



**Instituto Potosino de Investigación  
Científica y Tecnológica, A. C.**

**Posgrado en Control y Sistemas Dinámicos.**

**Modelo anatómico funcional del cerebro completo,  
asociado a la edad, basado en el proceso  
multivariado Ornstein Uhlenbeck.**

Tesis que presenta

**Sara Yadira Martínez Cano**

Para obtener el grado de

**Maestra en Control y Sistemas Dinámicos**

**Directores de la tesis:**

Dr. Cesar Octavio Maldonado Ahumanda

Dr. Cesaré Moisés Ovando Vázquez

San Luis Potosí, S.L.P., abril de 2026



# Constancia de aprobación de la tesis

---

Dr. César Octavio Maldonado Ahumanda  
Director de la tesis

---

Dr. Cesaré Moisés Ovando Vázquez  
Director de tesis

---

Dr. Rodrigo Alejandro Cofré Torres  
Sinodal

---

Dr. Juan Gonzalo Barajas Ramírez  
Sinodal

# Créditos Institucionales

La presente tesis fue elaborada en la División de Control y Sistemas Dinámicos y del Centro Nacional de Supercomputo, del Instituto Potosino de Investigación Científica y Tecnológica, A. C, bajo la dirección del Dr. César Octavio Maldonado Ahumada y del Dr. Cesaré Moisés Ovando Vázquez.

Durante la realización de la presente tesis, recibí beca de la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) con el CVU: 1321068 y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

Aprobación de tesis.

# Dedicatoria

La presente tesis la dedico a mi madre Sara Martínez Cano por su apoyo incondicional y amor constante.

# Agradecimientos

Agradezco al Dr. Rodrigo Cofré por la confianza depositada en mí al proporcionarme los datos necesarios para realizar este trabajo. Al Dr. Cesaré, le agradezco mucho su paciencia, conocimiento compartido y empatía durante todo el desarrollo de la tesis y al Dr. Cesar Maldonado por su apoyo, paciencia, orientación, conocimiento compartido, ánimo y sobre todo la confianza en la asignación de mi tema de tesis. Y al Dr. Tonámetl Sánchez, a quien admiro y a quien tengo como un referente como futura investigadora.

Expreso mi más profundo agradecimiento a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo otorgado mediante la beca con número de CVU 1321068, así como al Instituto Potosino de Investigación Científica y Tecnológica, A. C., por el respaldo y apoyo brindado durante la realización de la presente tesis.

# Resumen

El efecto del envejecimiento a nivel del cerebro completo es un enfoque que se aborda desde la neurociencia. Dilucidar la intrincada relación entre la estructura y la función cerebral, tanto en condiciones sanas como patológicas, constituye un reto para la neurociencia moderna [29]. En este contexto, comprender la dinámica subyacente de la actividad cerebral requiere de un marco teórico capaz de describir sistemas complejos, es por ello que desde la perspectiva de la física estadística, la actividad cerebral puede entenderse como la dinámica colectiva de redes cerebrales sometidas a fluctuaciones [14].

Bajo este enfoque, se emplea un modelo anatómico funcional basado en el proceso multivariado Ornstein Uhlenbeck, para describir la dinámica multivariada de series BOLD (por sus siglas en inglés, Blood Oxygen Level Dependent, dependientes del nivel de oxigenación sanguínea), las cuales constituyen una medida indirecta de la actividad neuronal, mientras que la estructura de la red cerebral está dada por datos de imágenes por tensor de difusión (DTI), lo que da lugar a un esqueleto de conexiones. Una vez ajustado el modelo proporciona una estimación de la conectividad dirigida que refleja el estado dinámico de la actividad BOLD ([21] y [58]). Lo que permite calcular la tasa de producción de entropía  $\Phi$ , la cual es una cantidad que caracteriza el grado de irreversibilidad de un proceso estocástico en estado estacionario fuera del equilibrio [10]. En el contexto de la actividad cerebral,  $\Phi$  puede interpretarse como una medida del grado en que la dinámica neuronal se aleja del equilibrio, reflejando la presencia de flujos dirigidos de interacción entre regiones cerebrales.

El modelo anatómico funcional se aplica a datos de conectividad funcional (FC) y conectividad estructural (SC) de 161 sujetos totalmente sanos en estado de reposo, con edades comprendidas entre 10 y 80 años. El análisis se desarrolla en tres etapas.

En la primera, el modelo anatómico funcional se aplica a los datos agrupados de acuerdo con la edad biológica de los sujetos. En la segunda etapa, se aborda el problema inverso, es decir, inferir la edad de los sujetos a partir de la dinámica cerebral, para este fin se emplean métodos de aprendizaje no supervisado sobre la concatenación de las matrices de FC y SC, como resultado se obtienen agrupaciones, denominadas clusters, derivadas de las propiedades intrínsecas de los datos, consiguiente se aplica el modelo anatómico funcional a los clusters. En la tercera etapa se buscan identificar biomarcadores a partir de características presentes en los datos de FC y SC. Con este propósito, se implementan modelos de aprendizaje de máquina supervisado utilizando la concatenación de FC y SC, así como modelos de aprendizaje profundo a partir

de las series temporales BOLD. Los biomarcadores se definen como aquellas características explicativas que presentan mayor relevancia en el desempeño de los modelos. Finalmente, el modelo anatómico funcional se aplica a los biomarcadores identificados. Lo anterior permite comparar la dinámica cerebral estructural y funcional asociada al envejecimiento.

**Palabras clave: envejecimiento cerebral, tasa de producción de entropía, dinámica cerebral y proceso multivariado Ornstein Uhlenbeck.**

# Summary

The effect of aging at the whole-brain level is a topic addressed within neuroscience. Elucidating the intricate relationship between brain structure and function, both in healthy and pathological conditions, remains a major challenge for modern neuroscience [29]. In this context, understanding the underlying dynamics of brain activity requires a theoretical framework capable of describing complex systems. From the perspective of statistical physics, brain activity can therefore be understood as the collective dynamics of brain networks subject to fluctuations [14].

Under this approach, an anatomical functional model based on the multivariate Ornstein Uhlenbeck process is employed to describe the multivariate dynamics of BOLD time series (Blood Oxygen Level Dependent), which constitute an indirect measure of neuronal activity. Meanwhile, the structure of the brain network is defined by diffusion tensor imaging (DTI) data, which provide a structural backbone of connections. Once fitted, the model provides an estimate of the directed connectivity that reflects the dynamical state of BOLD activity ([21] and [58]). This allows the computation of the entropy production rate  $\Phi$ , a quantity that characterizes the degree of irreversibility of a stochastic process in a stationary state out of equilibrium [10]. In the context of brain activity,  $\Phi$  can be interpreted as a measure of the extent to which neuronal dynamics deviate from equilibrium, reflecting the presence of directed interaction flows between brain regions.

The anatomical functional model is applied to functional connectivity (FC) and structural connectivity (SC) data from 161 healthy subjects in resting state, with ages ranging from 10 to 80 years. The analysis is carried out in three stages.

In the first stage, the anatomical functional model is applied to data grouped according to the biological age of the subjects. In the second stage, the inverse problem is addressed, namely inferring the age of the subjects from brain dynamics. For this purpose, unsupervised learning methods are applied to the concatenation of FC and SC matrices. As a result, clusters derived from the intrinsic properties of the data are obtained, and the anatomical functional model is subsequently applied to these clusters. In the third stage, the aim is to identify biomarkers from features present in the FC and SC data. To achieve this, supervised machine learning models are implemented using the concatenation of FC and SC matrices, as well as deep learning models based on BOLD time series. Biomarkers are defined as explanatory features that show the greatest relevance in the performance of the models. Finally, the anatomical functional model is applied to the identified biomarkers. This approach allows the comparison of

structural and functional brain dynamics associated with aging.

**Keywords:** brain aging, entropy production rate, brain dynamics, multivariate Ornstein–Uhlenbeck process.

# Índice general

<b>1. Introducción.</b>	<b>1</b>
1.1. Objetivos . . . . .	1
1.2. Hipótesis. . . . .	2
1.3. Organización de la tesis. . . . .	2
<b>2. Preliminares</b>	<b>4</b>
2.1. Definiciones de probabilidad y procesos estocásticos. . . . .	4
2.1.1. Momentos de una variable aleatoria . . . . .	6
2.1.2. Movimiento Browniano . . . . .	7
2.1.3. Ecuación diferencial estocástica. . . . .	8
2.1.4. Proceso de Itô . . . . .	8
2.1.5. Ecuaciones de Langevin . . . . .	9
2.1.6. Proceso de Wiener . . . . .	9
2.1.7. Ecuación de Fokker-Planck . . . . .	10
2.1.8. Entropía . . . . .	10
<b>3. Proceso Multivariado Ornstein Uhlenbeck</b>	<b>12</b>
3.1. Ornstein Uhlenbeck OU . . . . .	13
3.1.1. Solución al proceso OU . . . . .	13
3.2. Ornstein Uhlenbeck Multivariado . . . . .	16
3.2.1. Solución al proceso MOU . . . . .	18
3.2.2. Evolución del estado estacionario del proceso MOU . . . . .	18
3.2.3. Covarianza sin desfase $t = t'$ . . . . .	19
3.2.4. Covarianza con desfase $t' > t$ . . . . .	20
3.3. Distribución de probabilidad . . . . .	20
3.3.1. Estado estacionario y corriente de probabilidad . . . . .	21
<b>4. Irreversibilidad y tasa de producción de entropía.</b>	<b>23</b>
4.1. Tasa de producción de entropía . . . . .	23
4.2. Reversibilidad e irreversibilidad . . . . .	26
<b>5. Datos y preprocesamiento</b>	<b>27</b>
5.1. Elementos de Neurociencia . . . . .	27
5.1.1. Datos funcionales. . . . .	28
5.1.2. Resonancia Magnética funcional (functional Magnetic Resonance Imaging fMRI) . . . . .	28
5.1.3. Datos estructurales . . . . .	29
5.1.4. Imágenes por Tensor de Difusión (Diffusion Tensor Imaging DTI)	29

5.1.5.	Parcelación. Módulos de estructura y función SFM . . . . .	29
5.1.6.	Conceptos . . . . .	30
5.2.	Preprocesamiento de datos . . . . .	32
5.2.1.	Transformada rápida de Fourier . . . . .	35
<b>6.</b>	<b>Estudios previos aplicados a los datos.</b>	<b>36</b>
6.1.	La información de interacción . . . . .	36
6.1.1.	Información de interacción en series de tiempo fMRI en estado de reposo . . . . .	37
6.1.2.	ROIs derivados de la información de interacción . . . . .	39
6.2.	Interdependencia de alto orden. . . . .	39
6.2.1.	Información sobre la estructura organizativa . . . . .	39
6.2.2.	$\Omega$ -Información en series de tiempo fMRI en estado de reposo. . . . .	40
6.2.3.	ROIs derivados de la $\Omega$ -Información. . . . .	41
6.3.	Modelo neurodegenerativo. . . . .	41
6.3.1.	Campo medio dinámico . . . . .	42
6.3.2.	Ajuste del modelo en $I_n$ . . . . .	44
6.3.3.	ROIs derivados del modelo neurodegenerativo. . . . .	44
<b>7.</b>	<b>Métodos.</b>	<b>45</b>
7.1.	Modelo anatómico funcional aplicado a grupos. . . . .	45
7.1.1.	Matrices de covarianza empírica $\hat{S}_{ij}(0)$ y $\hat{S}_{ij}(1)$ . . . . .	47
7.1.2.	Estimación de parámetros . . . . .	50
7.1.3.	Validación del modelo . . . . .	54
7.1.4.	Clasificación . . . . .	55
7.1.5.	Tasa de producción de entropía e (Ir)reversibilidad . . . . .	57
7.2.	Aprendizaje de máquina no supervisado. . . . .	58
7.2.1.	Clusters. . . . .	65
7.2.2.	Modelo anatómico funcional aplicado a los clusters. . . . .	67
7.3.	Aprendizaje de máquina supervisado. . . . .	69
7.3.1.	Aprendizaje profundo. . . . .	80
7.3.2.	Biomarcadores. . . . .	91
7.3.3.	Modelo anatómico funcional aplicado a biomarcadores. . . . .	92
<b>8.</b>	<b>Resultados y Discusión.</b>	<b>95</b>
<b>9.</b>	<b>Conclusiones</b>	<b>99</b>
	<b>Anexos</b>	<b>110</b>
	<b>Referencias</b>	<b>117</b>

# Índice de figuras

3.1. Proceso Ornstein Uhlenbeck . . . . .	15
3.2. Esperanza del proceso Ornstein Uhlenbeck $\mathbb{E}(X_t) = x_0 e^{-\alpha t}$ . . . . .	16
3.3. Varianza del proceso Ornstein Uhlenbeck . . . . .	16
5.1. Partes generales del cerebro . . . . .	27
5.2. Preprocesamiento de la conectividad funcional y estructural . . . . .	34
5.3. Concatenación horizontal de la FC y SC. . . . .	34
5.4. Transformada rápida de Fourier FFT . . . . .	35
7.1. Series BOLD filtradas de los 20 ROIs. . . . .	47
7.2. Matrices de covarianza empíricas $\hat{S}(0)_{ij}$ y $\hat{S}(1)_{ij}$ . . . . .	48
7.3. Distribución de valores de la SC . . . . .	49
7.4. Matriz de conectividad estructural (SC) de 20 ROIs y su correspondiente máscara booleana. . . . .	50
7.5. Máscara booleana $D$ para los 20 ROIs. . . . .	52
7.6. Representación esquemática del procedimiento de optimización de Lyapunov . . . . .	53
7.7. Matrices de covarianza, empírica $\hat{S}(0)_{ij}$ y teórica $S(0)_{ij}$ de los grupos $I_n$ con 20 ROIs. . . . .	54
7.8. Error del modelo y bondad de ajuste para $I_n$ con 20 ROIs. . . . .	55
7.9. Clasificación MLR y 1-NN para $I_n$ con 20 ROIs. . . . .	57
7.10. Matriz $Q_{ij}$ promedio de los 161 sujetos . . . . .	58
7.11. $Q_{ij}$ promedio de cada ROI por cada sujeto. . . . .	58
7.12. Gráficas de PCA en 2D y 3D . . . . .	60
7.13. Matriz de covarianza de los componentes principales obtenidos por PCA	61
7.14. t-SNE . . . . .	63
7.15. Método del codo . . . . .	64
7.16. Método de KMeans. . . . .	66
7.17. Comparación de los métodos de t-SNE y kMeans. . . . .	66
7.18. Matrices de covarianza empíricas $\hat{S}(0)_{ij}$ y $\hat{S}(1)_{ij}$ . . . . .	67
7.19. Matriz de covarianza empírica $\hat{S}(0)_{ij}$ y teórica $S(0)_{ij}$ de los clusters $C_n$ con 20 ROIs. . . . .	68
7.20. Error del modelo y bondad de ajuste para $C_n$ con 20 ROIs. . . . .	68
7.21. Clasificación MLR y 1-NN para $C_n$ con 20 ROIs . . . . .	69
7.22. Reporte del método de Random Forest Classifier. . . . .	71
7.23. Matriz de confusión para el método de Random Forest . . . . .	72
7.24. 50 características predominantes de acuerdo al modelo de random forest classifier. . . . .	72

7.25. Grafica SHAP de barras del modelo de random forest classifier. . . . .	73
7.26. ROIs predominantes en cada cluster $C_n$ . Gráfica de barras. . . . .	74
7.27. Descripción de los valores SHAP con el Random Forest Classifier . . . . .	75
7.28. Características predominantes en cada cluster $C_n$ . Gráfica de decisión, para el modelo de Random Forest Classifier. . . . .	76
7.29. Random Forest Regressor . . . . .	77
7.30. SHAP Random Forest Regressor . . . . .	78
7.31. Características predominantes generales para el modelo de Random Fo- rest Regressor. Gráfica de decisión. . . . .	79
7.32. Precisión y pérdida de los modelos densos. Deep Neural Networks, DNN	81
7.33. Gráficas de los modelos densos. Deep Neural Networks, DNN . . . . .	82
7.34. Precisión y pérdida de los modelos CNN-1D . . . . .	83
7.35. Gráficas de los modelos. Convolutional Neural Networks 1D . . . . .	83
7.36. Precisión y pérdida de los modelos CNN-2D. . . . .	84
7.37. Gráficas de los modelos. Gráficas de los modelos. Convolutional Neural Networks 2D . . . . .	85
7.38. Precisión y pérdida de los modelos recurrentes . . . . .	85
7.39. Gráficas de los modelos recurrentes . . . . .	86
7.40. SHAP para cada cluster $C_n$ importancia de las características del modelo Conv1D-three . . . . .	87
7.41. SHAP para cada cluster $C_n$ importancia de las características del modelo RNN . . . . .	88
7.42. Matrices de confusión para entrenamiento y validación para los modelos Conv 1D y recurrente. . . . .	89
7.43. Biomarcadores predominantes en los modelos aprendizaje de máquina y aprendizaje profundo. . . . .	90
7.44. Modelo híbrido . . . . .	91
7.45. Matriz de conectividad estructural (SC) de 5 ROIs y su correspondiente máscara booleana. . . . .	92
7.46. Matrices de covarianza empíricas $\hat{S}(0)_{ij}$ y $\hat{S}(1)_{ij}$ . . . . .	92
7.47. Matriz de covarianza empírica $\hat{S}(0)_{ij}$ y teórica $S(0)_{ij}$ de los clusters $C_n$ con 5 ROIs. . . . .	93
7.48. Error del modelo y bondad de ajuste para $C_n$ con 20 ROIs. . . . .	93
7.49. Clasificación MLR y 1-NN para $C_n$ con 20 ROIs . . . . .	94
8.1. Tasa de producción de entropía de los grupos $I_n$ , clusters $C_n$ y con los biomarcadores. . . . .	98



# 1

## Introducción.

El declive asociado al envejecimiento del cerebro completo en la relación entre estructura y función, es un enfoque que ha sido ampliamente estudiado desde la neurociencia. La tasa de producción de entropía  $\Phi$  permite caracterizar cómo la organización funcional y estructural del cerebro evoluciona con la edad. Estudios previos han analizado de manera independiente la función o la estructura cerebral, sin embargo aún persiste la necesidad de investigar el cerebro sano y completo, que integre la función y estructura desde los enfoques de la física y la matemática. En particular en el estado de reposo, desde un punto de vista clínico, el estudio del estado de reposo es crucial y tiene un enorme impacto y futuro, ya que los sujetos no necesitan comprender ni memorizar tareas cognitivas complejas que deban realizar, lo cual supone un reto para algunos de los trastornos cerebrales más comunes [7] o bien enfermedades neurodegenerativas.

Las enfermedades neurodegenerativas no se presentan de manera difusa ni aleatoria, sino que afectan redes específicas distribuidas en regiones de interés (ROIs). Las redes neuronales humanas pueden definirse por una actividad basal sincrónica, éstas muestran una susceptibilidad particular frente a las enfermedades neurodegenerativas [89]. En este contexto, la construcción de un modelo basal que integre la estructura y la función del cerebro en estado de reposo, basado en el proceso multivariado de Ornstein Uhlenbeck, puede constituir un punto de partida para comprender cómo se organizan estas redes en condiciones normales y cómo dicha organización puede modificarse con la edad. Asimismo, aún persisten interrogantes acerca de cuáles son las ROIs que cambian y como lo hacen a lo largo de la vida. En este sentido, futuros estudios podrían contribuir a esclarecer cómo se configuran estos sistemas a lo largo de la vida y de qué forma se ven alterados ante la aparición de enfermedades.

### 1.1. Objetivos

- Determinar la tasa de producción de entropía de 161 sujetos sanos agrupados de acuerdo a la edad biológica en  $I_1$  de 10-20 años,  $I_2$  de 20-40 años,  $I_3$  de 40-60 años e  $I_4$  de 60-80 años, utilizando la parcelación de módulos de estructura y función que toma 20 regiones de interés (ROIs). Con el fin de investigar la evolución de la dinámica cerebral fuera de equilibrio desde los 10 a los 80 años, aplicando tres clases de redes cerebrales, conectividad estructural, conectividad funcional y

conectividad efectiva.

- Determinar clusters y biomarcadores que permiten identificar patrones asociados al envejecimiento de sujetos sanos y en estado de reposo para evaluar la (ir)reversibilidad de la dinámica cerebral mediante la tasa de producción de entropía.

## 1.2. Hipótesis.

- La tasa de producción de entropía a nivel del cerebro completo muestra una tendencia creciente con la edad biológica, reflejando un incremento en la irreversibilidad y generando un sistema dinámico más alejado del equilibrio conforme la edad.
- La agrupación de sujetos sanos en estado de reposo puede basarse en las propiedades intrínsecas de la conectividad funcional y estructural, las cuales podrían mostrar una correspondencia con la edad cronológica o biológica.

## 1.3. Organización de la tesis.

La siguiente tesis se estructura de la siguiente manera.

El **Capítulo 1** corresponde a la introducción, en la cual se presentan el contexto y descripción general, la hipótesis y los objetivos de la presente investigación.

El **Capítulo 2** presenta los preliminares del proceso multivariado Ornstein Uhlenbeck, así como los elementos matemáticos y físicos necesarios para la comprensión de los estudios previos de los datos aplicados en esta tesis.

El **Capítulo 3** describe el proceso multivariado Ornstein Uhlenbeck, que constituye la base del modelo anatómico funcional utilizado en este trabajo.

El **Capítulo 4** aborda los conceptos de (ir)reversibilidad y tasa de producción de entropía, que constituyen el fundamento teórico para la interpretación de los resultados obtenidos mediante el modelo.

El **Capítulo 5** describe los datos estructurales y funcionales empleados en el modelo, así como el preprocesamiento de los mismos.

El **Capítulo 6** presenta tres enfoques aplicados a los mismos datos, cuyos resultados se utilizan como referencia comparativa en los resultados de esta tesis.

El **Capítulo 7** describe el modelo anatómico funcional, el cual se desarrolla en tres etapas metodológicas.

El **Capítulo 8** presenta los resultados obtenidos en cada una de las tres etapas, así como su comparación y contraste. Así como la discusión de los mismos.

Finalmente, el **Capítulo 9** expone las conclusiones de la tesis.

# 2

## Preliminares

En este capítulo se presentan definiciones y conceptos esenciales obtenidos de [12], [40], [41], [45] y [88] empleados en el desarrollo de la presente tesis.

### 2.1. Definiciones de probabilidad y procesos estocásticos.

Consideremos un experimento aleatorio asociado a un espacio de probabilidad.

