks and human disease. *Cell*, 144(6):986–998.
- World Health Organization, International League Against Epilepsy, and International Bureau for Epilepsy (2019). *Epilepsy: A public health imperative*. World Health Organization, Geneva.