**Definición 2.1.1** (Experimento aleatorio). *Es un experimento en el que al repetirse bajo las mismas condiciones el resultado que se obtiene no siempre es el mismo y tampoco es predecible.*

**Definición 2.1.2** (Espacio de probabilidad). *Es una terna  $(\Omega, F, P)$  en donde  $\Omega$  es un conjunto arbitrario no vacío, denominado espacio muestral que representa todos los posibles resultados de un experimento aleatorio, un elemento de  $\Omega$  se denota por  $\omega$ .  $F$  es una colección no vacía de subconjuntos de  $\Omega$ , denominada  $\sigma$ -álgebra,  $P$  es la función  $P : F \rightarrow [0, 1]$  es una medida de probabilidad. A la pareja  $(\Omega, F)$  se le llama espacio medible.*

**Definición 2.1.3** ( $\sigma$ -álgebra). *Una colección de  $F$  de subconjuntos de un espacio muestral  $\Omega$  es una  $\sigma$ -álgebra si cumple las siguientes tres condiciones:*

- $\Omega \in F$ .
- Si  $A \in F \Rightarrow A^c \in F$ .
- Si  $A_1, A_2, \dots \in F \Rightarrow \bigcup_{n=1}^{\infty} A_n \in F$ .

A los elementos de  $F$  se les denomina eventos o conjuntos medibles.

**Definición 2.1.4** (Variable aleatoria). *Es una transformación  $X$  del espacio de resultados  $\Omega$  al conjunto de números reales  $\mathbb{R}$ , esto es,*

$$X : \Omega \rightarrow \mathbb{R} \quad \forall x \in \mathbb{R} \quad \omega \in \{ \Omega : X(\omega) \leq x \in F \}$$

Decimos que una variable aleatoria  $X$  es discreta cuando el conjunto de valores que ésta toma es un conjunto discreto, es decir un conjunto finito o numerable. Por otro lado, una variable aleatoria  $X$  es continua, si toma todos los valores dentro de un intervalo  $(a, b) \subseteq \mathbb{R}$ .

**Definición 2.1.5** (Función de masa de probabilidad). *Sea  $X$  una variable aleatoria discreta con valores  $x_0, x_1, \dots$ . La función de masa de probabilidad de  $X$ , denotada por  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  se define como:*

$$f(x) = \begin{cases} P(X = x), & \text{si } x = x_0, x_1, \dots \\ 0 & \text{otro caso.} \end{cases}$$

**Definición 2.1.6** (Función de densidad). *Sea  $X$  una variable aleatoria continua. Decimos que la función integrable y no negativa  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$  es la función de densidad de  $X$  si para cualquier intervalo  $[a, b]$  de  $\mathbb{R}$  se cumple la igualdad.*

$$P(X \in [a, b]) = \int_a^b f(x) dx.$$

**Definición 2.1.7** (Función de distribución conjunta). *La función de distribución conjunta de dos variables aleatorias  $X$  e  $Y$ , denotada por  $F(x, y) : \mathbb{R}^2 \rightarrow [0, 1]$  para el caso discreto, se define como:*

$$F(x, y) = P(X \leq x, Y \leq y).$$

Para el caso continuo se define como:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du,$$

donde  $f(u, v)$  es una función no negativa.

**Definición 2.1.8** (Distribución Normal). *La variable aleatoria  $X$  tiene una distribución normal si su función de densidad está dada por la siguiente expresión:*

$$f(x) = \frac{1}{\sqrt{2\pi \text{Var}(X)}} e^{-\frac{(x - \mathbb{E}[x])^2}{2\text{Var}(X)}}, x \in \mathbb{R},$$

en donde  $\mathbb{E}[x]$  es la esperanza y  $\text{Var}(X)$  es la varianza.

**Definición 2.1.9** (Distribución normal estándar). *La variable aleatoria  $X$  tiene una distribución normal estándar si sigue una distribución normal con valor promedio cero y desviación estándar igual a uno, es decir,  $\mathbb{E}[x] = 0$  y  $\text{Var}(X) = 1$ , denotada como  $\mathcal{N}(0, 1)$ . En este caso la función de densidad se expresa como:*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}.$$

**Definición 2.1.10** (Estandarización). *Es el procedimiento de transformar una variable aleatoria normal no estándar en una estándar mediante la siguiente operación:*

$$A = \frac{X - \mathbb{E}[x]}{\text{Var}(X)} \sim \mathcal{N}(0, 1).$$

### 2.1.1. Momentos de una variable aleatoria

A continuación se describen cantidades numéricas asociadas a variables aleatorias. Para cada número natural  $n$ , se define el  $n$ -ésimo momento de una variable aleatoria  $X$  como el número  $\mathbb{E}[X^n]$ , suponiendo que tal momento exista.

**Definición 2.1.11** (Momentos de una variable aleatoria). *Los momentos de una variable aleatoria  $X$  son la colección de números  $\mathbb{E}(X)$ ,  $\mathbb{E}(X^2)$ ,  $\mathbb{E}(X^3)$ , ..., correspondientes al primer momento, segundo momento, ..., cuando tales cantidades existen. Para variables aleatorias continuas, el  $n$ -ésimo momento se calcula como sigue:*

$$\mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx.$$

El primer momento es el valor promedio o esperanza de  $X$ , el segundo momento está relacionado con la dispersión de  $X$  y la varianza es el tercer momento que está relacionado con la simetría de la correspondiente distribución de probabilidad [41].

**Definición 2.1.12** (Esperanza). *Sea  $X$  una variable aleatoria continua con función de densidad  $f(x)$ . La esperanza de  $X$  denotada por  $\mathbb{E}(X)$  se define como:*

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

*cuando esta integral sea absolutamente convergente, es decir cuando la suma de los valores absolutos es convergente y en tal caso se dice que  $X$  es integrable o que tiene esperanza finita.*

**Definición 2.1.13** (Varianza). *Sea  $X$  una variable aleatoria continua con función de densidad  $f(x)$ . La varianza de  $X$  se define como:*

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) dx,$$

*cuando esta integral es convergente.*

**Definición 2.1.14** (Covarianza). *Sean  $X$  e  $Y$  dos variables aleatorias cualesquiera definidas en el mismo espacio de probabilidad. La covarianza de  $X$  e  $Y$  se define como*

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) (Y - \mathbb{E}[Y])],$$

*siempre que existan  $\mathbb{E}(X)$  y  $\mathbb{E}(Y)$ .*

**Definición 2.1.15** (Covarianza de varias variables). *En caso de variables variables, la matriz de covarianza se define como:*

$$\mathbb{E}[X_i, X_j] \equiv \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] \equiv \mathbb{E}[X_i X_j] - \mathbb{E}(X_i) \mathbb{E}(X_j).$$

*Si las variables son independientes en pares, la matriz de covarianza es diagonal.*

**Definición 2.1.16** (Proceso estocástico). *Un proceso estocástico es una colección de variables aleatorias  $\{X_t : t \in T\}$  parametrizadas por un conjunto  $T$ , usualmente interpretado como un conjunto de tiempos y llamado espacio parametral. Por lo tanto un proceso estocástico es una función de dos variables*

$$X : [0, \infty] \times \Omega \rightarrow \mathbb{R}, \quad \forall t \geq 0 \quad \omega \rightarrow X_t(\omega), \quad (2.1)$$

donde la función  $\omega \rightarrow X_t(\omega)$  es una variable aleatoria mientras que  $\forall \omega \in \Omega$ , la función  $t \rightarrow X_t(\omega)$  es una trayectoria del proceso.

**Definición 2.1.17** (Proceso estocástico continuo). *Es una colección de variables aleatorias  $\{X_t : t \in T\}$  parametrizadas por un conjunto  $T$  en donde las variables toman valores en un conjunto  $\Omega$ . El espacio parametral puede tomarse como un conjunto continuo  $T = [0, \infty)$  se denota como  $\{X_t : t \geq 0\}$ .*

**Definición 2.1.18** (Proceso estacionario). *Un proceso estocástico a tiempo continuo  $\{X_t : t \geq 0\}$  es estacionario en el sentido estricto si para cualquier conjunto de tiempos  $t_1, \dots, t_n$  la distribución del vector  $(X_{t_1+h}, \dots, X_{t_n+h})$  para cualquier valor de  $h > 0$ , la distribución de  $X_t$  es la misma que  $X_{t+h}$ .*

**Definición 2.1.19** (Proceso estacionario reversible.). *Un proceso estocástico estacionario  $\{x(t); t \in \mathbb{R}\}$  es reversible en el tiempo si para cada entero positivo  $m$  y cada  $t_1, t_2, \dots, t_m \in \mathbb{R}$  la probabilidad conjunta está dada por:  $P(x(t_1), x(t_2), \dots, x(t_m)) = P(x(-t_1), x(-t_2), \dots, x(-t_m))$ .*

**Definición 2.1.20** (Proceso Gaussiano). *Un proceso estocástico a tiempo continuo  $\{X_t : t \geq 0\}$  es un proceso Gaussiano si para cualquiera colección finita de tiempos  $t_1, \dots, t_n$  el vector  $X_{t_1}, \dots, X_{t_n}$  tiene una distribución normal o Gaussiana multivariada.*

**Definición 2.1.21** (Distribución  $t$ ). *Decimos que la variable aleatoria continua  $X$  tiene una distribución  $t$  con  $n > 0$  grados de libertad si su función de densidad  $f(x)$  está dada por la siguiente expresión.*

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}, \quad (2.2)$$

donde  $\Gamma$  es la función gamma.

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \quad z > 0.$$

Para enteros positivos  $k$  :  $\Gamma(k) = (k-1)!$ .

## 2.1.2. Movimiento Browniano

El objeto matemático que llamamos movimiento browniano es un proceso estocástico de tiempo continuo y estado continuo, también llamado proceso de Wiener, en honor al matemático estadounidense Norbert Wiener. Un proceso estocástico de tiempo continuo  $(\mathcal{B}_t)_{t \geq 0}$  es un movimiento browniano estándar, si cumple con las siguientes propiedades [67].

1.  $\mathcal{B}_0 = 0$ .
2. Distribución normal. Para  $t \geq 0$ ,  $\mathcal{B}_t$  tiene una distribución normal con media 0 y varianza  $t$ .
3. Incrementos estacionarios. Para  $s, t \geq 0$ ,  $\mathcal{B}_{t+s} - \mathcal{B}_s$  tiene la misma distribución que  $\mathcal{B}_t$ . Es decir,

$$P(\mathcal{B}_{t+s} - \mathcal{B}_s \leq z) = P(\mathcal{B}_t \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx,$$

para  $-\infty < x < \infty$ .

4. Incrementos independientes. Si  $0 \leq q < r \leq s < t$ , entonces  $\mathcal{B}_t - \mathcal{B}_s$  y  $\mathcal{B}_r - \mathcal{B}_q$  son variables aleatorias independientes.
5. Caminos continuos. La función  $t \mapsto \mathcal{B}_t$  es continuo, con probabilidad 1.

### 2.1.3. Ecuación diferencial estocástica.

Sea  $(\Omega, \mathcal{F}, P)$  un espacio de probabilidad y sea  $\{\mathcal{B}_t : t \geq 0\}$  un movimiento Browniano unidimensional adaptado a la filtración  $(\mathcal{F}_t)_{t \geq 0}$ .

**Definición 2.1.22** (Filtración).  $\{\mathcal{F}_t\}_{t \geq 0}$  es una colección de  $\sigma$ -álgebras tal que  $\mathcal{F}_n \subseteq \mathcal{F}_m$  cuando  $n \leq m$ . En particular, la filtración natural o canónica de un proceso  $\{X_n : n \geq 1\}$  es aquella sucesión de  $\sigma$ -álgebras definida por:

$$\mathcal{F}_n = \sigma\{X_1, \dots, X_n\}, \quad n \geq 1.$$

**Definición 2.1.23** (Ecuación diferencial estocástica.). Sean  $b(t, x)$  y  $\zeta(t, x)$  dos funciones de  $[0, T] \times \mathbb{R} \in \mathbb{R}$ . Una ecuación estocástica es una ecuación de la forma:

$$dX_t = b(t, X_t)dt + \zeta(t, X_t)dB_t, \quad (2.3)$$

definida para valores de  $t$  en el intervalo  $[0, T]$  y con condición inicial la variable aleatoria  $X_0$  que se presupone  $\mathcal{F}_0$ -medible e independiente del movimiento Browniano. El coeficiente  $b(t, x)$  se conoce como coeficientes de tendencia (drift en inglés o deriva en español) y el coeficiente  $\zeta(t, x)$  se denomina difusión.

### 2.1.4. Proceso de Itô

**Definición 2.1.24** (Proceso de Itô). La solución a la ecuación (2.3) se interpreta como la integral

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t \zeta(s, X_s)dB_s. \quad (2.4)$$

La primera integral es una integral de Riemann, mientras que la segunda es una integral estocástica de Itô. Al proceso  $\{X_t\}$  se le llama proceso de Itô.

La solución de (2.4) se interpreta como el estado de un sistema que evoluciona de manera determinista gobernado por la parte no aleatoria de la ecuación (la tendencia o deriva), pero alterado por un ruido aditivo dado por la integral estocástica (la difusión) [42].

**Definición 2.1.25** (Fórmula de Proceso de Itô). Si  $\{X_t\}$  es un proceso de Itô dado por (2.4) y  $f(t, x)$  es una función de clase  $C^1$  en  $t$  y de clase  $C^2$  en  $x$ , entonces el proceso  $\{Y_t = f(t, X_t)\}$  es también un proceso de Itô y satisface la ecuación estocástica:

$$dY_t = f_t(t, X_t) dt + f_x(t, X_t) dX_t + \frac{1}{2} f_{xx}(t, X_t) (dX_t)^2. \quad (2.5)$$

### 2.1.5. Ecuaciones de Langevin

**Definición 2.1.26** (Ecuación de Langevin en una dimensión). La ecuación de Langevin es una ecuación diferencial estocástica de primer orden. Para una variable escalar  $X_t \in \mathbb{R}$  tenemos,

$$dX_t = f(t, X_t)dt + g(t, X_t)dW(t), \quad (2.6)$$

donde  $f(t, X_t) \in \mathbb{R}$  es la deriva, representa la fuerza sistemática o tendencia del sistema,  $g(t, X_t) \in \mathbb{R}$  es la difusión, que determina la magnitud de las fluctuaciones aleatorias y  $W(t)$  es un proceso de Wiener unidimensional.

**Definición 2.1.27** (Ecuación de Langevin multivariada). Para un sistema de dimensión  $N$ , con  $X_t \in \mathbb{R}^N$ , la ecuación de Langevin se generaliza como:

$$dX_t = f(t, X_t) + g(t, X_t)dW(t), \quad (2.7)$$

donde  $f(t, X_t) \in \mathbb{R}^N$  es el vector de deriva,  $g(t, X_t) \in \mathbb{R}^{N \times N}$  y  $W(t) := \eta(t) \in \mathbb{R}^N$  es denominado vector de ruido blanco,  $dW(t) := \eta(t)dt$  es un proceso multivariado de Wiener, fuertemente fluctuante con media cero y una autocorrelación de:

$$\mathbb{E}[\eta_i(t)\eta_j(t')]_x = \delta_{ij}(t - t'), \quad (2.8)$$

donde  $\delta_{ij}$  se denomina delta de Kronecker,  $\delta_{ij}(t - t')$  es  $\delta$ -distribución,  $\mathbb{E}[\cdot]_x$  es el promedio sobre  $\Omega$ .

### 2.1.6. Proceso de Wiener

El proceso de Wiener es definido por la ecuación de Langevin.

$$dW(t) = \eta(t) \rightarrow W(t) = w_0 + \int_{t_0}^t \eta(t)dt, \quad (2.9)$$

donde  $\eta(t)$  ruido blanco. El proceso de Wiener  $W(t)$  es un proceso aleatorio continuo.

### 2.1.7. Ecuación de Fokker-Planck

Las ecuaciones de Fokker-Planck describen la evolución de los sistemas estocásticos. La ecuación general unidimensional

$$\frac{\partial P(x, t)}{\partial t} = - \frac{\partial}{\partial x} [B(x, t) P(x, t)] + \frac{\partial^2}{\partial x^2} [D(x, t) P(x, t)] \quad (2.10)$$

donde  $B(x, t)$  termino de deriva y  $D(x, t)$  es el termino de difusión y  $P(x, t)$  es la función de densidad.

### 2.1.8. Entropía

Es una medida de incertidumbre de una variable aleatoria discreta.

**Definición 2.1.28** (Entropía de una variable aleatoria). *La entropía de una variable aleatoria  $X$  con función de masa de probabilidad  $p(x)$  se define como:*

$$H(X) = - \sum_x p(x) \log_2 p(x). \quad (2.11)$$

La entropía expresada en bits se obtiene con  $\log_2(x)$ , donde  $x$  representa un estado posible de la variable aleatoria  $X$  [61]. Si la base de log es  $e$  entonces la entropía se expresa en nats [45].

**Definición 2.1.29** (Entropía conjunta). *La entropía conjunta  $H(X, Y)$  de un par de variables aleatorias discretas  $(X, Y)$  con una función de distribución de masa conjunta  $p(x, y)$  se define como:*

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log_2 p(x, y). \quad (2.12)$$

**Definición 2.1.30** (Entropía condicional). *Si  $(X, Y) \sim p(x, y)$ , entonces la entropía condicional  $H(X, Y)$  se define como:*

$$H(Y|X) = - \sum_x \sum_y \log_2 p(Y|X). \quad (2.13)$$

La naturalidad de la definición de entropía conjunta y entropía condicional se demuestra por el hecho de que la entropía de un par de variables aleatorias es la entropía de una más la entropía condicional de la otra [45].

**Definición 2.1.31** (Información mutua). *Consideremos dos variables aleatorias  $X$  e  $Y$  con función de masa de probabilidad conjunta  $p(x, y)$  y funciones de masa de probabilidad marginal  $p(x)$  y  $p(y)$ . La información mutua  $I(X; Y)$  se define como:*

$$I(X; Y) = \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}. \quad (2.14)$$

Relación entre entropía e información mutua.

$$\begin{aligned} I(X; Y) &= \sum_x \sum_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x, y) \log_2 \frac{p(x|y)}{p(x)} \\ &= - \sum_{x,y} p(x, y) \log_2 p(x) + \sum_{x,y} p(x, y) \log_2 p(x|y) \\ &= - \sum_x p(x) \log_2 p(x) - \left( - \sum_{x,y} p(x, y) \log_2 p(x|y) \right) \\ &= H(X) - H(X|Y). \end{aligned} \quad (2.15)$$

**Definición 2.1.32** (Cópula). *Una cópula es una función de distribución multivariante con marginales uniformes en  $[0, 1]$ . Toda función de distribución conjunta  $F_{X,Y}$  de dos variables aleatorias puede representarse en términos de sus distribuciones marginales  $F_X$ ,  $F_Y$  y una cópula  $C$  tal que:*

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y)), \quad (2.16)$$

*de esta manera, la cópula captura exclusivamente la dependencia entre las variables, independientemente de sus distribuciones marginales.*

**Definición 2.1.33** (Correlación de Pearson). *Sea  $X$  e  $Y$  dos variables aleatorias, el coeficiente de correlación de Pearson se define como:*

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (2.17)$$

*donde  $\text{Cov}(X, Y)$  es la covarianza entre  $X$  e  $Y$  y  $\sigma$  es la desviación estándar.*

# 3

## Proceso Multivariado Ornstein Uhlenbeck

El modelo fue propuesto en 1930 por los físicos L. S. Ornstein y G. E. Uhlenbeck para estudiar el comportamiento de los gases. Su análisis se basó en la teoría cinética de los gases, la cual establece que las moléculas se encuentran en constante movimiento y que la velocidad promedio del conjunto de partículas está determinada por las leyes de la presión y la temperatura. Ornstein y Uhlenbeck argumentaron que una molécula individual podía acelerar o desacelerar, no obstante para mantener la coherencia con la velocidad promedio constante, determinaron que la diferencia entre la velocidad de la partícula individual y la media debía tender a revertirse hacia cero [63].

Los sistemas lineales de Langevin, son sistemas con estados estacionarios de no equilibrio éstos pueden considerarse extensiones de mayor dimensión del proceso Ornstein Uhlenbeck, que describe la velocidad de una partícula Browniana en una dimensión [10], la cual se escribe en su forma unidimensional como:

$$m\dot{v}(t) = -\gamma v(t) + \eta(t), \quad (3.1)$$

donde  $\gamma$  es el coeficiente de fricción y  $\eta(t)$  es el ruido blanco de Gauss con amplitud  $\sigma$ , i.e :

$$\mathbb{E}[\eta(t)\eta(t')] = 2\sigma^2\delta(t - t').$$

El proceso Ornstein Uhlenbeck es reversible y satisface el principio de balance detallado con respecto a la distribución de probabilidad de equilibrio de la velocidad. Dicha distribución corresponde a la distribución de Maxwell-Boltzman, es decir, una ley de Gauss con varianza.

$$\mathbb{E}(v^2) = \frac{\sigma^2}{m\gamma} = \frac{kT}{m}, \quad (3.2)$$

donde  $\gamma$  es el coeficiente de fricción  $m$  masa,  $k$  es la constante de Boltzman y  $T$  temperatura. La relación de Einstein  $\sigma^2 = kT\gamma$ , es un ejemplo de una relación de fluctuación-disipación de equilibrio [10]. El trabajo propuesto por Ornstein y Uhlenbeck se emplea para modelar la variación de la velocidad en el movimiento difuso de una partícula para tiempos pequeños [40].

### 3.1. Ornstein Uhlenbeck OU

El proceso de Ornstein Uhlenbeck (OU), puede describirse de la siguiente forma:

$$dX_t = -\alpha X_t dt + \sigma dW_t, \quad X_0 = x_0. \quad (3.3)$$

donde  $X_t$  representa la velocidad de la partícula al tiempo  $t$ , además  $\alpha$  y  $\sigma$  son dos constantes positivas,  $-\alpha X_t$  es el término determinista o de deriva, que corresponde a la fuerza de fricción y  $\sigma dW_t$  es el término de difusión, que corresponde al ruido aleatorio. El término de deriva  $-\alpha X_t dt$  es negativo cuando  $X_t > 0$  y es positivo cuando  $X_t < 0$ , por lo que, aunque el proceso nunca estará libre de fluctuaciones aleatorias, se puede esperar que  $X_t$  vuelva a cero cada vez que se haya alejado [63]. El proceso muestra un decaimiento conforme el tiempo avanza y ello debido al factor de fricción del modelo [40]. Se utiliza la notación estándar de Itô para el proceso OU.

#### 3.1.1. Solución al proceso OU

Consideremos la siguiente solución al proceso Ornstein Uhlenbeck (3.3) con una condición inicial  $x_0$ :

$$X_t = a(t) \left\{ x_0 + \int_0^t b(s) dW_s \right\}, \quad (3.4)$$

donde  $a(t)$  y  $b(s)$  son funciones diferenciables [63]. Derivando (3.4) y usando la fórmula de Itô para el producto de dos procesos estocásticos, tenemos que  $X_t = U_t V_t$  con  $U_t$  y  $V_t$ :

$$U_t = a(t), \quad dU_t = a'(t) dt, \quad V_t = x_0 + \int_0^t b(s) dW_s, \quad dV_t = b(t) dW_t.$$

Aplicando la regla del producto de Itô

$$dX_t = U_t dV_t + V_t dU_t + d[U, V]_t,$$

donde  $d[U, V]_t$  es el diferencial del proceso de covarianza cuadrática. Dado que  $dt dW_t = 0$  luego  $d[U, V]_t = 0$ . Se sigue que:

$$dX_t = a'(t) \left\{ x_0 + \int_0^t b(s) dW_s \right\} dt + a(t) b(t) dW_t. \quad (3.5)$$

Asumimos que  $a(0) = 1$ , y  $a(t) > 0$ ,  $\forall t \geq 0$  entonces el proceso definido por (3.4) es una solución de una ecuación diferencial estocástica (EDE):

$$dX_t = \frac{a'(t)}{a(t)} X_t dt + a(t) b(t) dW_t, \quad \text{con } X_0 = x_0. \quad (3.6)$$

La relación entre las ecuaciones (3.4) y (3.6) nos da una solución sencilla para una EDE, con un coeficiente de deriva lineal en  $X_t$  y el coeficiente de difusión que dependa sólo de  $t$ . Para que los términos de la ecuación (3.6) coincidan con los de la ecuación (3.3) se satisfacen las siguientes ecuaciones:

$$\frac{a'(t)}{a(t)} = -\alpha \quad a(t) b(t) = \sigma,$$

$$a(0) = 1 \rightarrow a(t) = e^{-\alpha t} \quad \text{y} \quad b(t) = \sigma e^{\alpha t}.$$

La solución al proceso OU 3.3 está dada por :

$$X_t = e^{-\alpha t} \left\{ x_0 + \sigma \int_0^t e^{\alpha s} dW_s \right\}.$$

$$\boxed{X_t = x_0 e^{-\alpha t} + \sigma \int_0^t e^{-\alpha(t-s)} dW_s.} \quad (3.7)$$

Se modifica la notación para el término de difusión  $dB_t$  en [63] por  $dW_t$ , para el proceso OU.

El proceso Ornstein Uhlenbeck tiene las siguiente propiedades:

- a) Esperanza  $\mathbb{E}(X_t) = x_0 e^{-\alpha t}$

En la solución del proceso OU (3.7), a medida que  $t \rightarrow \infty$ , la influencia del valor inicial  $X_0 = x_0$  decae exponencialmente, luego tenemos que  $\mathbb{E}(X_t) = x_0 e^{-\alpha t}$ . Dado que la integral de  $dW_s$  tiene media cero la media de  $\mathbb{E}(X_t)$  tiende a cero rápidamente [63]:

$$\mathbb{E}(X_t) = x_0 e^{-\alpha t}. \quad (3.8)$$

- b) La varianza:

$$\text{Var}(X_t) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}).$$

$$\begin{aligned} \text{Var}(X_t) &= \sigma^2 \text{Var} \left( \int_0^t e^{-\alpha(t-s)} dW_s \right) \\ &= \sigma^2 \mathbb{E} \left( \int_0^t e^{-\alpha(t-s)} dW_s \right)^2 \\ &= \sigma^2 \left( \int_0^t e^{-2\alpha(t-s)} ds \right) \\ &= \sigma^2 e^{-2\alpha t} \left[ \frac{1}{2\alpha} e^{2\alpha s} \right]_0^t \\ &= \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t}). \end{aligned} \quad (3.9)$$

- c) La covarianza:

$$\text{Cov}(X_t, X_s) = \frac{\sigma^2}{2\alpha} (e^{-\alpha(t-s)} - e^{-\alpha(t+s)})$$

$$\begin{aligned} \text{Cov}(X_t, X_s) &= \mathbb{E}(X_t, X_s) - \mathbb{E}(X_t) \mathbb{E}(X_s) \\ &= \mathbb{E} \left[ \left( x_0 e^{-\alpha t} + \sigma \int_0^t e^{-\alpha(t-u)} dW_u \right) \left( x_0 e^{-\alpha s} + \sigma \int_0^s e^{-\alpha(s-u)} dW_u \right) \right] - x_0^2 e^{-\alpha(t+s)} \\ &= \sigma^2 \mathbb{E} \left( \int_0^t e^{-\alpha(t-u)} dW_u \int_0^s e^{-\alpha(s-u)} dW_u \right). \end{aligned}$$

La primera integral puede descomponerse en la suma de dos integrales, una sobre el intervalos  $[0, s]$  y otra sobre  $(s, t]$ . Dada la propiedad de incrementos independientes del movimiento Browniano, el segundo sumando desaparece [42].

$$\begin{aligned} &= \sigma^2 e^{-\alpha(t+s)} \mathbb{E} \left( \int_0^s e^{\alpha u} dW_u \right)^2 \\ &= \sigma^2 e^{-\alpha(t+s)} \int_0^s e^{2\alpha u} du \\ &= \frac{\sigma^2}{2\alpha} (e^{-\alpha(t-s)} - e^{-\alpha(t+s)}). \end{aligned} \tag{3.10}$$

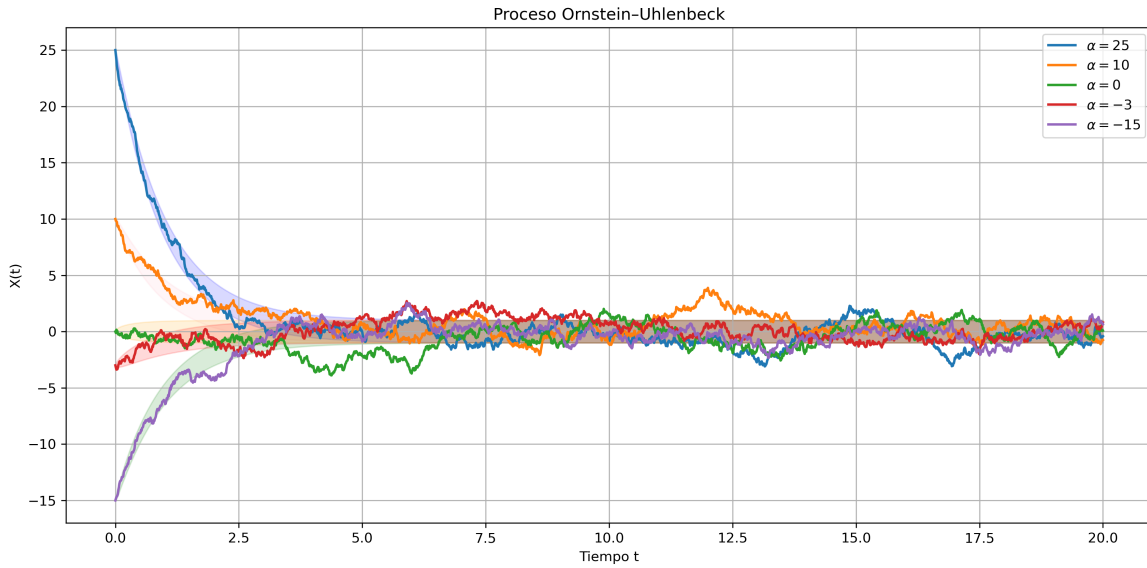


Figura 3.1: De acuerdo a la ecuación (3.3), donde:  $X_t$  es el estado del proceso,  $\alpha \neq 0$  es el retorno a la media,  $\sigma > 0$  es la intensidad del ruido y  $W_t$  es el proceso de Wiener.

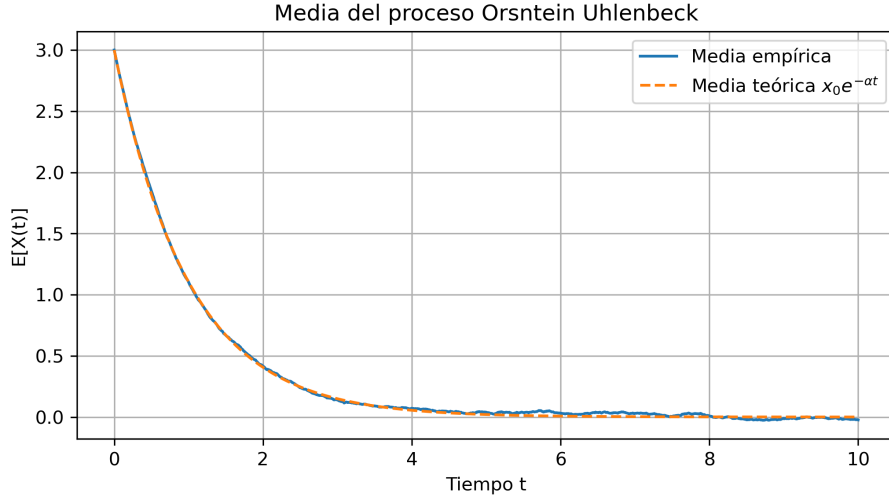


Figura 3.2: De acuerdo a la ecuación (3.8), donde:  $X_0 = 3$  y  $\alpha = 1$ .

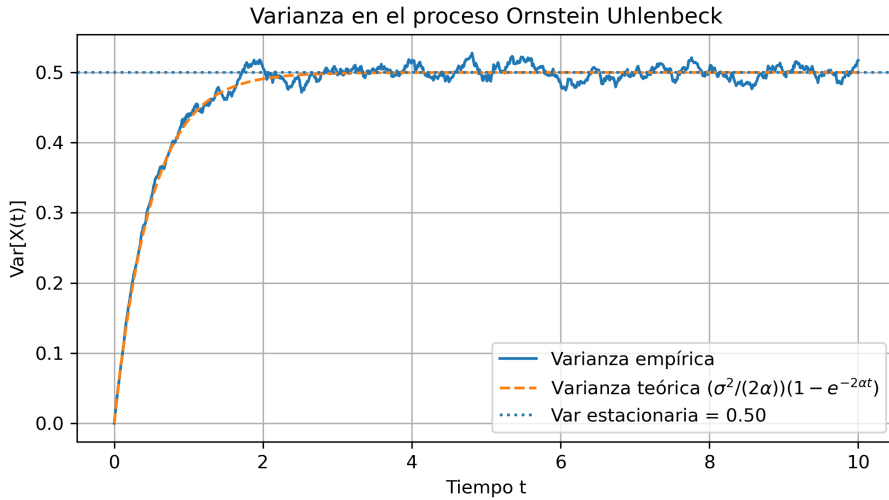


Figura 3.3: De acuerdo a la ecuación (3.9) donde:  $\alpha = 1$  y  $\sigma = 1$ , la varianza inicia en 0 va aumentando en  $t$  debido a la difusión.

## 3.2. Ornstein Uhlenbeck Multivariado

El proceso multivariado de Ornstein Uhlenbeck (MOU) es un proceso de difusión definido por  $N$  ecuaciones de Langevin que describen la evolución de variables dinámicas acopladas. De manera general para  $N$  ecuaciones de Langevin lineales acopladas se expresa como:

$$\dot{x}_m(t) = - \sum_n B_{mn} x_n(t) + \eta_m(t), \quad \forall m, n = 1, \dots, N, \quad (3.11)$$

donde  $\eta_m(t)$  es el vector de ruido blanco Gaussiano, tal que:

$$\mathbb{E}[\eta_m(t)\eta_m(t')] = 2D_{mn}\delta(t-t').$$

$$i.e. \quad \mathbb{E}[\eta(t)\eta^T(t')] = 2D\delta(t-t'). \quad (3.12)$$

El proceso MOU se define mediante dos matrices reales  $D$  y  $B \in \mathbb{R}^{N \times N}$ . La matriz  $D$ , denominada como matriz de covarianza o de difusión, es simétrica por construcción de acuerdo a la ecuación (3.12). En [10] se asume además que  $D$  es positiva definida, de lo cual se deduce que es no singular.

$D \in \mathbb{R}^{N \times N}.$ <i>(i)</i> $D = D^T,$ simetría, <i>(ii)</i> $x^T D x > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\},$ $D$ definida positiva, Si $D$ cumple <i>(i)</i> e <i>(ii)</i> $\Rightarrow \det(D) > 0 \Rightarrow D$ es no singular e invertible.	(3.13)
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------

La matriz de fricción  $B$  es simétrica en general, y no cumple con la condición de Hurwitz, es decir, todos sus valores propios  $B_k$  en general son complejos [10].

$B \in \mathbb{R}^{N \times N}.$ <i>(i)</i> $B = B^T,$ $B_k \Rightarrow Re(\tilde{B}_k) > 0, (k = 1 \dots N).$	(3.14)
----------------------------------------------------------------------------------------------------------------------	--------

De acuerdo a (3.13) y (3.14) se garantiza que el proceso cae exponencialmente hasta  $x = 0$  en ausencia de ruido, y hasta un estado estacionario fluctuante con estadística gaussiana en presencia de ruido [10]. En notación vectorial y matricial se expresa como:

$$\dot{x}(t) = -B x(t) + \eta(t), \quad (3.15)$$

$$D, B \in \mathbb{R}^{N \times N}, \quad x(t) \in \mathbb{R}^N.$$

$$B = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix}, \quad x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \quad \eta(t) = \begin{bmatrix} \eta_{11} & \dots & \eta_{1m} \\ \vdots & \ddots & \vdots \\ \eta_{n1} & \dots & \eta_{nm} \end{bmatrix}.$$

### 3.2.1. Solución al proceso MOU

La solución para el proceso MOU (3.15) está dada de la siguiente manera. Consideremos las matrices constantes  $B$  y  $D$  y el término del ruido  $\eta(t) = dW(t)$ . Aplicando la regla del producto de Itô, definimos a  $y(t) = x(t)e^{Bt}$  donde  $x(t)$  es un proceso estocástico y  $e^{Bt}$  es un proceso determinista con  $B$  una matriz constante.

$$\begin{aligned}\dot{y}(t) &= e^{Bt}\dot{x}(t) + Bx(t)e^{Bt} + 0 \\ &= e^{Bt}[-B(x) + D dW(t)] + Bx(t)e^{Bt} \\ &= -Bx(t)e^{Bt} + De^{Bt}dW(t) + Bx(t)e^{Bt} \\ &= De^{Bt}dW(t).\end{aligned}$$

$$y(t) = x(t)e^{Bt}$$

$$x(t)e^{Bt} = \int_0^t De^{Bs}dW(s)$$

$$x(t) = x(0)e^{-Bt} + D \int_0^t e^{-B(t-s)}dW(s).$$

$$\boxed{x(t) = x(0)e^{-Bt} + D \int_0^t e^{-B(t-s)}dW(s).} \quad (3.16)$$

### 3.2.2. Evolución del estado estacionario del proceso MOU

A partir de un valor inicial determinista conocido  $x(0)$ , la evolución del proceso MOU (3.15) admite una solución analítica que pueda obtenerse mediante la función de Green [10], dicha solución satisface la ecuación diferencial 3.18.

$$G(t) = e^{-Bt}, \quad (3.17)$$

$$\dot{G}(t) = -BG(t) = -G(t)B, \quad (3.18)$$

con  $G(0) = 1$ . La hipótesis planteada sobre la matriz de fricción  $B$  asegura que  $G(t)$  disminuye exponencialmente hasta cero, por lo que la solución del proceso se expresa como:

$$x(t) = G(t)x(0) + \int_0^t G(t-s)\eta(s)ds. \quad (3.19)$$

El proceso  $\dot{x}(t)$  es de naturaleza Gaussiana, por lo que puede expresarse en términos lineales del ruido blanco Gaussiano. Sus propiedades estadísticas en el instante  $t$  quedan determinadas por  $\mathbb{E}[x(t)] = G(t)x(0)$ , y por su matriz de correlación de tiempos iguales [10].

### 3.2.3. Covarianza sin desfase $t = t'$

La covarianza sin desfase no sigue la regla determinista dado que el producto de incrementos estocásticos es del orden  $dt$ , siguiendo la regla de Itô que establece que:

$$dW dW^T = I dt,$$

se sigue que la covarianza sin desfase está definida como:

$$S(t, t) = \mathbb{E}[x(t) x^T(t)]. \quad (3.20)$$

Derivamos la ecuación (3.20) y por la fórmula de Itô para el producto tenemos que:

$$\begin{aligned} \dot{S}(t, t) &= \dot{\mathbb{E}}[xx^T] \\ &= \mathbb{E}[\dot{xx}^T + x\dot{x}^T + d[\dot{xx}^T]] \\ &= \mathbb{E}[(-Bx + \eta)x^T + x(-Bx + \eta)^T + d[(-Bx + \eta)(-Bx + \eta)^T]] \\ &= \mathbb{E}[-Bxx^T + \eta x^T - xx^T B^T + x\eta^T + [BB^T xx^T - Bx\eta^T - B^T x^T \eta + \eta\eta^T] dt] \end{aligned}$$

$$\text{Si } \mathbb{E}[x\eta^T] = \mathbb{E}[\eta x^T] = 0 \text{ entonces } \mathbb{E}[-Bxx^T - xx^T B^T + \eta\eta^T].$$

Por linealidad de  $\mathbb{E}$ :

$$\mathbb{E}[-Bxx^T] - \mathbb{E}[xx^T B^T] + \mathbb{E}[\eta\eta^T] = -B\mathbb{E}[xx^T] - B^T\mathbb{E}[xx^T] + \mathbb{E}[\eta\eta^T].$$

De acuerdo a la ecuación (3.18) para  $S(t, t)$  y (3.12) para la matriz  $D$ , tenemos que:

$$\dot{S}(t, t) = -BS(t, t) - S(t, t)B^T + 2D.$$

$$\boxed{\dot{S}(t, t) = -BS(t, t) - S(t, t)B^T + 2D.} \quad (3.21)$$

Para  $\lim_{t \rightarrow \infty} \dot{S}(t, t) = 0$ , entonces:

$$\boxed{BS(0) + S(0)B^T = 2D.} \quad (3.22)$$

La ecuación (3.21) relaciona la matriz de covarianza estacionaria  $S$  con las matrices de fricción y difusión  $B$  y  $D$  que definen el proceso. Las ecuaciones matriciales lineales de este tipo se denominan ecuaciones de Sylvester (o de Lyapunov) [10]. Relacionando la ecuación (3.19) con la ecuación (3.20) tenemos que:

$$\begin{aligned} S(t) &= \mathbb{E} \left[ \left( G(t)x(0) + \int_0^t G(t-s) \eta(s) ds \right) \left( G(t)x(0) + \int_0^t G(t-u) \eta(u) du \right)^T \right] \\ &= G(t) \mathbb{E}[x(0)x(0)^T] G^T(t) + \mathbb{E} \left[ \int_0^t \int_0^t G(t-s) \eta(s) \eta(u)^T G^T(t-u) ds du \right]. \end{aligned}$$

Los términos cruzados se anulan debido a  $\mathbb{E}[\eta(t)] = 0$ .

Si  $S(0) = \mathbb{E}[x(0)x(0)^T]$ ,  $\mathbb{E}[\eta(s)\eta(u)^T] = 2D \delta(s-u)$  entonces

$$S(t) = G(t) S(0) G^T(t) + 2 \int_0^t G(t-s) D G^T(t-s) ds.$$

El estado estacionario del proceso es, por tanto, gaussiano, con media cero, y se caracteriza por la matriz de covarianza estacionaria [10].

$$S = \lim_{t \rightarrow \infty} S(t) = 2 \int_0^{\infty} G(t) D G^T(t) dt. \quad (3.23)$$

### 3.2.4. Covarianza con desfase $t' > t$

Determinamos la covarianza con desfase como la evolución de  $x(t')$  a  $x(t)$ . A partir de la ecuación (3.19) con  $t' > t$  y la ecuación (3.20) [21].

$$\begin{aligned} S(t, t') &= \mathbb{E} \left[ x(t) \left( G(t' - t)^T x^T(t) + \int_t^{t'} G(t' - s) \eta(s) ds \right)^T \right], \\ &= \mathbb{E} [x(t) G(t' - t)^T x^T(t)] + \mathbb{E} \left[ \int_t^{t'} G(t' - s) \eta(s) ds \right]^T. \end{aligned}$$

Si  $\mathbb{E} \left[ \left( \int_t^{t'} G(t' - s) \eta(s) ds \right)^T \right] = 0$  entonces  $\mathbb{E} [x(t) G(t' - t)^T x^T(t)]$ ,

$$\begin{aligned} S(t, t') &= G(t' - t)^T \mathbb{E} [x(t) x^T(t)] \\ &= \text{Por (3.17) y (3.20) con } t' > t \\ &= S(t, t) e^{-B^T(t'-t)}. \end{aligned}$$

$$\boxed{S(t, t') = S(t, t) e^{-B^T(t'-t)}}. \quad (3.24)$$

## 3.3. Distribución de probabilidad

Las ecuaciones de Fokker-Planck describen la evolución de sistemas estocásticos desde el punto de vista de la evolución de la densidad de probabilidad. En general, se aplican a sistemas en equilibrio y en desequilibrio [12]. La evolución de densidad de probabilidad  $P(x, t)$  del proceso MOU (3.15) se describe mediante la siguiente ecuación de Fokker-Planck [21].

$$\frac{\partial P(x, t)}{\partial t} = \nabla [Bx(t)P(x, t) + D \nabla P(x, t)]. \quad (3.25)$$

donde  $\nabla$  denota la derivada parcial respecto de  $x$ . La ecuación (3.25) puede reescribirse como una ecuación de continuidad de la forma:

$$\frac{\partial P(x, t)}{\partial t} + \nabla \cdot J(x, t) = 0. \quad (3.26)$$

donde  $J(x, t)$  es el flujo o corriente de probabilidad.

$$J(x, t) = -D \nabla P(x, t) - Bx(t)P(x, t). \quad (3.27)$$

### 3.3.1. Estado estacionario y corriente de probabilidad

La propiedad de Gauss-Markov del proceso (3.15) garantiza que tanto la media como las covarianzas convergen exponencialmente hacia sus respectivos puntos fijos, siempre que los valores propios de la matriz  $B$  (3.14) (que pueden ser complejos) tengan una parte real positiva. El estado estacionario del proceso MOU exhibe fluctuaciones gaussianas alrededor de una media nula, lo que corresponde a la densidad de probabilidad multivariante independiente del tiempo [21].

$$P(x) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det S}} e^{-\frac{1}{2}x^T S^{-1}x}, \quad (3.28)$$

donde  $S$  denota el punto fijo de la matriz de covarianza de retardo cero con la ecuación (3.22)  $S(t, t)$ . El gradiente de la ecuación (3.28) se determina de la siguiente manera:

$$P(x) = C e^{-\frac{1}{2}x^T S^{-1}x},$$

donde

$$C = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det S}} \quad \text{y} \quad P'(x) = e^{-\frac{1}{2}x^T S^{-1}x}.$$

Con ello podemos escribir que:

$$\begin{aligned} P(x) &= C \cdot P'(x). \text{ Sacando el gradiente:} \\ \nabla P(x) &= \nabla[C \cdot P'(x)], \\ &= C \nabla[P'(x)], \\ &= C \left[ e^{-\frac{1}{2}x^T S^{-1}x} \nabla \cdot \left[ -\frac{1}{2}x^T S^{-1}x \right] \right]. \end{aligned}$$

Si  $S^{-1}$  es simétrica, entonces

$$\nabla(x^T S^{-1}x) = 2S^{-1}x.$$

Con ello:

$$\begin{aligned} \nabla P(x) &= C \left[ e^{-\frac{1}{2}x^T S^{-1}x} [-S^{-1}x] \right], \\ &= \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det S}} \left[ e^{-\frac{1}{2}x^T S^{-1}x} [-S^{-1}x] \right], \\ &= P(x) [-S^{-1}x], \\ &= -P(x) S^{-1}x. \end{aligned}$$

$$\nabla P(x) = -P(x) S^{-1}x.$$

$$\boxed{\nabla P(x) = -P(x) S^{-1}x.} \quad (3.29)$$

La corriente de probabilidad estacionaria  $J(x)$  ecuación (3.27) puede reescribirse de la siguiente manera [21]:

$$J(x) = DS^{-1}xP(x) - BxP(x) = \mu xP(x). \quad (3.30)$$

$$\mu = DS^{-1} - B. \quad (3.31)$$

Se presentó el fundamento teórico del proceso multivariado Ornstein Uhlenbeck, que constituye la base del modelo anatómico funcional. En el siguiente capítulo, se presenta una descripción y explicación de la irreversibilidad y de la tasa de producción de entropía, que servirán como soporte para la interpretación de los resultados obtenidos.

# 4

## Irreversibilidad y tasa de producción de entropía.

En este capítulo se aborda la equivalencia entre simetría y la reversibilidad temporal. Se proporciona una condición de reversibilidad suficiente y necesaria para un proceso Gaussiano y Markoviano [21], características del proceso Multivariado Ornstein-Uhlenbeck MOU (3.15) expresada en términos de las matrices  $B$  y  $D$ , que cumplen las propiedades enunciadas en (3.13) y (3.14).

### 4.1. Tasa de producción de entropía

El trabajo es el producto de la fuerza por el desplazamiento; la fuerza es el producto del coeficiente de fricción  $D^{-1}$  y la velocidad  $Bx(t)$ , para el caso del proceso Ornstein-Uhlenbeck. Esta es la ley de conservación de la energía. De ello se deduce que la tasa media de disipación de calor, (hdr) es decir, la ley de conservación de la energía [72].

Partimos de la función de disipación de calor instantánea, en esta sección se denota como  $W(t)$ :

$$\begin{aligned}\dot{W}(t) &= D^{-1}Bx(t) \cdot \dot{x}(t) + \nabla B\dot{x}(t), \\ \dot{W}(t) &= \int_{\mathbb{R}^n} D^{-1}Bx(t)\nabla J(x, t),\end{aligned}\tag{4.1}$$

con

$$J(x, t) = -D \nabla P(x, t) - B x(t), P(x, t).\tag{4.2}$$

En términos mecánicos, la tasa media de disipación de calor es el producto de la fuerza  $D^{-1}Bx(t)$ . En términos de la ecuación 3.25  $\Pi$  se denomina fuerza termodinámica [72] consiguiente se puede expresar en términos de probabilidades de transición, lo que conduce a la ecuación (4.3):

$$\begin{aligned}\dot{\Pi}(x) &= \log \left( \frac{P(x + \dot{x}(t), t | x)P(x, t)}{P(x, t | x + \dot{x}(t))P(x + \dot{x}(t))} \right) \\ &\approx \{D^{-1}Bx(t) - \nabla \log P(x, t)\}.\end{aligned}\tag{4.3}$$

donde  $\approx$  es válido para  $\dot{x}(t)$  pequeño.

La relación  $\Pi$  con el flujo  $J$  esta dado por ([21],[72]):

$$\begin{aligned} J &= DP\Pi, \\ \Pi &= \frac{D^{-1}J}{P}. \end{aligned} \quad (4.4)$$

donde la matriz  $D$  es simétrica. Por lo tanto  $J$  y  $\Pi$  refleja cierta simetría [72].

Utilizando la conocida definición de entropía para la distribución de probabilidad  $P(x, t)$ , considerando su versión dependiente del tiempo, tenemos que [21].

$$e[P] = - \int_{\mathbb{R}^n} P(x, t) \log P(x, t) dx. \quad (4.5)$$

La tasa de aumento de la entropía a lo largo del tiempo puede descomponerse en dos factores.

$$\begin{aligned} \dot{e}[P] &= \int_{\mathbb{R}^n} (\log P + 1) \nabla J dx \\ &= - \int_{\mathbb{R}^n} P^{-1} (\nabla P - D^{-1} Bx(t)P) J dx - \int_{\mathbb{R}^n} D^{-1} Bx(t) J dx \\ &= \int_{\mathbb{R}^n} \Pi J dx - \int_{\mathbb{R}^n} D^{-1} Bx(t) J dx \\ &= \Phi - hdr. \end{aligned} \quad (4.6)$$

En el contexto de los procesos de difusión MOU (3.15) estacionarios, una expresión general para la tasa de producción de entropía por unidad de tiempo en el estado estacionario es la siguiente [21]:

$$\Phi = \int (\nabla \log P(x) - DBx(t))^T D (\nabla P(x) - D^{-1} Bx) P(x) dx, \quad (4.7)$$

donde  $\Phi$  es no negativo.

Para la distribución independiente del tiempo  $P(x)$ , tenemos que  $\Phi$  está dada por:

$$\Phi = \int \frac{J^T(x) D^{-1} J(x)}{P(x)} dx = \mathbb{E}[\Pi^T D \Pi], \quad (4.8)$$

donde  $\Pi$  es llamado la fuerza termodinámica y se relaciona con  $J$  por la relación reciproca de Onsager [76]. De acuerdo a [21], a partir de la ecuación (3.31)

$$\begin{aligned} \mu &= DS^{-1} - B \\ x D^{-1} [\mu = DS^{-1} - B] x D^{-1} \\ D^{-1} \mu x P &= (S^{-1} - D^{-1} B) x P \end{aligned}$$

Por (3.30):

$$D^{-1} J = (S^{-1} - DB) x P$$

Por (4.4)

$$\begin{aligned}\Pi &= (S^{-1} - D^{-1}B)x \\ \Pi &= S^{-1}x - D^{-1}Bx.\end{aligned}\tag{4.9}$$

De la ecuación (4.7)  $\nabla \log P(x) = S^{-1}x$ . De la ecuación (4.8) tenemos que:

$$\Phi = \mathbb{E}[\Pi^T D \Pi] = \mathbb{E}\left[x^T (D^{-1}B - S^{-1})^T D (D^{-1}B - S^{-1})x\right].\tag{4.10}$$

De la relación de las ecuaciones (4.10) (4.18) y (3.31) tenemos que:

$$L = BS = D + Q \rightarrow B = DS^{-1} + QS^{-1}$$

Reemplazar en  $D^{-1}B - S^{-1}$ :

$$\begin{aligned}D^{-1}[DS^{-1} + QS^{-1}] - S^{-1} \\ - D^{-1}\mu = D^{-1}QS^{-1}\end{aligned}$$

Por (3.31)

$$-D^{-1}\mu = D^{-1}QS^{-1}.\tag{4.11}$$

Se reemplaza (4.11) en (4.10)

$$\begin{aligned}\Phi &= \mathbb{E}\left[x^T (D^{-1}QS^{-1})^T D (D^{-1}QS^{-1})x\right], \\ &= -\mathbb{E}[x^T \mu^T D^{-1}\mu x].\end{aligned}\tag{4.12}$$

La tasa de producción de entropía es no negativa. Es estrictamente positiva si el proceso es irreversible y se anula solo si es reversible. Dado que el estado estacionario del MOU es gaussiano con matriz de covarianza S, tenemos la siguiente propiedad:

$$\mathbb{E}[x^T Ax] = \text{tr}(SA),\tag{4.13}$$

que puede escribirse en las siguientes expresiones equivalentes que no involucran la matriz de covarianza S ni su inversa explícitamente:

$$\Phi = \text{tr}(S \mu^T D^{-1}\mu),$$

Por (4.11)

$$= \text{tr}\left(S (QS^{-1})^T D^{-1} (QS^{-1})\right).$$

Si  $Q$  es antisimétrica, entonces  $Q^T = -Q$ , luego por (4.18),

$$= \text{tr}(B^T D^{-1}Q) = -\text{tr}(D^{-1}BQ).\tag{4.14}$$

## 4.2. Reversibilidad e irreversibilidad

El producto  $AB$  simétrico es una condición necesaria y suficiente para la reversibilidad de la solución estacionaria del proceso MOU (3.15) [72].

$$\boxed{BD = DB^T}. \quad (4.15)$$

Una solución gaussiana estacionaria reversible de proceso estocástico de la forma  $\dot{x}(t) = -Bx(t) + D\eta(t)$  tiene una matriz de covarianza  $S = B^{-1}D$  donde  $D = \eta(t)\eta^T(t)$  y  $B$  debe tener todos sus valores propios reales y positivos. Tiene una matriz de covarianza en dos tiempos  $S(t, t') = e^{-B|t-t'|}S$  y su  $\Phi$  es cero [72] se sigue que la ecuación de Lyapunov (3.22) (se cumple cuando el proceso es reversible).

$$S = B(B^{-1}D) + (B^{-1}D)B^T,$$

Por (4.11)

$$B(B^{-1}D) + B^{-1}(DB) = 2D. \quad (4.16)$$

En este sentido la ecuación de Lyapunov, sin desfase temporal (3.22) es separable igual a  $D$ . La matriz de covarianza estacionario se reescribe como :

$$S = B^{-1}D = D(B^T)^{-1}. \quad (4.17)$$

Se reparametriza las matrices  $B$   $D$  y  $S$  como

$$\boxed{L = BS = D + Q}, \quad \boxed{L^T = SB^T = D - Q}. \quad (4.18)$$

donde la matriz  $L$  es la matriz de Onsager, su parte antisimétrica,  $Q$  proporciona una medida del grado de irreversibilidad del proceso. Si el proceso es reversible,  $L = D$  es simétrico y  $Q$  se anula [10]. La condición necesaria y suficiente para que un proceso gaussiano estacionario sea irreversible es que la matriz  $B$  en (3.15) tenga todos los valores propios con partes reales positivas y  $D^{-1}B$  sea no simétrica [72].

Los conceptos presentados se aplicarán para fundamentar parte de los resultados obtenidos del modelo anatómico funcional. A continuación, se proporciona una descripción de los datos utilizados en dicho modelo, así como del preprocesamiento de los mismos.

## Datos y preprocesamiento

En este capítulo se presenta una descripción de los datos aplicados así como el preprocesamiento de los mismos.

### 5.1. Elementos de Neurociencia

El cerebro es un sistema jerárquico organizado en múltiples niveles, es responsable de integrar la información sensorial, generar funciones motoras y generar procesos superiores que caracterizan la conducta humana [73].

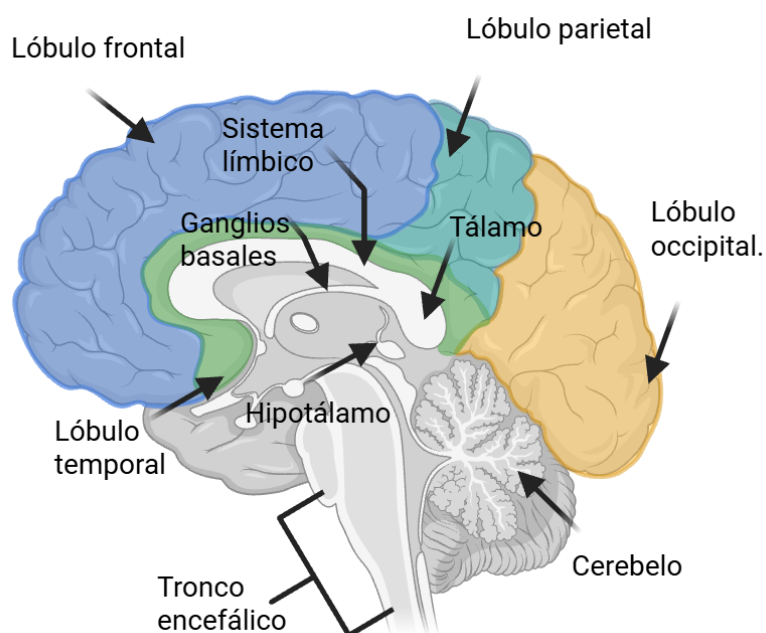


Figura 5.1: Las partes principales y generales del cerebro se pueden dividir en corteza cerebral: lóbulo frontal, parietal, temporal y occipital. Estructuras subcorticales: tálamo, hipotálamo, ganglios basales y sistema límbico. El tronco encefálico: mesencéfalo y bulbo raquídeo y el cerebelo.

### 5.1.1. Datos funcionales.

Los datos funcionales corresponden a las series de tiempo BOLD, a partir de las cuales se estima la conectividad funcional, la cual se denota como FC por su término en inglés Functional Connectivity obtenida mediante resonancia magnética funcional.

### 5.1.2. Resonancia Magnética funcional (functional Magnetic Resonance Imaging fMRI)

Señal BOLD es la abreviación del término en inglés Blood Oxygen Level Dependent, que significa señal que depende del nivel de oxígeno en la sangre. Es la respuesta dependiente de los cambios en el flujo sanguíneo cerebral, el volumen sanguíneo cerebral y la tasa metabólica cerebral de consumo de oxígeno.

Las series BOLD se utilizan para generar mapas *in vivo* en tiempo real de la oxigenación sanguínea cerebral en condiciones fisiológicas normales [81]. Un aumento en el flujo sanguíneo cerebral causa un aumento en la señal BOLD, mientras que los aumentos en el volumen sanguíneo cerebral venoso y la tasa metabólica cerebral de consumo de oxígeno causan una disminución. Típicamente, el efecto del flujo sanguíneo cerebral es dominante, creando una respuesta BOLD positiva. El cambio máximo de la señal BOLD está determinado por el estado fisiológico basal y aumenta con el aumento de hematocrito sanguíneo, la fracción de extracción de oxígeno en reposo y el volumen sanguíneo cerebral [65].

La resonancia magnética es una técnica muy flexible en la que las secuencias utilizadas pueden adaptarse para proporcionar datos óptimos. Las secuencias de resonancia magnética han demostrado ser una aplicación en el estudio de la función cerebral humana al monitorizar los cambios localizados en la intensidad de la señal que acompañan a la actividad neuronal [70].

La resonancia magnética funcional (fMRI) es una técnica no invasiva que permite detectar cambios dinámicos en la actividad cerebral a partir de las series BOLD, el cual se origina en las propiedades magnéticas de la desoxihemoglobina y en el aumento desproporcionado del flujo sanguíneo local respecto al consumo de oxígeno ante incrementos de la actividad neuronal [[6] [69]], la forma de fMRI resulta de la diferencia de susceptibilidad magnética entre la hemoglobina oxigenada diamagnética (Hb) y la hemoglobina desoxigenada paramagnética (dHb) [6].

La adquisición y procesamiento de las imágenes fMRI son elaboradas por [7], en el que los participantes fueron reclutados de la población general en las cercanías de Lovaina y Hasselt (Bélgica) a través de anuncios en sitios web, distribución de folletos y comunicación en eventos públicos (Investigador principal: Stephan Swinnen). La mues-

tra incluyó 164 voluntarios sanos (81 mujeres) con edades entre 10 y 80 años, todos diestros según el Inventario de Lateralidad de Edimburgo. Ninguno de los participantes tenía antecedentes de enfermedades oftalmológicas, neurológicas, psiquiátricas o cardiovasculares que pudieran influir en las pruebas de imagen o las medidas clínicas. Se obtuvo el consentimiento informado antes de la prueba. El estudio fue aprobado por el comité de ética local para la investigación biomédica y se realizó de conformidad con la Declaración de Helsinki.

### **5.1.3. Datos estructurales**

Los datos estructurales permiten estimar la conectividad estructural, la cual se denota como SC por su término en inglés Structural Connectivity, obtenida a partir de imágenes por tensor de difusión.

### **5.1.4. Imágenes por Tensor de Difusión (Diffusion Tensor Imaging DTI)**

El tensor de difusión describe la magnitud, el grado de anisotropía y la orientación principal de la difusión de agua en los tejidos. A partir de estas propiedades, la tractografía basada en DTI permite estimar los patrones de conectividad de la sustancia blanca utilizando la anisotropía de difusión y las direcciones principales de difusión. La fibrotractografía por DTI proporciona información cuantitativa y cualitativa valiosa para la visualización y el estudio de la arquitectura de los tractos fibrilares en el cerebro y en otros tejidos fibrosos [32]. La difusión del agua en los tejidos se ve modulada por la microestructura y la organización tisular; en consecuencia, las técnicas de resonancia magnética ponderadas por difusión, incluida la DTI, constituyen herramientas eficaces para caracterizar los efectos de la enfermedad y del envejecimiento sobre la microestructura tisular [39].

### **5.1.5. Parcelación. Módulos de estructura y función SFM**

La correspondencia entre la estructura y función del cerebro completo se considera un mapeo de uno a muchos, lo que significa que para una anatomía fija, el repertorio funcional debe ser vasto para facilitar la percepción, la acción, la memoria, la cognición y los comportamientos complejos [29]. Por ello, es indispensable contar con una parcelación del cerebro completo que defina unidades de análisis tanto estructural como funcionalmente.

Los datos funcionales y estructurales tienen la parcelación módulos de estructura y función, denominado SFM por su término en inglés Structure Function Modules, proporcionados por el Dr. Rodrigo Cofré, mismos que fueron empleados en [54] y [55] por otros enfoques descritos en el capítulo 6.

La parcelación del cerebro completo es de  $M = 20$ , donde  $M$  se designa como módulos, en este trabajo se denominan regiones de interés (ROIs), es una región cerebral que agrupa voxels con características anatómicas y funcionales comunes. SFM es una partición común óptima compartida por estructura y función maximizando una cantidad novedosa, llamada modularidad cruzada, que divide el cerebro en regiones inconexas, en el que se emplea el algoritmo de Newman [29], el cual dada cualquier red, produce una partición de los vértices en comunidades [33]. Adicionalmente, dicha partición se caracteriza por optimizar un índice definido como la media geométrica entre tres cantidades: la modularidad estructural, la modularidad funcional y la similitud media de Sorensen, ésta última corresponde a la medida de similitud entre dos conjuntos de acuerdo a los elementos que comparten [7]. Por lo tanto, la parcelación SFM es una representación consistente para el análisis integrando conectividad estructural y funcional.

**Definición 5.1.1** (Algoritmo de Newman). *Sea  $e_{ij}$  la fracción de aristas en la red que conectan los vértices del grupo  $i$  con los del grupo  $j$  y sea  $a_i = \sum_j e_{ij}$  tenemos que:*

$$Q = \sum_{i=1}^M (e_{ii} - a_i^2),$$

donde  $M$  es el número de ROIs,  $e_{ii}$  es la fracción de enlaces que conectan dos ROIs  $i$ , y  $a_i$  representa la fracción de enlaces que conectan la ROI  $i$  con el resto de las ROIs de la red.

SFM representa una partición cerebral distinta con otras parcelaciones cerebrales a macroescala, que suelen basarse únicamente en la función o estructura, como es el caso de atlas anatómico automatizado (AAL) que divide en 90 ROIs, que se puede agrupar en 45 áreas homólogas, plegando la ROI izquierda y derecha de cada hemisferio y por otro lado el atlas cerebral funcional de redes de estado en reposo (RSNs) se construye a partir del análisis de componentes independientes de datos funcionales. SFM incorpora regiones cerebrales estructurales y funcionales distintas en una única red/unidad operativa. Pueden superponerse y compartir regiones cerebrales tanto anatómicas como funcionales [29]. La descripción anatómica de cada uno de los 20 ROIs se encuentra en la sección de anexos 9.

### 5.1.6. Conceptos

- **Conectividad Estructural (SC)** La SC mide la densidad o probabilidad de vías anatómicas que conectan dos regiones de interés (ROIs) principalmente a través de la sustancia blanca [58]. Codifica conexiones anatómicas o enlaces entre ROIs [83].

- **Conectividad Funcional (FC)** Es la dependencia estadística temporal entre regiones del cerebro.
- **Conectividad Efectiva (EC)** Se denota como EC por su termino en inglés Effective Connectivity. La EC identifica las interacciones causales que subyacen a la activación ordenada temporalmente o al flujo de información [38].
- **Biomarcador** En términos generales, es una característica medible que funciona como indicador de procesos biológicos normales, procesos patológicos o respuestas a una exposición o intervención, incluidas las intervenciones terapéuticas. Las características moleculares, histológicas, radiográficas y fisiológicas constituyen distintos tipos de biomarcadores [19].
- **Clasificador** Es un modelo que asigna patrones de entrada a clases según una función de decisión aprendida.
- **Regresión logística multinomial** La regresión logística multinomial (MLR) es un clasificador lineal multiclase que asigna un coeficiente a cada variable de entrada para estimar la probabilidad de cada etiqueta.
- **k-vecinos más cercanos** El método de k-vecinos más cercanos (k-NN, por sus siglas en inglés K-Nearest Neighbors) es un clasificador no lineal que asigna la categoría a una nueva muestra en función de la clase mayoritaria entre sus  $k$  vecinos más cercanos en el conjunto de entrenamiento, de acuerdo con una métrica de distancia o una medida de similitud.

El análisis de series de tiempo consiste en el estudio de los modelos matemáticos para representar, controlar y pronosticar el comportamiento de la variable en estudio [43].

- **Serie de tiempo.** Es una sucesión de observaciones de una variable de interés realizada a lo largo del tiempo. Para su representación se usa la siguiente notación,  $\{x_t : t \geq 0\}$  ó  $\{x_t\}$ , en donde se usan letras minúsculas para indicar que se trata de cantidades numéricas.
- **Modelo generativo.** Modelo de ecuaciones (dinámicas) que genera una señal para ajuste a datos empíricos.
- **Optimización de Lyapunov o descenso de gradiente natural** Es un procedimiento matemático para actualizar los pesos de un sistema dinámico, de tal manera que el sistema mantenga estabilidad y reduzca la diferencia entre el modelo y datos reales. En este marco se ajusta los pesos de la EC en el modelo anatómico funcional que ajusta las matrices de covarianza FC del modelo a sus contrapartes empíricas [58].
- **Entropía** La entropía en neurociencia se refiere al grado de desorden o incertidumbre dentro de la actividad neuronal, cuantificando a menudo la aleatoriedad de los estados cerebrales [74].

## 5.2. Preprocesamiento de datos

La conectividad funcional (FC) y la conectividad estructural (SC) constituyen los datos de entrada para el procesamiento. Ambas se representan mediante una matriz de diseño, en la cual cada fila corresponde a una muestra distinta y cada columna corresponde a una característica diferente [28].

Se tienen 161 sujetos, denominados muestras, representados como  $x_i$ , de edades de entre 10 y 80 años ordenados y distribuidos de manera ascendente de acuerdo a la tabla 5.1. Tanto FC como SC adoptan la parcelación del cerebro completo de módulos de estructura y función (SFM). La FC es una matriz  $FC(20, 200)$  en donde 20 son las ROIs y 200 tiempos de repetición denotado como  $TR$ . La SC es una matriz  $SC(20, 20)$  donde 20 son las ROIs.

Rango de edad por años	Número de muestras
10–20	28
20–30	35
30–40	11
40–50	12
50–60	17
60–70	29
70–80	29

Cuadro 5.1: Distribución de las 161 muestras por rangos de edad.

La calidad de los datos y la cantidad de información útil que contienen son factores clave que determinan qué tan bien puede aprender un algoritmo, denominado como procedimiento bien definido que toma un conjunto de valores de entrada y produce un conjunto de valores de salida [28].

Para el procesamiento de las 161 muestras de FC y SC se efectúa el siguiente procedimiento: Para cada muestra de FC.

- Filtrado de las 20 series BOLD mediante el algoritmo de la FFT.
- Homogenización de las series BOLD a 159 tiempos de repetición ( $TR$ ).
- Se aplica la transposición a la matriz de cada muestra, es decir de  $FC(20, 159)$  a  $FC(159, 20)$  de esta manera las ROIs están en las columnas.
- Estandarización.
- Cálculo de la matriz de correlación entre las 20 ROIs ( $20, 20$ ).
- Vectorización de la matriz de correlación, generando el vector de dimensión 400.
- Matriz de correlación funcional de todas las muestras  $X_{FC}(161, 400)$ .

- La matriz de correlación es simétrica, por ello se considera únicamente la parte superior sin la diagonal principal, obteniendo la matriz  $X_{\text{FC}}(161, 190)$ . Cada muestra normalmente se representa como un vector [28]. Por consiguiente tenemos que:

Cada muestra  $x_i \in X_{\text{FC}}$  con  $i\{1, \dots, 161\}$  y  $j\{1, \dots, 190\}$  que representan las 190 características de correlación funcional, está representado por la matriz:

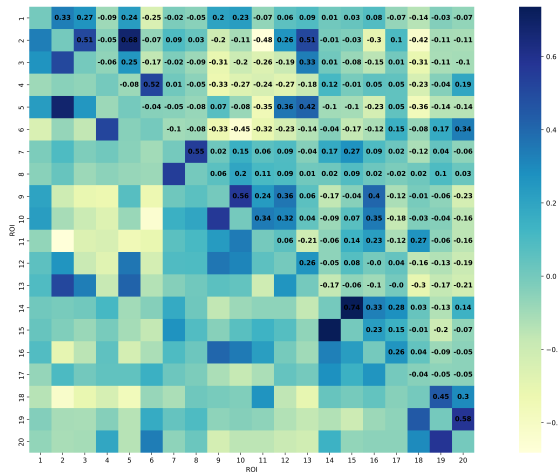
$$X_{\text{FC}} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,190} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,190} \\ \vdots & \vdots & \ddots & \vdots \\ x_{161,1} & x_{161,2} & \cdots & x_{161,190} \end{bmatrix}.$$

Para los datos estructurales  $SC(20, 20)$ , se realiza el siguiente procedimiento.

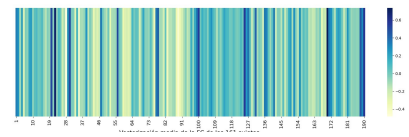
- Estandarización.
- Matriz entre las 20 ROIs (20, 20)
- Vectorización de la matriz, generando el vector de dimensión 400
- Análogamente a FC, se obtiene  $X_{\text{SC}}(161, 190)$  es decir para cada muestra  $x_i$  se tiene  $j'\{1, \dots, 190\}$  representa a las 190 características estructurales.

Consiguientemente se concatenan horizontalmente  $X_{\text{FC}}$  y  $X_{\text{SC}}$  que determina la entrada a los modelos de aprendizaje de máquina,  $X_{\text{FC+SC}}(161, 380)$ .

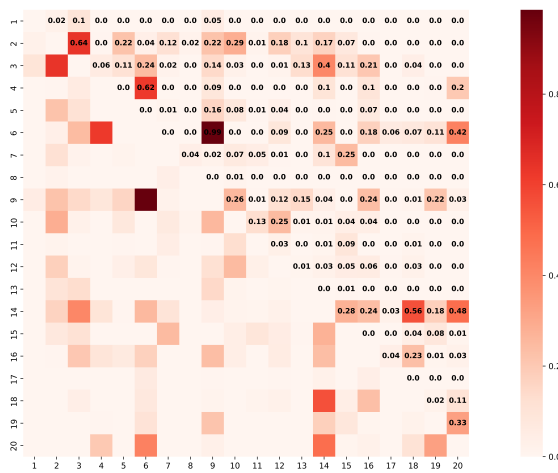
$$X_{\text{FC+SC}} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,190} & x_{1,191} & \cdots & x_{1,380} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,190} & x_{2,191} & \cdots & x_{2,380} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{161,1} & x_{161,2} & \cdots & x_{161,190} & x_{161,191} & \cdots & x_{161,380} \end{bmatrix}. \quad (5.1)$$



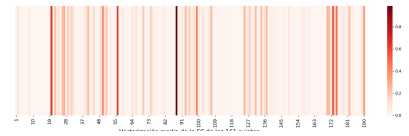
(a) Conectividad funcional (FC).



(b) Vectorización de la FC.



(c) Conectividad estructural (SC).



(d) Vectorización de la SC.

Figura 5.2: Datos de la FC y SC preprocesados. La concatenación serán aplicados a los modelos de aprendizaje de máquina no supervisados y supervisados.

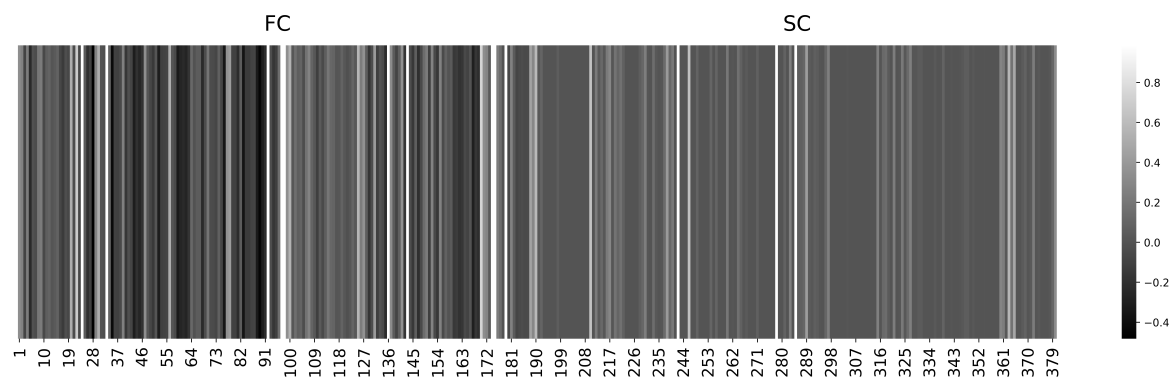


Figura 5.3: Para cada muestra  $x_i \in X_{FC+SC}$  (ver 5.1), se tienen 380 características: 190 de correlación funcional y 190 estructurales.

### 5.2.1. Transformada rápida de Fourier

La transformada rápida de Fourier (denominada como FFT por su nombre en inglés Fast Fourier Transform), funciona descomponiendo una señal de dominio temporal de  $n$  puntos en  $n$  señales de dominio temporal, cada una compuesta por un solo punto. El segundo paso consiste en calcular los  $n$  espectros de frecuencia correspondientes a estas  $n$  señales de dominio temporal. Finalmente, los  $n$  espectros se sintetizan en un único espectro de frecuencia [85].

En el presente trabajo se aplica FFT para el aumento de datos, dado que el conjunto inicial es de 161 muestras, una cantidad limitada para entrenar arquitecturas de aprendizaje profundo. A partir del espectro de magnitud se identificó, para cada sujeto y ROI, la frecuencia dominante correspondiente a la máxima amplitud espectral. Posteriormente, esta frecuencia se transformó en su periodo temporal utilizando la relación:  $T = \frac{1}{f}$ , donde  $T$  representa el período de oscilación y  $f$  la frecuencia dominante. Finalmente, para aumentar el número de muestras, se aplicó el aumento de datos basada en ventanas deslizantes, segmentando cada serie temporal en ventanas solapadas de 13 TR (Tiempo de repetición), con un desplazamiento de uno, generando múltiples segmentos temporales a partir de cada señal original, obteniendo así (23667, 13, 20), donde 23667 son muestras, 13 el tamaño de ventana y 20 ROIs.

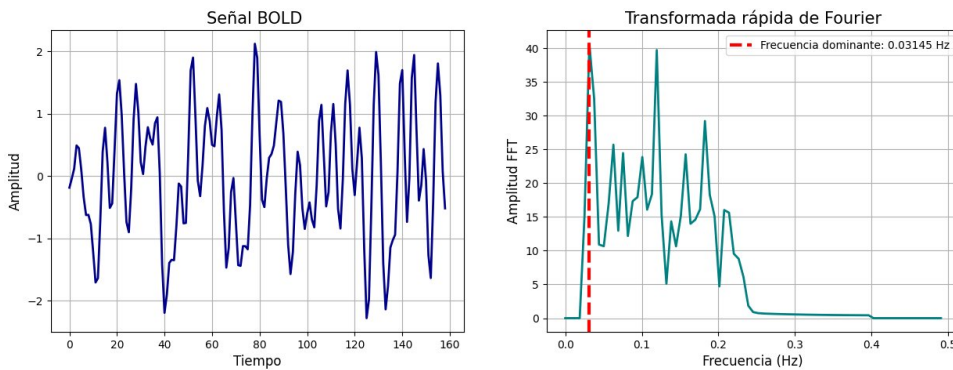


Figura 5.4: Se aplica la FFT para el aumento de datos de las series BOLD, al tener un conjunto inicial de (161, 20, 159) se aumenta a (23667, 13, 20)

El preprocesamiento del conjunto de datos funcionales y estructurales permite transformar los datos en información organizada y útil, facilitando tanto su comprensión como la aplicación en los modelos. En el siguiente capítulo se describen tres enfoques en los que se emplearon estos mismos datos.

# 6

## Estudios previos aplicados a los datos.

En este capítulo se realiza una descripción general del estado del arte y estudios previos con diferentes enfoques aplicados a los datos de fMRI y DTI distribuidos en cuatro diferentes grupos de edades  $I_1, I_2, I_3$  e  $I_4$ .

### 6.1. La información de interacción

Los sistemas neuronales se componen de unidades que interactúan entre sí y la información relevante sobre su función o mal funcionamiento puede inferirse analizando las dependencias estadísticas entre la actividad de cada unidad. Si bien el estudio de las correlaciones y la información mutua se utilizan comúnmente para caracterizar estas dependencias, extender las interacciones a tripletes de variables para detectar y caracterizar mejor la transferencia dinámica de información es un enfoque para la medición de la información de interacción, referida como  $II$  por sus siglas en inglés Interaction Information [4].

En [7] se calcula la información de interacción en datos de fMRI de 164 sujetos sanos en estado de reposo, con un rango de edad de 10 a 80 años, distribuidos de la siguiente manera  $I_1$  de 10-20 años con 30 sujetos,  $I_2$  de 20-40 años con 46 sujetos,  $I_3$  de 40-60 años con 29 sujetos e  $I_4$  de 60-80 años con 59 sujetos, con la parcelación módulos de estructura y función (SFM). En el que se estudia como se distribuyen los valores de redundancia (R) y sinergia (S) entre las áreas cerebrales, denominado módulos, a lo largo de la vida. En el que se encuentra un equilibrio sólido entre R y S entre los diferentes grupos  $I_n$ , lo que indica mecanismos compensatorios de información en las redes cerebrales.

La información de interacción, se mide de acuerdo a la entropía de Shannon, lo que nos permite analizar las interacciones entre tripletes, a partir de la ecuación (2.11) se puede generalizar a dos y tres dimensiones, para el caso de tres se expresa como:

$$H(X, Y, Z) = - \sum_x \sum_y \sum_z p(x, y, z) \log_2 p(x, y, z). \quad (6.1)$$

La entropía es expresada en *bits*, cada *bit* de información proporcionado por las

neuronas reduce la incertidumbre general sobre el estímulo a la mitad. Si los estímulos y las respuestas son independientes, la información mutua es igual a cero. De lo contrario, toma valores positivos [76].

Para cualquier triplete de la forma  $(X, Y, Z)$ . A partir de la ecuación (2.12), la información de interacción se define de la siguiente manera [4]:

$$II(X; Y; Z) \equiv I(X; Y) - I(X; Y|Z), \quad (6.2)$$

donde  $I(X; Y)$  es la información mutua entre  $X$  e  $Y$ , que puede definirse en términos de las entropías marginal y conjunta, i.e. :

$$I(X; Y) = H(X) + H(Y) - H(X, Y), \quad (6.3)$$

donde además  $I(X; Y|Z)$  es la información mutua condicional entre  $X$  e  $Y$  condicionada a  $Z$ . Análogamente  $I(X; Y|Z)$  puede escribirse como una función de las entropías marginal y conjunta, i.e.:

$$I(X; Y|Z) = H(X, Z) - H(Z) + H(Y, Z) - H(X, Y, Z). \quad (6.4)$$

donde además, utilizando (6.2), (6.3) y (6.4),  $II$  se puede expresar como una función de las entropías de Shannon [4] i.e.:

$$II(X, Y, Z) \equiv H(X, Y, Z) + H(X) + H(Y) + H(Z) - H(X, Y) - H(X, Z) - H(Y, Z). \quad (6.5)$$

### 6.1.1. Información de interacción en series de tiempo fMRI en estado de reposo

En la ecuación (6.1) el triplete  $(X, Y, Z)$  representa tres series de tiempo de dinámica funcional en estado de reposo. A partir de (6.2) se obtiene la información mutua para una cópula Gaussiana y se expresa como:

$$I(X; Y; Z) = \frac{1}{2 \ln 2} \ln \left[ \frac{[\Sigma_X][\Sigma_Y][\Sigma_Z]}{[\Sigma_{XYZ}]} \right], \quad (6.6)$$

en donde  $\Sigma_X, \Sigma_Y, \Sigma_Z$  son las varianzas de  $X, Y$  y  $Z$  así como  $\Sigma_{XYZ}$  es la matriz de covarianza conjunta [2].

La información de interacción  $II$  mide el cambio en la información mutua entre dos variables cualesquiera tras añadir una tercera. Si el cambio es positivo, se produce una interacción sinérgica en el triplete, mientras que si es negativo, surgen interacciones redundantes [7]. Para calcular  $II$  se usan los tripletes  $(X, Y, Z)$  de series de tiempo y la ecuación (6.2). Se estima  $I(X; Y)$  e  $I(X; Y|Z)$  usando la información mutua para la cópula gaussiana (6.6) ([2] [7][30]).

Los valores de redundancia y sinergia por módulo se obtuvieron sumando (para un módulo fijo  $m$ ) todos los pares de modo que  $II$  fuera positivo y negativo.

$$R_m \equiv \frac{1}{\mathcal{N}^+} \sum_y \sum_z \Pi^+(X = m, Y, Z), \quad (6.7)$$

$$S_m \equiv \frac{1}{\mathcal{N}^-} \sum_y \sum_z |\Pi^-(X = m, Y, Z)|, \quad (6.8)$$

donde  $\Pi^+$  y  $\Pi^-$ , representan valores positivos y negativos de  $II$ ,  $\mathcal{N}^+$ ,  $\mathcal{N}^-$  es el número total de elementos negativos y positivos,  $|\cdot|$  es el valor absoluto.

Para (6.7) y (6.8) se consideran tripletes en los que las tres variables son distintas entre sí, es decir, que cumplen con  $m \neq y$ ,  $y \neq z$  y  $y \neq z$ , posteriormente se normalizan los valores de R y S. Código disponible en [30].

Para estudiar los efectos de la edad a lo largo de la vida, se realiza una comparación grupal de R, S y R/S con  $I_1, I_2, I_3$  e  $I_4$  se efectúan cuatro etapas [7]:

1. Promedio de todos los módulos R, S y R/S.
2. Prueba de Kruskal-Wallis entre los valores correspondientes a los diferentes intervalos de edad. La prueba Kruskal-Wallis es una prueba robusta no paramétrica basada en rangos para clasificación de diferentes categorías o grupos. La prueba se calcula con la ecuación (6.9).

Sea una muestra aleatoria de tamaño  $N$  de una población grande que consta de  $k \geq 2$  grupos o categorías disjuntos, con proporciones de población  $\pi_i$ ,  $i = (1, \dots, k)$  y  $\sum_{i=1}^k \pi_i = 1$ , sea  $G$  la variable categórica que toma valores en  $G = \{1, 2, \dots, k\}$ . Cuando se dispone de la asignación precisa de la categoría de  $G$ , se realiza la prueba de Kruskal-Wallis para la clasificación de todas las observaciones juntas y comparando la suma de los rangos de cada grupo. Sea  $r_j$  el rango de  $Y_j$  en la muestra general y definamos la variable indicadora  $Z_{ij} = 1(G_j = i)$  para la pertenencia al grupo. La prueba estadística de Kruskal-Wallis se define como [16]:

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1), \quad (6.9)$$

donde  $n_i = \sum_{j=1}^N Z_{ij}$  y  $R_i = \sum_{j=1}^N Z_{ij} r_j$ .

3. Prueba de suma de rangos de Wilcoxon, prueba no paramétrica que compara dos muestras relacionadas o emparejadas, como análisis post-hoc entre todos los pares de comparaciones.

4. Corrección de Bonferroni, método estadístico para controlar el error de tipo 1, es decir, falsos positivos, estableciendo un umbral igual a 0,0083, ya que se contaban con cuatro intervalos de edad, lo que proporcionó seis comparaciones por pares.

### 6.1.2. ROIs derivados de la información de interacción

De acuerdo a este enfoque, realizado por [7] los valores más altos de redundancia se encontraron en los ROIs 3, 6 y 9, la función asociada a dichos ROIs es una superposición de tres redes de estado de reposo, la red por defecto (Default Mode Network DMN), la red sensoriomotora y la auditiva [7]. La actividad intrínseca de la DMN está relacionada con el envejecimiento [80]. En [13] fMRI en estado de reposo mostró una disminución de la actividad de la DMN en pacientes con Alzheimer en comparación con controles sanos de la misma edad.

Los valores más altos de sinergia encontrados son los ROIs son 3, 8 y 18, se asocian con diferentes dominios cognitivos, como la memoria espacial y temporal, el procesamiento de las emociones y las habilidades motoras [7].

## 6.2. Interdependencia de alto orden.

En [55] se analiza cómo las interdependencias estadísticas de alto orden se ven afectadas por la edad y cómo estos cambios se asocian con una disminución en la segregación de los sistemas cerebrales funcionales. Dicha segregación se define a partir de los patrones de correlación funcional en estado de reposo [62]. Para ello, se emplean  $\Omega$ -Información y  $\Sigma$ -Información, debido a que la mayoría de los análisis de FC se centran únicamente en las estadísticas por pares y descuidan las interacciones de orden superior.

Se muestra cómo cambia la organización de la información de alto nivel del cerebro con la edad, a través de las interacciones sinérgicas y redundantes para diferentes órdenes de interacción, por medio de datos de fMRI en estado de reposo de  $N = 164$  sujetos sanos de entre 10 y 80 años, con la misma distribución del enfoque de [7].

### 6.2.1. Información sobre la estructura organizativa

La  $\Omega$ -Información es una forma abreviada de información sobre la estructura organizativa. Se aplica como métrica capaz de caracterizar sistemas dominados por la sinergia y la redundancia [17]. Denotado como  $\Omega(X^n)$  la  $\Omega$ -Información con  $n > 3$  mantiene la capacidad de discriminar entre sistemas dominados por redundancia y sinergia [55]. La  $\Omega$ -Información de un grupo de  $n$  variables aleatorias  $X^n = (X_1, \dots, X_n)$  se define como:

$$\Omega(X^n) = TC(X^n) + DTC(X^n) \quad (6.10)$$

donde  $TC$  es la correlación total, y  $DTC$  es la correlación total dual.  $\Sigma$ -Información denotada como  $\Sigma(X^n)$ , cuantifica la fuerza de las correlaciones multivariadas. La  $\Sigma$ -Información de un grupo de  $n$  variables aleatorias  $X^n = (X_1, \dots, X_n)$  se define como:

$$\Sigma(X^n) = TC(X^n) + DTC(X^n). \quad (6.11)$$

Un sistema de  $n$  variables aleatorias es dominado por la redundancia si  $\Omega(X_1, \dots, X_n) > 0$  y dominado por la sinergia si  $\Omega(X_1, \dots, X_n) < 0$  [17]. Por lo que se aplica una medida de dependencia multivariada que cuantifica cuánta información se comparte en un conjunto de variables aleatorias.

Sean  $TC$  y  $DTC$  que representan la correlación total y la correlación total dual respectivamente ([27] [82]).

$$TC(X^n) = \sum_{i=1}^n H(X_i) - H(X^n), \quad (6.12)$$

$$DTC(X^n) = H(X^n) - \sum_{i=1}^n H(X_i | X_{-j}^n), \quad (6.13)$$

donde  $H$  representa la entropía de Shannon y  $X_{-j}^n$  representa el vector de  $n-1$  variables compuestas por todas menos  $X_j$  i.e.  $(X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_n)$ . Tanto la  $TC$  como la  $DTC$  son generalizaciones no negativas de información mutua, en el sentido de que son cero si y sólo si todas las variables  $X_1, \dots, X_n$  son estadísticamente independientes entre sí.

### 6.2.2. $\Omega$ -Información en series de tiempo fMRI en estado de reposo.

El cálculo de  $\Sigma$ -Información,  $\Omega$ -Información, se obtuvo para cada sujeto y para cada orden de interacción las siguientes cantidades [55]:

$$\Omega_n^m = \frac{1}{Z_n} \sum_{x_2} \dots \sum_{x_n} \Omega_n(X_1 = m, X_2 = x_2, \dots, X_n = x_n), \quad (6.14)$$

$$\Sigma_n^m = \frac{1}{Z_n} \sum_{x_2} \dots \sum_{x_n} \Sigma_n(X_1 = m, X_2 = x_2, \dots, X_n = x_n), \quad (6.15)$$

donde  $m$  representa módulo  $m = \{1, \dots, M\}$ ,  $n$  el orden de interacción y  $Z_n$  es el número total de subconjuntos de tamaño  $n-1$  en un conjunto  $M$ . Estas cantidades representan el promedio  $\Omega$ -Información y  $\Sigma$ -Información por sujeto y módulo [55].

$$Z_n \equiv \binom{M-1}{n-1}, \quad (6.16)$$

$$\Omega_n = \frac{1}{M} \sum_{m=1}^M \Omega_n^m, \quad \Sigma_n = \frac{1}{M} \sum_{m=1}^M \Sigma_n^m, \quad (6.17)$$

$$R_n = \frac{1}{M} \sum_{m=1}^M R_n^m, \quad S_n = \frac{1}{M} \sum_{m=1}^M S_n^m. \quad (6.18)$$

Se compara  $I_4$  con  $I_1, I_2, I_3$  agrupados. El código se puede consultar en [53]. Las diferencias entre grupos se evalúa mediante la prueba estadística no paramétrica de suma de rangos de Wilcoxon. Cuando fue apropiado, los niveles significativos para las pruebas de hipótesis se corrigieron para comparaciones múltiples, controlando la tasa de falsos descubrimientos (TFD) [55].

### 6.2.3. ROIs derivados de la $\Omega$ -Información.

Se obtiene un aumento de la  $\Sigma$ -Información, lo que implica incremento de interdependencias entre varias regiones cerebrales y  $\Omega$ - Información de  $I_4$ , lo que sugiere que las correlaciones observadas, en general, están dominadas por la redundancia y se intensifican en órdenes grandes. La redundancia entre los ROIs 1-3, 5, 13-15 y 18-20 presentan diferencias significativas en todos los órdenes de interacción, mientras que los demás sólo presentan diferencias significativas para órdenes grandes. Esto sugiere la existencia de un núcleo de redundancia. Por el contrario, el patrón de sinergia entre áreas cerebrales es muy heterogéneo, y sólo los ROIs 15, 17, 18 y 19 presentan diferencias significativas.

El núcleo redundante existente en los participantes jóvenes, pero no en los mayores, está compuesto por los ROIs 9, 10 y 16. En particular los 9 y 16 están presentes en todos  $I_n$ . Por lo que se podría plantear la hipótesis de una doble función redundante-sinérgica de la DMN, en donde los participantes de mayor edad, la contribución redundante parece verse afectada. Sin embargo, esta afirmación requiere mayor investigación [55].

## 6.3. Modelo neurodegenerativo.

El conectoma, se refiere al conjunto completo de conexiones anatómicas del cerebro y a su organización como una red estructural, el conectoma proporcionará una herramienta importante para el modelado mecanicista y la interpretación de datos funcionales del cerebro humano [66].

La relación precisa entre las interacciones funcionales de alto orden y el conectoma, así como sus variaciones con la edad, son en gran medida desconocidas, en parte debido a la ausencia de un modelo mecanicista que pueda mapear el conectoma cerebral con la conectividad funcional en el envejecimiento, el modelo neurodegenerativo, constituye un modelo computacional neurobiológicamente realista del cerebro completo utilizando datos estructurales y funcionales de 161 participantes sanos de entre 10 y 80 años en estado de reposo.

En éste modelo se demuestra que las diferencias de edad en las interacciones funcionales de alto orden pueden explicarse por variaciones en el conectoma, la mayor redundancia observada en el grupo  $I_4$  sujetos de 60-80 años, puede explicarse, por reglas que afectan la conectividad anatómica, lo que enfatiza el papel crucial que desempeña el conectoma cerebral en la configuración de interacciones complejas y la eficiencia de la comunicación global del cerebro completo [54].

La distribución de los datos de edades, tiene una pequeña variación en relación a los dos anteriores, 161 sujetos sanos, agrupados en  $I_1$  de 10-20 años con 28 sujetos,  $I_2$  de 20-40 años con 46 sujetos,  $I_3$  de 40-60 años con 29 sujetos e  $I_4$  de 60-80 años con 58 sujetos.

### 6.3.1. Campo medio dinámico

Campo medio dinámico denominado DMF por su nombre en inglés Dynamic Mean Field. El modelado DMF se utiliza para perturbar sistemáticamente las características del conectoma, al tiempo que se evalúan los efectos resultantes en la dinámica y la función cerebral a macroescala, abriendo así la puerta a interpretaciones mecanicistas de los datos obtenidos a partir de estudios de lesiones y envejecimiento [9]. DMF simula la dinámica neuronal de mesoescala mediante ecuaciones diferenciales estocásticas acopladas que incorporan aspectos realistas de la neurofisiología [22]. Así como en [8] se aplica modelo de campo medio y modelo estocástico lineal simple.

DMF simula la actividad neuronal de cada región, cada ROI se modela mediante la interacción de poblaciones neuronales inhibitoras  $I$  y excitatorios  $E$ . DMF asume que las corrientes  $I^I$  están mediadas por receptores ácido gamma-aminobutírico (GABA) y las corrientes  $I^{(E)}$  por receptores N-metil-D-aspartato, (NDMA) ([22],[24],[75]). La conectividad entre dos nodos se expresa como  $C_{np}$  en donde  $n$  y  $p$  son nodos [54] se iguala  $C_{np}$  al elemento de la matriz estructural  $SC_{np}$ . Donde  $S^{(E)}$  poblaciones excitadores y  $S^{(I)}$  poblaciones inhibitorias. Y las tasas de activación excitatoria e inhibitoria respectivamente  $r_n^{(E)}$  y  $r_n^{(I)}$  y el peso de control inhibitorio de retroalimentación, es decir  $J_n^{FIC}$ , ajustado a cada nodo  $n$ . i.e. 6.19 6.1 ([54][79]).

$$\begin{aligned}
I_n^{(E)} &= W_E I_0 + w_+ J_{NMDA} S_n^{(E)} + G J_{NMDA} \sum_{p=1}^N C_{np} S_p^{(E)} - J_n^{FIC} S_n^{(I)}, \\
I_n^{(I)} &= W_I I_0 + J_{NMDA} S_n^{(E)} - S_n^{(I)}, \\
r_n^{(E)} &= F(I_n^{(E)}) = \frac{g_{(E)}(I_n^{(E)} - I_{thr}^{(E)})}{1 - \exp\left\{-d_E g_{(E)}(I_n^{(E)} - I_{thr}^{(E)})\right\}}, \\
r_n^{(I)} &= F(I_n^{(I)}) = \frac{g_{(I)}(I_n^{(I)} - I_{thr}^{(I)})}{1 - \exp\left\{-d_I g_{(I)}(I_n^{(I)} - I_{thr}^{(I)})\right\}}, \\
\frac{dS_n^{(E)}(t)}{dt} &= -\frac{S_n^{(E)}}{\tau_{NMDA}} + (1 - S_n^{(E)})\gamma r_n^{(E)} + \sigma u_n(t), \\
\frac{dS_n^{(I)}(t)}{dt} &= -\frac{S_n^{(I)}}{\tau_{GABA}} + r_n^{(I)} + \sigma u_n(t).
\end{aligned} \tag{6.19}$$

Símbolo	Parámetro	Valor
$I_0$	Corriente externa	0.38nA
$W_E$	Factor de escala excitatorio para $I_0$	1
$W_I$	Factor de escala inhibitorio para $I_0$	0.7
$w_+$	Recurrencia excitatoria local	1.4
$J_{NMDA}$	Unión sináptica excitatoria	0.15 nA
$I_{thr}^E$	Umbral para F $I_n^E$	0.403nA
$I_{thr}^I$	Umbral para F $I_n^I$	0.288nA
$g_E$	Ganancia para el factor F $I_n^E$	310 n $C^{-1}$
$g_I$	Ganancia para el factor F $I_n^I$	615 n $C^{-1}$
$d_E$	Forma de F $I_n^E$ alrededor de $I_{thr}^E$	0.16s
$d_I$	Forma de F $I_n^I$ alrededor de $I_{thr}^I$	0.087s
$\gamma$	Parámetro cinético excitatorio	0.641
$\sigma$	Amplitud del ruido gaussiano no correlacionado $v_n$	0.01nA
$\tau_{NMDA}$	Constante de tiempo para NMDA	100 ms
$\tau_{GABA}$	Constante de tiempo para GABA	10 ms

Cuadro 6.1: Parámetros del modelo DMF

El código está disponible en [48]. Posteriormente se utiliza un modelo hemodinámico biofísico [37] para transformar la dinámica de la tasa  $r_n^{(E)}$  de disparo del modelo DMF en señales similares a BOLD. La señal tipo BOLD del nodo  $n$ , denotada como  $B_n(t)$ , es una función no lineal de  $q_n(t)$  (contenido de desoxihemoglobina) y  $v_n(t)$  (volumen sanguíneo), que puede escribirse como:

$$B_n = V_0[k_1(1 - q_n) + k_2(1 - q_n/v_n) + k_3(1 - v_n)], \tag{6.20}$$

donde  $V_0$  representa la fracción de sangre venosa (desoxigenada) en estado de reposo, y  $k_1 = 2,77$ ,  $k_2 = 0,2$   $k_3 = 0,5$  son constantes cinéticas, elegidas entre [15]. Se aplica la *FC* y la *SC*. El DMF se calibra optimizando un parámetro libre, denominado  $G$ , que permite que el modelo se aproxime mejor a la actividad funcional por pares *FC*. Para cada  $G$  se compara los datos empíricos de los simulados utilizando la distancia Kolmogorow-Smirnov sobre la distribución de los valores de las matrices *FC*, seleccionando  $G$  óptimo que minimiza la distancia [54].

### 6.3.2. Ajuste del modelo en $I_n$

Se construye un modelo DMF para cada  $I_n$  utilizando el promedio de *SC* por cada grupo. Cada modelo es calibrado por separado obteniendo un valor de  $G$  por grupo. Y se calcula las interdependencias de alto orden a partir de datos simulados, metodología en y de [55]. Finalmente se realizan pruebas de Wilcoxon para comparar los valores promedios de R y S en  $I_1, I_2, I_3$  con  $I_4$  [54].

### 6.3.3. ROIs derivados del modelo neurodegenerativo.

El modelo reprodujo el aumento de las interdependencias, de  $I_4$  en concordancia de [55]. Estos cambios funcionales son impulsados por procesos no lineales localizados por la neurodegeneración en el conectoma. A partir de los coeficientes de correlación se construyeron una nueva matriz, obteniendo tres comunidades, de las cuales dos resultaron significativas tras la corrección múltiple, indicando una reducción de la *SC* con la edad. La segunda comunidad estuvo dominada por conexiones de los ROIs 15 y 18, asociadas a estructuras subcorticales, mientras que en la tercera comunidad destacaron los ROIs 6, 9 y 20 por su mayor fuerza de nodo [54].

Los estudios previos descritos muestran el estado del arte: en los dos primeros se aplicó la conectividad funcional, mientras que el tercero integra la conectividad estructural. Consiguiente se presenta el modelo anatómico funcional del cerebro completo asociado a la edad, basado en el proceso multivariado de Ornstein Uhlenbeck, que se desarrolla en tres etapas.

# 7

## Métodos.

En el presente capítulo se describen los métodos desarrollados en esta tesis, la cual se estructura en tres etapas. En la primera etapa se presenta el modelo anatómico funcional, establecido por [21], el cual permite calcular la tasa de producción de entropía  $\Phi$  de sujetos sanos en estado de reposo, denominado así debido a que integra la matriz de conectividad funcional (FC) y la matriz de conectividad estructural (SC), descritos en el capítulo 5 y aplicados con otros enfoques presentados en el capítulo 6. El modelo anatómico funcional está basado en el proceso multivariado Ornstein Uhlenbeck, descrito en el capítulo 3, el cual se aplica a grupos de sujetos distribuidos de acuerdo a su edad biológica, los cuales se denotan como  $I_n$ , los datos son procesados por ([7] y [54]) y proporcionados por el Dr. Rodrigo Cofré.

En la segunda etapa, se implementan métodos de aprendizaje de máquina no supervisados a partir de la concatenación horizontal de la FC y SC (ver 5.2 y 5.3). Como resultado, se obtienen grupos de sujetos denominados clusters, que corresponden a conjuntos de muestras con características similares [5], los cuales se denotan como  $C_n$ . Posteriormente, se aplica el modelo anatómico funcional a  $C_n$ . Para ambas agrupaciones  $C_n$  e  $I_n$ , se utiliza la parcelación módulos de estructura y función (SFM) que considera 20 regiones de interés (ROIs).

Finalmente, en la tercera etapa se implementan modelos de aprendizaje de máquina supervisados a partir de la concatenación horizontal de la FC y SC y modelos de aprendizaje profundo a partir de las series BOLD (por sus siglas en inglés, Blood Oxygen Level Dependent, dependientes del nivel de oxigenación sanguínea), para obtener biomarcadores, que se definen como características explicativas que presentan mayor relevancia en el desempeño de los modelos. Posteriormente, a éstos biomarcadores se aplica el modelo anatómico funcional.

### 7.1. Modelo anatómico funcional aplicado a grupos.

En esta primera etapa, se presenta el modelo anatómico funcional aplicado a grupos de sujetos ordenados de acuerdo a su edad biológica, denotados como  $I_n$ , que se distribuyen de la siguiente manera:  $I_1$  con 28 sujetos de 10 a 20 años,  $I_2$  con 46 sujetos de 20 a 40 años,  $I_3$  con 29 sujetos de 40 a 60 años, e  $I_3$  con 58 sujetos de 60 a 80

años. En la primera etapa, se emplea un código en python, se toma del modelo de ([21], [31], [57]) en el que se desarrolla la tasa de producción de entropía  $\Phi$  para niveles de conciencia, que constituye la base para el modelo anatómico funcional, éste modelo amplía el enfoque de  $\Phi$  asociada a la edad biológica.

Existen diferentes formas de modelar los sistemas dinámicos neuronales. Entre ellos los modelos abstractos, que son aquellos que no simulan ningún parámetro real aunque sí toman algún aspecto cualitativo del comportamiento cerebral [68], los modelos de integración y disparo, como el modelo de neurona única [78], o los más realistas, que son los modelos de estructura múltiple, como los modelos del cerebro completo en el que se integran parámetros fisiológicos reales a nivel funcionamiento ([8] [23] [24]). Aunado a el registro de la actividad neuronal realizado por los neurofisiólogos, que evidencia el carácter estocástico [18].

Se han desarrollado modelos dinámicos de la actividad cerebral, que superan las limitaciones de las medidas originales, como el modelo MOU-EC (EC conectividad efectiva), que proporciona una estimación de la EC que refleja el estado dinámico de la actividad BOLD [58], la estimación directa de la tasa de producción de entropía a partir de registros de actividad neuronal se ve socavada por un muestreo espaciotemporal insuficiente, dado que algunas clases de señales neuronales dependen de series temporales largas [26] lo que lleva a la adopción de heurísticas y aproximaciones que carecen de una justificación rigurosa, así como en [84] en el que se demuestra un modelo generativo que puede inferir en características de conectividad conocida, en una red cerebral completa, que permite deducir la EC en redes cerebrales completas.

En [11] se muestra que, tanto en tareas como en reposo, la dinámica neuronal genera entropía, lo que confirma que el cerebro rompe el equilibrio detallado a gran escala. Además, la producción de entropía observada supera de manera significativa el ruido de fondo asociado a datos finitos. Futuras investigaciones deberían explorar cómo esta ruptura del equilibrio detallado se relaciona con las asimetrías de la conectividad estructural y funcional entre las distintas regiones cerebrales, por lo que en [21] se describe una dinámica estable de la actividad fluctuante y una propagación no conservativa [46].

Por lo que construir un modelo matemático es eficaz para entender un fenómeno o sistema y poder explicar su funcionamiento [43]. Para modelar series de tiempo estacionarias, se sigue un método general de acuerdo a [20] que describe las siguientes partes: identificación del modelo, estimación de parámetros, verificación-diagnóstico y pronóstico. En esta tesis se siguen las siguientes: matrices de covarianza empírica, estimación de parámetros, validación, clasificación y resultados.

### 7.1.1. Matrices de covarianza empírica $\hat{S}_{ij}(0)$ y $\hat{S}_{ij}(1)$ .

El modelo anatómico funcional se ajusta para reproducir las dos matrices de covarianza calculadas a partir de las series empíricas BOLD [21]. Se caracteriza el estado estacionario de no equilibrio del proceso MOU genérico, descrito en el capítulo 3, consecuente, la irreversibilidad del proceso se codifica en la parte antisimétrica de la matriz de Onsager, descrito en el capítulo 4, mientras que la linealidad de las ecuaciones de Langevin nos permite derivar una expresión cerrada para tasa de producción de entropía  $\Phi$  en función de las matrices que definen el proceso MOU, como resultado se obtiene una estimación basada la tasa de producción de entropía  $\Phi$  ajustado a datos fMRI [21].

Las series temporales BOLD se filtran en el dominio de la frecuencia mediante la transformada rápida de Fourier (FFT por sus siglas en inglés fast Fourier transform), el cual permite aislar las oscilaciones lentas de la señal, que son las más relevantes para el estudio de la conectividad funcional en estado de reposo y que se concentran principalmente en el rango de frecuencias de  $0.1 - 0.2 \text{ Hz}$ .

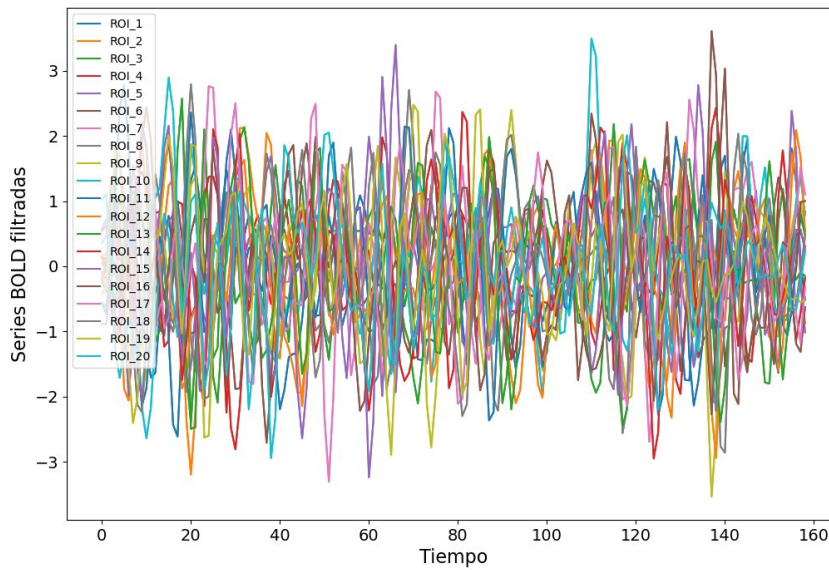


Figura 7.1: Se muestran 20 series temporales BOLD filtradas mediante FFT, correspondientes a un sujeto, una por cada región de interés (ROI). Las series fueron adquiridas con  $\text{TR} = 3000 \text{ ms} = 3 \text{ s}$  donde TR (Tiempo de Repetición), corresponde al intervalo en el que se adquiere un volumen completo del cerebro [7]. Luego de homogenizar las series de los 161 sujetos, cada serie contiene 159 puntos temporales.

En [35] se demuestra que la correlación lineal de datos normalizados captura adecuadamente la FC completa para datos típicos de fMRI en estado de reposo, en el que

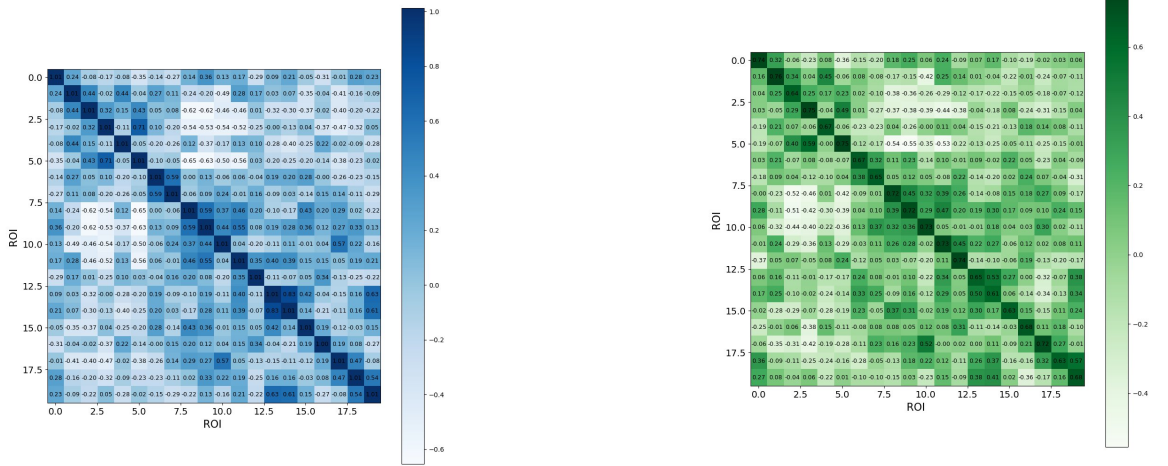
la mayor parte de la información se encuentra en las estadísticas de segundo orden de las series BOLD. El modelo se ajusta para reproducir las dos matrices de covarianza empíricas obtenidas a partir de las series BOLD con una autocorrelación y con un desfase temporal de uno [21], éstas se definen, respectivamente, como:

$$\hat{S}_{ij}(0) = \frac{1}{T-2} \sum_{1 \leq t \leq T-1} [x_i(t) - \bar{x}_i] [x_j(t) - \bar{x}_j] \quad (7.1)$$

$$\hat{S}_{ij}(1) = \frac{1}{T-2} \sum_{1 \leq t \leq T-1} [x_i(t) - \bar{x}_i] [x_j(t+1) - \bar{x}_j] \quad (7.2)$$

donde  $\bar{x}_i$  denota la señal empírica media,  $\bar{x}_i = \frac{1}{T} \sum_t x_i(t)$  para todo  $i$ , lo cual se utiliza para centrar los datos, ya que todas las variables  $x_i$  tienen media cero en el modelo. Estas son las contrapartes empíricas de las covarianzas del modelo  $S_{ij}(0)$  y  $S_{ij}(1)$ , promediadas en el tiempo  $t$  [21]. La contraparte del modelo anatómico funcional está desarrollado en el capítulo ??, es decir, la covarianza en  $t = t'$  siguiendo la ecuación (3.22) y la covarianza en  $t' > t$  con la ecuación (3.24). El uso de covarianzas con desplazamiento temporal tanto cero como no cero es la clave para inferir la EC [60]. La covarianza estacionaria de las fluctuaciones espontáneas se relaciona explícitamente con la conectividad y la dinámica subyacente [25]. La correlación de Pearson a partir de las covarianzas 7.1 y 7.2.

$$P_{ij} = \frac{\hat{S}_{ij}(0)}{\sqrt{\hat{S}_{ij}(0)\hat{S}_{ij}(1)}} \quad (7.3)$$



a)  $\hat{S}(0)_{ij}$  Matriz de conectividad funcional (FC) sin desfase temporal.

b)  $\hat{S}(1)_{ij}$  Matriz de conectividad funcional (FC) con desfase temporal de uno.

Figura 7.2: La mayor parte de la información relevante se encuentra en las estadísticas de segundo orden de las series BOLD,  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$  son clave para inferir en la conectividad efectiva (EC).

La SC se establece como el promedio de los 161 sujetos lo cual da lugar a un esqueleto de conexiones en el que los pesos corresponden a enlaces ausentes por lo que se mantienen fijos en 0 en todo momento, dando lugar a una matriz booleana, mientras que los pesos de las conexiones existentes se estiman a partir de las matrices *FC* durante el proceso de optimización [87].

La SC mide la densidad de fibras neuronales y no es suficiente para explicar completamente la estructura de la FC, que también depende de la dinámica de los ROIs. Las interacciones dinámicas entre las áreas corticales se reflejan en la EC [58]. La EC y SC se relacionan de forma muy significativa, se han observado fenómenos contraintuitivos en diversos sistemas, como en redes de osciladores acoplados que destacan que en ciertas condiciones, el ruido puede ayudar a reconstruir la conectividad estructural a partir de la EC basada en correlación [36].

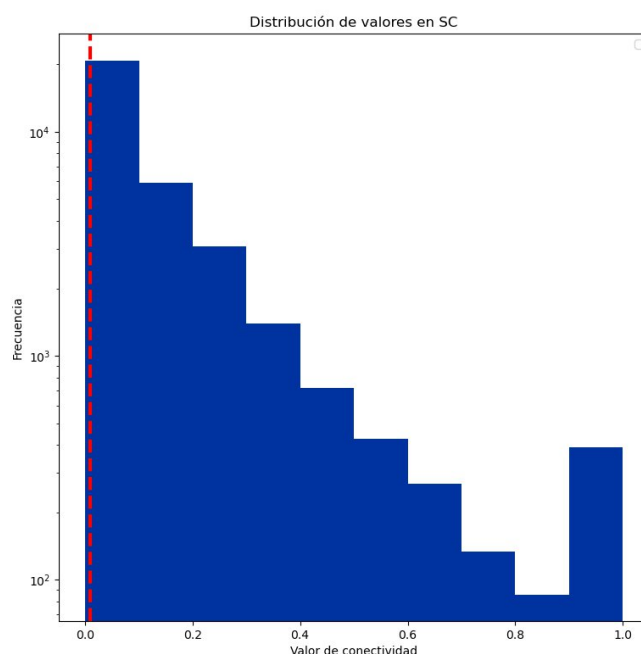
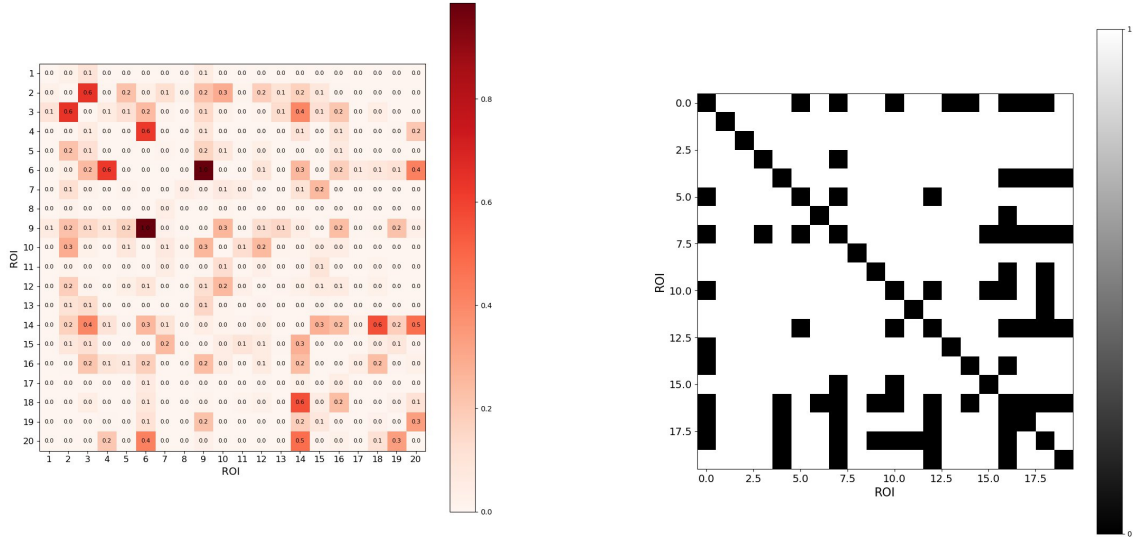


Figura 7.3: Los valores de la SC de los 161 sujetos son previamente normalizados, aquellos que superan el umbral de 0.00001 se consideran conexiones existentes.



a) Media de la matriz de conectividad estructural (SC) de los 161 sujetos, la cual forma un esqueleto de conexiones.

b) Máscara booleana con una densidad de conexiones del 81.05 %.

Figura 7.4: Matriz de conectividad estructural (SC) de 20 ROIs y su correspondiente máscara booleana.

### 7.1.2. Estimación de parámetros

La dinámica de la red está determinada por dos conjuntos de parámetros: matriz Jacobiana,  $-B_{ij} = C_{ij}$ , que describe la interacción causal entre regiones cerebrales  $ROI_i \rightarrow ROI_j$  y la matriz de covarianza de entrada  $D_{ij}$  que representa la actividad espontánea de cada ROI ([21] y [59]).

Conforme a trabajos previos: la inferencia de la topología de interacción a partir de la dinámica de la red de [52], el modelo estocástico dinámico de redes neuronales basado en ecuaciones diferenciales estocásticas ordinarias acopladas no lineales que satisfacen relaciones de media, varianza y covarianza [77] así como el modelo de picos y el modelo del campo medio dinámico y sus aproximaciones lineales [25], se derivan las ecuaciones de autoconsistencia para la media  $\mathbb{E}[x]$  y la covarianza  $S_{ij}$ . Se supone la estacionariedad del proceso, cuando el sistema tiene un punto fijo estable, este viene dado por los ceros del sistema matricial lineal [60].

#### Matriz Jacobiana $C_{ij}$

Consideremos una red neuronal interconectada. La actividad de cada ROI está determinada por:

$$\dot{x}_i(t) = \left[ -\frac{x_i^t}{\tau_x} + \sum_{j \neq i} C_{ij} x_j^t + e_i \right] + dW_i^t. \quad (7.4)$$

Las variables de la actividad se describen mediante un proceso MOU,  $x_i^t$  representa la actividad de cada ROI,  $i$  decae exponencialmente según la constante de tiempo  $\tau_x$  donde  $\tau \in \{0, 1\}$  [59] y es excitado por otros cuyas contribuciones están escaladas por los pesos recurrentes  $C_{ij}$ . Las fluctuaciones son generadas por el ruido blanco Gaussiano  $dW_i$  con varianza  $(\sigma_i)^2$ , formalmente  $W_i$  es un proceso de Wiener escalado por un factor  $\sigma_i$ . Sin pérdida de generalidad, la entrada  $e_i$  es homogénea para todos los ROIs [60].

Se calcula el Jacobiano de la ecuación (7.4), definimos  $f_i(x)$  como:

$$f_i(x) = -\frac{x_i^t}{\tau_x} + \sum_{j \neq i} C_{ij} x_j^t + e_i, \quad (7.5)$$

$$\begin{aligned} \frac{\partial f_i}{\partial x_j} &= \frac{\partial}{\partial x_j} \left( -\frac{x_i}{\tau_x} \right) + \frac{\partial}{\partial x_j} \left( \sum_{j \neq i} C_{ij} x_j \right) + \frac{\partial}{\partial x_j} e_i \\ &= -\frac{1}{\tau_x} \delta_{ij} + C_{ij}. \end{aligned} \quad (7.6)$$

Para  $n = 3$  tenemos la matriz  $J_{ij}$

$$J_{ij} = \begin{pmatrix} -\frac{\delta_{11}}{\tau_x} + C_{11} & -\frac{\delta_{12}}{\tau_x} + C_{12} & -\frac{\delta_{13}}{\tau_x} + C_{13} \\ -\frac{\delta_{21}}{\tau_x} + C_{21} & -\frac{\delta_{22}}{\tau_x} + C_{22} & -\frac{\delta_{23}}{\tau_x} + C_{23} \\ -\frac{\delta_{31}}{\tau_x} + C_{31} & -\frac{\delta_{32}}{\tau_x} + C_{32} & -\frac{\delta_{33}}{\tau_x} + C_{33} \end{pmatrix}.$$

donde  $\delta_{ij}$  es el delta de Kronecker, si  $i = j$  entonces  $\delta_{ij} = 1$  y 0 en caso contrario [60].  $C_{ij}$  tiene diagonal nula, es decir, no se consideran autoconexiones ([60], [21]) y los elementos fuera de la diagonal, es decir  $C_{ij} = -B_{ij}$  corresponden al concepto de conectividad efectiva  $CE$  de ROI  $j$  al ROI  $i$ , excitatorio cuando  $C_{ij} > 0$  [21]. Las condiciones para la construcción de la matriz de conectividad efectiva está dada de la siguiente manera:

$$C_{ij} = \begin{cases} 0, & si \ i = j, \\ C_{ij} & si \ i \neq j. \end{cases} \quad \delta_{ij} = \begin{cases} 0, & si \ i \neq j, \\ 1 & si \ i = j. \end{cases} \quad (7.7)$$

$$C_{ij} = -B_{ij} = \begin{pmatrix} -\frac{1}{\tau_x} & C_{12} & C_{13} \\ C_{21} & -\frac{1}{\tau_x} & C_{23} \\ C_{31} & C_{32} & -\frac{1}{\tau_x} \end{pmatrix}.$$

El sistema tiene un único punto fijo siempre que  $B_{ij}$  sea invertible.

El modelo se calibra primero calculando la constante de tiempo  $\tau_x$  a partir de las señales empíricas (7.1) y (7.2),  $\tau_x$  es idéntico para todas las ROIs [21].

$$\tau = -\frac{N}{\sum_{1 \leq i \leq N} a(u_i|u)} = -\frac{N}{\sum_{1 \leq i \leq N} [\log(\hat{S}_{ii}(0)), \log(\hat{S}_{ii}(1))]}, \quad (7.8)$$

donde  $a(u_i|u)$  es la pendiente de la regresión lineal,  $u = [0, 1]$ ,  $N$  es el número de ROIs ([21], [59]).

### Matriz de covarianza $D_{ij}$

Se le denomina conectividad matriz de covarianza  $D_{ij}$ . A partir de la ecuación de Lyapunov (3.22).  $D_{ii}$  corresponde a una matriz de covarianza diagonal, en donde  $dW_i$  representan entradas del ruido independientes [59].

$$D_{ij} = \delta_{ij} \mathbb{E}[dW_i^t dW_j^t], \quad (7.9)$$

donde  $\delta_{ij}$  es delta de Kronecker.

$$\delta_{ij} = \begin{cases} 1, & \text{si } i = j, \\ 0, & \text{si } i \neq j. \end{cases} \quad (7.10)$$

Para  $n = 3$  una matriz D, se representa de la siguiente manera:

$$D = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}.$$

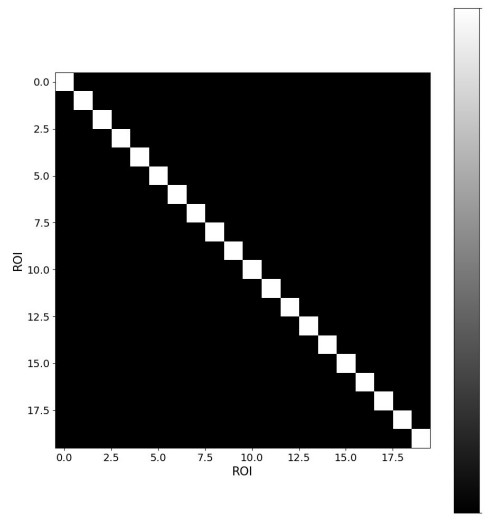


Figura 7.5: La matriz de varianza  $D_{ij}$  refleja la amplitud de la fluctuación de la ROI  $i$  [21].

## Optimización de Lyapunov

Se realiza una optimización de Lyapunov o descenso de gradiente para ajustar iterativamente las matrices  $C_{ij} = B_{ij}$  y  $D_{ij}$  hasta alcanzar el mejor ajuste [60], descrito en el esquema 7.6. En cada paso de optimización, se calculan las contrapartes del modelo de las matrices de covarianza en la ecuación (7.1)  $S_{ij}(0)$  y  $S_{ij}(1)$ , asumiendo estacionariedad en cada sesión de fMRI [21].

MOU optimizado extrae información espaciotemporal sobre la dinámica BOLD. Para inferir la conectividad a partir de observaciones empíricas, es necesario definir los observables de la actividad de los ROIs que el modelo anatómico funcional debe replicar [60], es decir, las matrices de covarianza teóricas con y sin desfase temporal denotados por  $S_{ij}(0)$  y  $S_{ij}(1)$  de las series BOLD en estado de reposo. Al realizar la optimización de Lyapunov se ajusta iterativamente la conectividad de la red para reproducir momentos de segundo orden de la actividad de cada ROI, es decir, las matrices de covarianza empíricas con y sin desfase temporal denotado por  $\hat{S}_{ij}(0)$  y  $\hat{S}_{ij}(1)$  siguiendo la figura 7.6 correspondientes a los 161 sujetos agrupados de acuerdo a  $I_n$ .

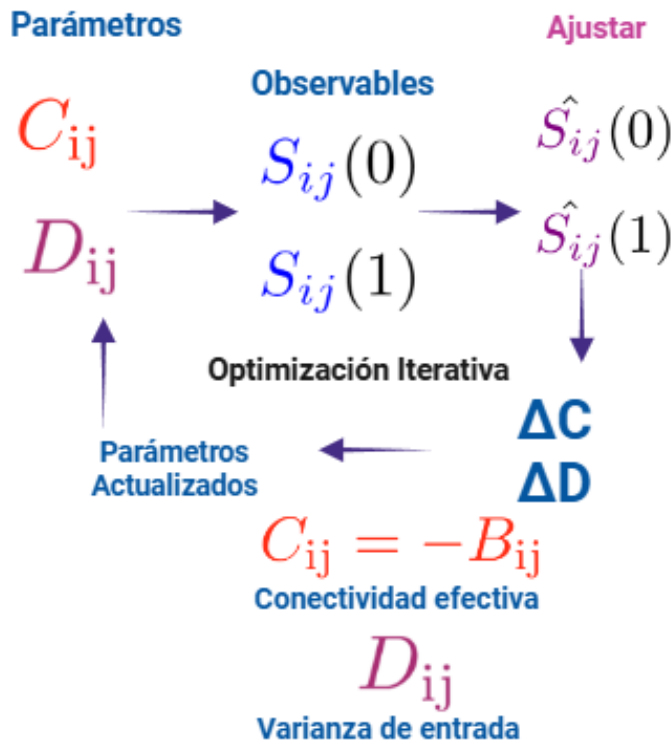


Figura 7.6: Los observables, es decir, las matrices de covarianza teóricas del modelo anatómico funcional con y sin desfase temporal  $S_{ij}(0)$  y  $S_{ij}(1)$ , se ajustan a las matrices empíricas  $\hat{S}_{ij}(0)$  y  $\hat{S}_{ij}(1)$ .

A partir de ecuación de Lyapunov, para  $t' = t$ , (3.22) y para tiempos diferentes  $t' > t$  (3.24). Se calculan las diferencias entre el modelo y las covarianzas empíricas [21].

$$\begin{aligned}\Delta B &= \epsilon_B [S(0)]^{-1} \left[ \Delta S(0) - \Delta S(1)e^{B^T} \right], \\ \Delta D &= \epsilon_D B \Delta S(0) + \epsilon_D \Delta S(0) B^T.\end{aligned}\tag{7.11}$$

con  $\epsilon_B$  y  $\epsilon_D$  pequeñas tasas de aprendizaje. El mejor ajuste corresponde a minimizar la norma al cuadrado tanto de  $\Delta S_{ij}(0)$  como de  $\Delta S_{ij}(1)$  [21]. La optimización de Lyapunov captura los efectos de la red debido a la retroalimentación recurrente [60].

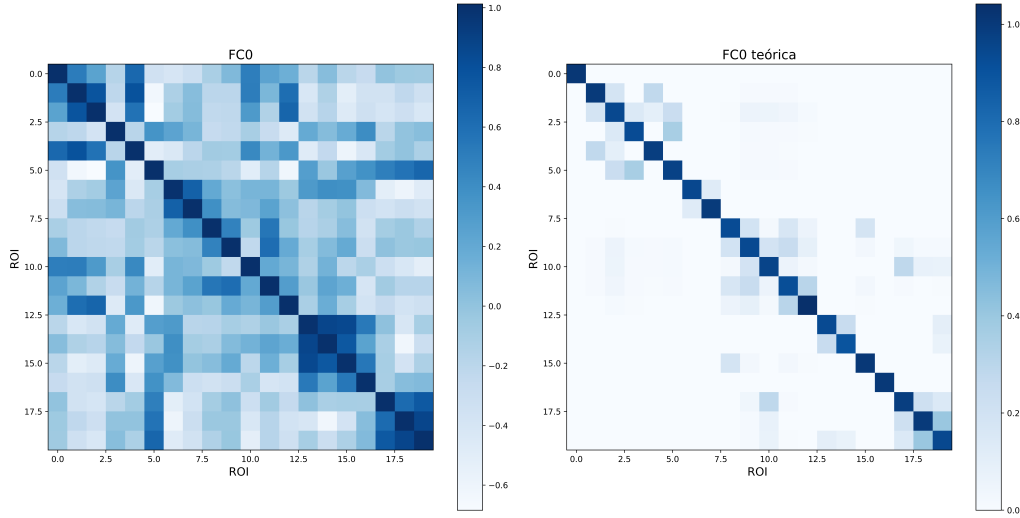


Figura 7.7: Matrices de covarianza, empírica  $\hat{S}(0)_{ij}$  y teórica  $S(0)_{ij}$  de los grupos  $I_n$  con 20 ROIs.

Los valores más altos en  $\hat{S}(0)_{ij}$  son los que el modelo logra estimar mejor.

### 7.1.3. Validación del modelo

#### Error del modelo y bondad de ajuste

La validación del modelo anatómico funcional se realiza mediante dos métricas complementarias: el error del modelo y la bondad de ajuste. El error del modelo se define como la distancia normalizada basada en la norma de Frobenius, ecuaciones (7.12) y (7.13) promediada entre la matriz de covarianza sin retardo  $\hat{S}_{ij}(0)$  y la matriz de covarianza teórica  $S_{ij}(0)$ , dada por la ecuación (7.14) ([21], [58] y [60]). La norma de Frobenius de una matriz  $A = [a_{ij}]$  se define como  $\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}^2|}$ , ésta métrica proporciona una medida de la discrepancia entre las  $\hat{S}(0)_{ij}$  y  $S(0)_{ij}$ .

$$\text{error}_{S_{ij}(0)} = \frac{\left\| \hat{S}_{ij}(0) - S_{ij}(0) \right\|}{\left\| \hat{S}_{ij}(0) \right\|},\tag{7.12}$$

$$\text{error}_{S_{ij}(1)} = \frac{\|\hat{S}_{ij}(1) - S_{ij}(1)\|}{\|\hat{S}_{ij}(1)\|}. \quad (7.13)$$

El error total del modelo se define como el promedio de ambas distancias como:

$$\text{Error} = \frac{1}{2} (\text{error}_{S_{ij}(0)} + \text{error}_{S_{ij}(1)}). \quad (7.14)$$

La bondad de ajuste se cuantifica mediante la correlación de Pearson de acuerdo a la ecuación (7.15) entre la matriz de covarianza sin retardo  $\hat{S}_{ij}(0)$  y la matriz de covarianza teórica  $S_{ij}(0)$  ([21], [58] y [60]).

$$\text{ajuste} = \frac{1}{2} \left[ \rho(\hat{S}_{ij}(0), S_{ij}(0)) + \rho(\hat{S}_{ij}(1), S_{ij}(1)) \right]. \quad (7.15)$$

donde  $\rho$  es el coeficiente de correlación de Pearson.

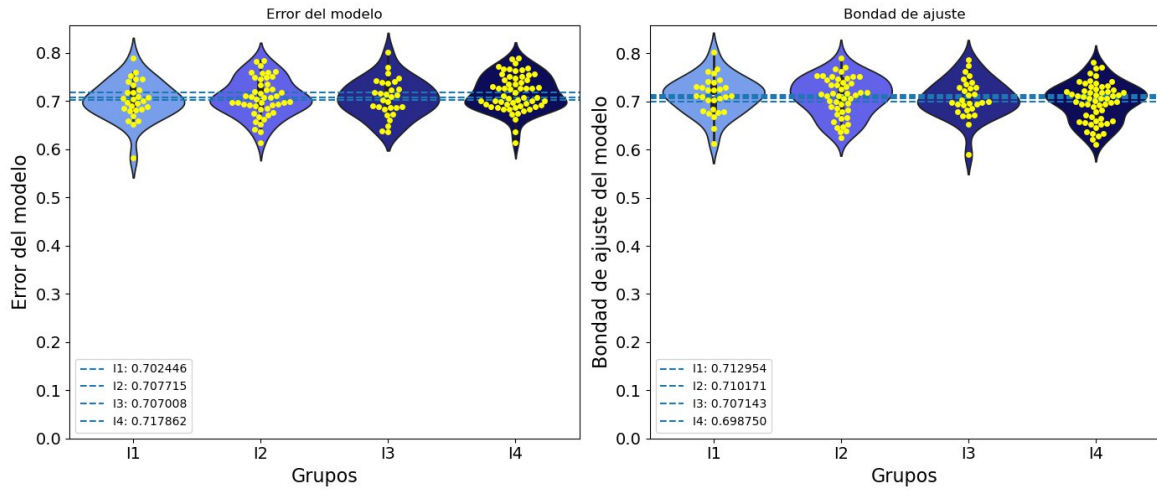


Figura 7.8:

Permite evaluar qué tan bien el modelo reproduce la matriz teórica  $S(0)_{ij}$ , al comparar los cuatro grupos se observa consistencia entre las medias, manteniendo un desempeño comparable entre los distintos grupos.

#### 7.1.4. Clasificación

##### Clasificador Regresión Logística Multinomial

El clasificador MLR denominado así por su nombre en inglés Multinomial Logistic Regression, es una herramienta canónica para la clasificación lineal de alta dimensión. Ajusta un peso para cada característica (una por dimensión), seleccionando así las importantes. Los parámetros (o regresores) del modelo se ajustan para predecir las probabilidades de que nuevas muestras pertenezcan a cada clase (o categoría). Se basa en la siguiente función logística, relacionada con la probabilidad de que la medida de

conectividad vectorizada  $v^l$  ( con elementos indexados por  $i$ ) para la sesión  $l$  pertenezca a la clase objeto  $s$  [87].

$$Pr(v^l \in s) = \psi \left( \sum_i w_i^s v_i^l \right), \quad (7.16)$$

donde  $\psi$  es la función sigmoide  $F(t) = \frac{1}{1+e^{-t}}$  con rango de  $(0, 1)$ .

El entrenamiento se realiza mediante una regresión para determinar el peso de clasificación  $w_i^s$  de tal manera que  $Pr(v^l \in s)$  discrimine la clase  $s$  del último sujeto  $s$ .

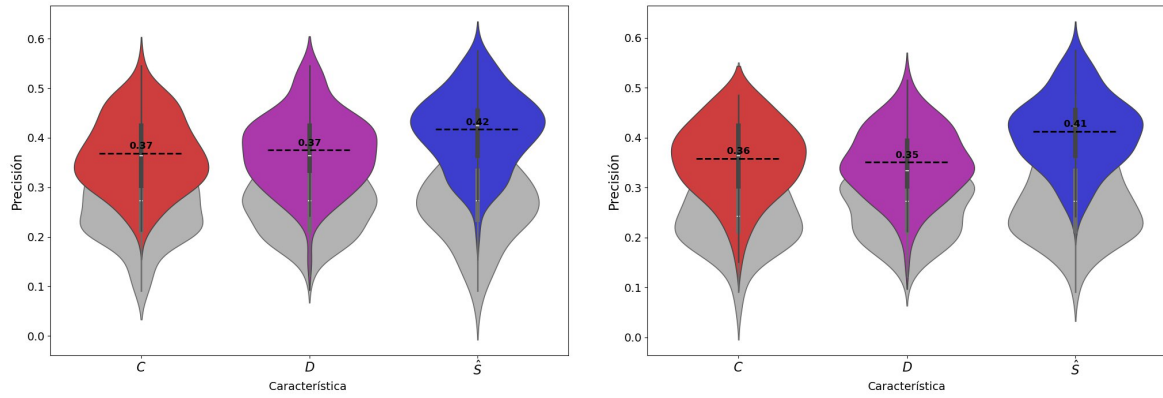
### Clasificador 1-NN

El clasificador kNN denominado por su nombre en inglés 1-Nearest Neighbor, es una técnica que asigna a una nueva muestra la clase a la que pertenecen la mayoría de sus  $k$  vecinos más cercanos. En este trabajo se asigna a  $k = 1$  de tal manera que cada sujeto se clasifica únicamente en función de su vecino más similar, lo que permite evaluar la proximidad entre pares de sujetos. Para cuantificar dicha similitud cada individuo se representa mediante el vector  $v^i$ . A partir de estos vectores, se define una medida de similitud entre sujetos expresada por:

$$S^{ij} = PCC(v^i, v^j), \quad (7.17)$$

donde  $S^{ij}$  representa la similitud entre sujetos,  $i \neq j$ ,  $PCC$  es el coeficiente de correlación de Pearson, el cual cuantifica el grado de asociación lineal entre dos vectores, lo cual permite construir la distribución de similitud entre sujetos (Between subject similarity BBS) [87].

La implementación en el modelo anatómico funcional del clasificador MLR y 1-NN, con las entradas de las matrices vectorizadas  $C_{ij}$ ,  $S_{ij}$ , y  $D_{ij}$  descartando elementos cero y elementos redundantes [21]. Se implementa una validación cruzada 80% de entrenamiento y 20% de prueba con 100 divisiones aleatorias y un máximo de 500 iteraciones.



a) Clasificador de regresión logística multinomial.

b) Clasificador 1-NN.

Figura 7.9: Las medias de cada características  $C$ ,  $D$  y  $\hat{S}$  se encuentran por encima del valor obtenido con shuffle (representado por los violines de color gris). Esto indica un buen desempeño del modelo, ya que logra reconocer dichas características y no se debe al azar.

### 7.1.5. Tasa de producción de entropía e (Ir)reversibilidad

La tasa de producción de entropía es siempre no negativa; es estrictamente positiva si el proceso es irreversible y se anula únicamente en procesos reversibles. Según la ecuación (4.14), se tiene:

$$\Phi = \text{tr}(B^T D^{-1} Q) = - \text{tr}(D^{-1} B Q),$$

donde  $B$  representa la matriz Jacobiana del sistema,  $D$  es la matriz de covarianza, y  $Q$  es la matriz que cuantifica el grado de irreversibilidad del proceso. Esta expresión permite evaluar de manera explícita cómo la dinámica del sistema y sus fluctuaciones contribuyen a la generación de entropía.

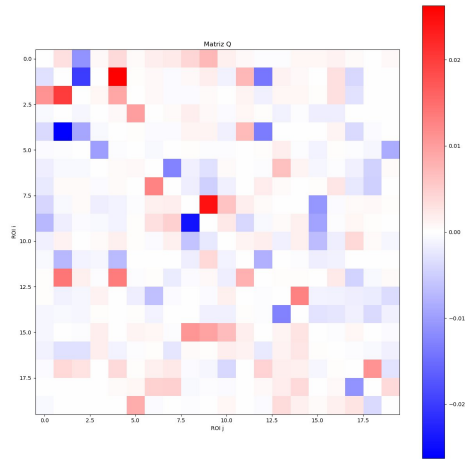


Figura 7.10: Matriz  $Q$  promedio de los 161 sujetos, se muestra la asimetría.

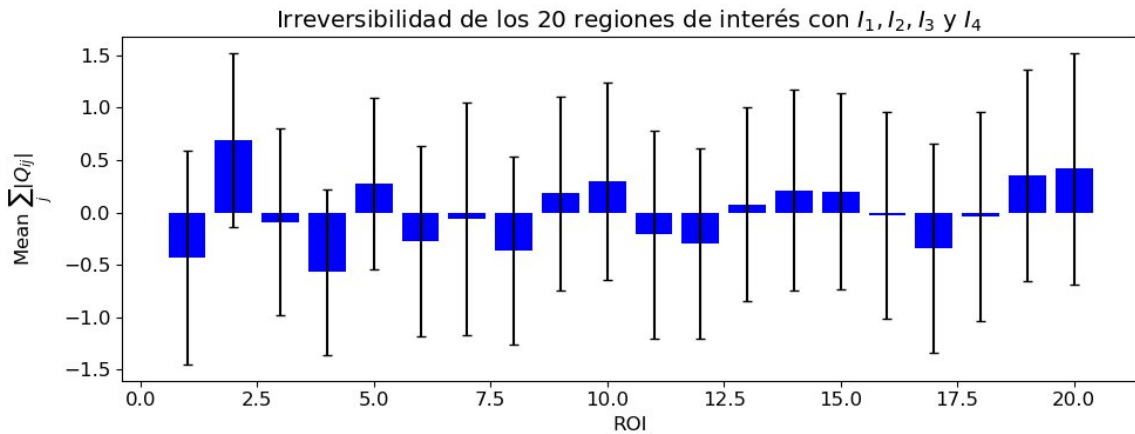


Figura 7.11: Media de irreversibilidad por ROI, calculada como la suma de los valores absolutos de  $Q_{ij}$  normalizados por sujeto.

## 7.2. Aprendizaje de máquina no supervisado.

En la segunda parte, se realiza el problema inverso, es decir inferir la edad de los sujetos a partir de la dinámica cerebral, para éste fin, se desarrollan modelos de aprendizaje de máquina no supervisados a partir de la concatenación horizontal de la FC y SC, denominada como  $X_{FC+SC}$  (ver 5.2), para analizar y explorar la estructura interna de los datos y extraer información significativa sin la guía de una variable conocida [64]. Como resultado, se obtienen grupos denominados clusters  $C_n$  y posteriormente se aplica el modelo anatómico funcional a dichos clusters.

Los métodos de aprendizaje de máquina se construyen utilizando las librerías del lenguaje de programación de scikit-learn en python con las siguientes versiones, Numpy 1.26.4 para el análisis numérico, Pandas 2.2.3 para manipulación de datos tabulares, Matplotlib 3.10.0 para la visualización básica de datos, Seaborn 0.13.2 para la visualización estadística, SHAP 0.48.0 para la interpretación de modelos y Scipy 1.13.1 para el análisis estadístico.

El aprendizaje de máquina es la ciencia (y el arte) de programar computadoras para que puedan aprender de datos [5] es decir, se basa en la construcción de modelos matemáticos que son capaces de identificar patrones en los datos mediante parámetros ajustados a las observaciones, permitiendo que el programa aprenda de ellos [34]. De esta manera, es posible extraer información de datos para mejorar gradualmente el rendimiento de los modelos y tomar decisiones basada en datos [64].

## Análisis de componentes principales PCA

Los sistemas de aprendizaje de máquina se pueden clasificar según la cantidad y el tipo de supervisión que reciben durante el entrenamiento, es decir en el proceso de ajuste de parámetros. Los modelos de aprendizaje no supervisado identifican la estructura interna en los datos no etiquetados, lo que implica modelar características de un conjunto de datos sin referencia a ninguna etiqueta [5].

PCA es un método lineal no supervisado que se aplica para la reducción de la dimensionalidad de datos, visualización de datos de alta dimensión, filtrado de ruido y para la selección de características dentro de un conjunto de datos de alta dimensión [34]. En este trabajo se aplica PCA para la reducción de dimensionalidad de la concatenación de las matrices  $FC$  y  $SC$  denominada como  $X_{FC+SC}$  (ver la matriz 5.1) y para la visualización del nuevo espacio. PCA localiza las direcciones de máxima varianza en datos de alta dimensión y los proyecta en un nuevo subespacio con dimensiones iguales o menores que el original. Lo ejes ortogonales (componentes principales) del nuevo subespacio pueden interpretarse como las direcciones de máxima varianza dada la restricción de que los nuevos ejes de características son ortogonales entre sí [64].

Se construye a  $d \times k$ - matriz de transformación dimensional  $\mathbb{W}$  que nos permite mapear un vector de muestra  $\mathbf{x}$  en un nuevo subespacio de características de  $k$ -dimensión que tiene menos dimensiones que el espacio de características de  $d$ -dimensión original.

$$\mathbf{x} = [x_1, x_2, \dots, x_d], \quad \mathbf{x} \in \mathbb{R}^d.$$

$$\mathbf{y} = \mathbf{x}\mathbb{W}, \quad \mathbb{W} \in \mathbb{R}^{d \times k}, \quad d \gg k.$$

$$\mathbf{y} = [y_1, y_2, \dots, y_k], \quad \mathbf{y} \in \mathbb{R}^k.$$

Como resultado de transformar los datos originales de  $d$ -dimensión en un nuevo subespacio  $k$ -dimensión, el primer componente principal tendrá la mayor varianza posible, y todos los componentes principales consecuentes tendrán la mayor varianza dada la restricción de que estos componentes son ortogonales a los otros componentes principales, incluso si las características de entrada están correlacionadas, los componentes principales resultantes serán mutuamente ortogonales (no correlacionados) [64].

En nuestro caso cada  $x_i \in X_{FC+SC}$ , tiene un vector de dimensión  $d = 380$ , al aplicar PCA se reduce a un nuevo espacio de dimensión  $k = 3$ .  $\mathbb{W}$  contiene los vectores propios.

Cada  $x_i$  se convierte en el vector  $y_i \in \mathbb{R}^3$ , es una combinación lineal de las 380 características iniciales, preservando la mayor cantidad de varianza posible, los tres componentes principales resultantes son mutuamente ortogonales y forman el nuevo espacio denominado  $Y_{FC+SC}(161, 3)$ , es decir:

$$x_i = [m_1, m_2, \dots, m_{380}], \quad x_i \in \mathbb{R}^{380}.$$

$$y_i = x_i \mathbb{W}, \quad \mathbb{W} \in \mathbb{R}^{380 \times 3}.$$

$$y_i = [n_1, n_2, n_3], \quad y_i \in \mathbb{R}^3.$$

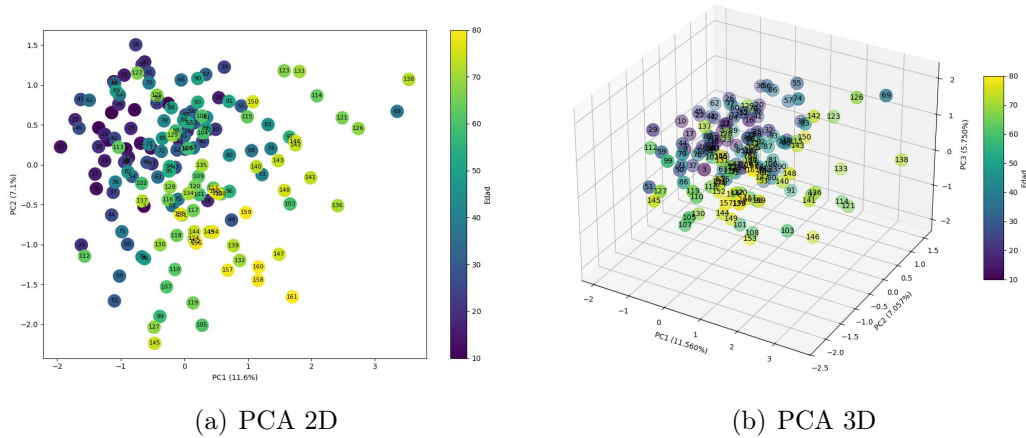


Figura 7.12: Gráficas de PCA en a)2D y b)3D del espacio denominado como:  $Y_{FC+SC}(161, 3)$ , como resultado de la reducción de la dimensionalidad, cada punto en las gráficas representa a el vector  $y_i$ , con  $i \in \{1, \dots, 161\}$ .

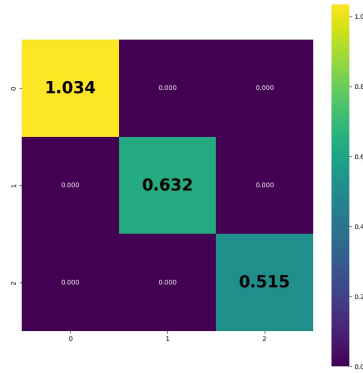


Figura 7.13: Matriz de covarianza del espacio  $Y_{FC+SC}$ , la diagonal principal corresponde a la varianza de los tres componentes principales.

El objetivo de la reducción de dimensionalidad es preservar la mayor parte posible de la estructura significativa de los datos de alta dimensión en un espacio de baja dimensión [50], es decir, de  $X_{FC+SC}(161, 380)$  a  $Y_{FC+SC}(161, 3)$ . Si bien el PCA es flexible, rápido y fácilmente interpretable, no funciona tan bien cuando existen relaciones no lineales dentro de los datos [34]. Consiguente se aplica el método t-SNE (siglas del término en inglés t-Distributed Stochastic Neighbor Embedding. Incrustación de vecinos estocásticos distribuidos en t), para la reducción de la dimensionalidad del espacio  $Y_{FC+SC}$ .

### Incrustación de vecinos estocásticos distribuidos en t. t-SNE

El método t-SNE reduce la dimensionalidad al intentar mantener las muestras similares cerca y las diferentes separadas. Se utiliza principalmente para visualización, en particular para visualizar grupos de instancias en un espacio de alta dimensión, llamado agrupamiento, que consiste en identificar muestras similares y asignarlas a grupos, es decir, grupos de muestras similares denominados clusters[5].

El método t-SNE es capaz de capturar muy bien gran parte de la estructura local de los datos de alta dimensión, al tiempo que revela la estructura global, como la presencia de clusters en varias escalas. Utiliza una distribución t de Student en lugar de una gaussiana para calcular la similitud entre dos muestras en el espacio de baja dimensión [50], es decir, la similitud entre dos puntos no es proporcional a una densidad gaussiana, sino a la de una distribución t de Student con un sólo grado de libertad (es decir, una distribución de Cauchy) [51]. El método t-SNE minimiza la divergencia entre dos distribuciones: una que mide las similitudes por pares de muestras de entrada y otra que mide las similitudes por pares de los puntos correspondientes a los de baja dimensión en la incrustación, es decir la representación de cada muestra de alta dimensión en un espacio de menor dimensión [49].

La entrada al método de t-SNE, es el espacio reducido por PCA, denotado como:  $Y_{FC+SC}(161, 3)$  con  $y_i \in \mathbb{R}^3$ , para calcular la distancia entre pares de muestras  $y_i$  se aplica una función, comúnmente la distancia euclidiana [49]. En esta tesis se aplica la distancia Manhattan (o L1) definida como:

$$d_{L1}(y_i, y_j) := \sum_{i=1}^N |y_i - y_j|, \quad (7.18)$$

donde  $N = 161$ ,  $y_i, y_j$  son dos muestras.

El propósito de t-SNE es aprender una incrustación de  $s$ -dimensión en la que cada  $y_i$  esté representado por una muestra fija,  $Z_{FC+SC}(z_1, \dots, z_N)$ , donde  $N = 161$  con  $z_i \in \mathbb{R}^s$  (valores típicos para  $s$  son 2 ó 3). En este trabajo se aplica para  $s = 2$ . Para este fin, t-SNE define probabilidades conjuntas  $p_{ij}$  que miden la similitud por pares entre las muestras  $y_i$  e  $y_j$  simetrizando dos probabilidades condicionales como sigue [49]:

$$p_{j|i} = \frac{\exp\left(-\frac{d_{L1}(y_i, y_j)^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{d_{L1}(y_i, y_k)^2}{2\sigma_i^2}\right)}, \quad p_{i|i} = 0. \quad (7.19)$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}, \quad (7.20)$$

donde  $d_{L1}$  es la distancia Manhattan,  $\sigma_i$  es la varianza gaussiana centrada en  $y_i$ , con  $N = 161$ .

En la ecuación (7.19), el ancho de banda de los núcleos gaussiano es decir  $\sigma_i$  éste se establece de tal manera que la perplejidad, definida como el número efectivo de vecinos para cada muestra, de la distribución condicional  $p_i$  coincida con un valor fijo denominado  $u$ , donde  $u$  representa el parámetro de perplejidad, con el valor de  $u = 22$ . En consecuencia, el valor óptimo de  $\sigma_i$  varía entre muestras: en regiones con alta densidad de muestras,  $\sigma_i$  tiende a ser menor que en regiones del espacio con menor densidad [49].

El espacio incrustado, denominado como:  $Z_{FC+SC}$ , la similitud de incrustación entre las muestras  $z_i$  e  $z_j$  es decir, las muestras de baja dimensión de  $y_i$  e  $y_j$ , se calcula como un t-Student normalizado con un solo grado de libertad, definido de la siguiente manera:

$$q_{ij} = \frac{(1 + \|z_i - z_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|z_k - z_l\|^2)^{-1}}, \quad q_{ii} = 0, \quad (7.21)$$

t-Student normalizado permiten modelar muestras de entrada distintos  $y_i$  e  $y_j$  mediante sus contrapartes de baja dimensión  $z_i$  e  $z_j$  que están demasiado separadas. Esto es deseable porque crea más espacio para modelar con precisión las pequeñas distancias entre pares, es decir la estructura local de las muestras, en la incrustación de baja

dimensión. Las ubicaciones de las muestras de incrustación  $z_i$  se determinan minimizando la divergencia de Kullback-Leibler entre las distribuciones conjuntas  $P$  y  $Q$  [49], definido como :

$$C(Z) = KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \quad (7.22)$$

Debido a la asimetría de la divergencia de Kullback-Leibler, la función objetivo se centra en modelar valores altos de  $p_{ij}$  (muestras similares) mediante valores altos de  $q_{ij}$  (muestras cercanas en el espacio de incrustación) La función objetivo no es convexa en la incrustación  $Z$  [49]. Normalmente se minimiza descendiendo a lo largo del gradiente:

$$\frac{\partial C}{\partial z_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) q_{ij} A(z_i - z_j), \quad (7.23)$$

donde  $A = \sum_{k \neq i} (1 + \|z_i - z_k\|^2)^{-1}$ .

Se modifica la notación de ([49] y [50]) como  $x = y_i$ ,  $y = z_i$  y  $Z = A$ .

De acuerdo a la figura 7.14 observamos que se separan la mayoría de  $z_i$  teniendo como similitud el rango de edad. Consiguiente se aplica kMeans al espacio reducido por t-SNE  $Z_{FC+SC}$ .

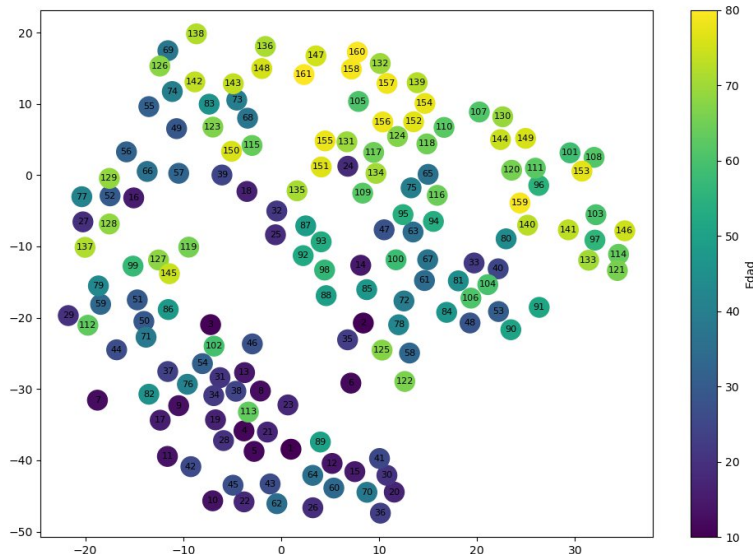


Figura 7.14: Se aplica t-SNE al espacio reducido por PCA  $Y_{FC+SC}$ . El nuevo espacio es denominado como  $Z_{FC+SC}$  con las nuevas muestras  $z_i \in \mathbb{R}^2$ . Las muestras están ordenadas de manera ascendente de acuerdo a la tabla 5.1.

## KMeans

El método de kMeans. Es un modelo que busca un número predeterminado de k clusters dentro de un conjunto de datos multidimensionales sin etiquetar [34]. Para

estimar el número óptimo de clusters  $k$ , se utiliza el método gráfico llamado método del codo, en el que intuitivamente, podemos decir que, si  $k$  aumenta, la distorsión disminuirá. Esto se debe a que las muestras estarán más cerca de los centroides a los que están asignadas. El método del codo consiste en identificar el valor de  $k$  donde la distorsión comienza a aumentar con mayor rapidez [64].

Se aplica kMeans al espacio obtenido por t-SNE, denominado  $Z_{FC+SC}$  kMeans sigue el siguiente algoritmo:

1. Se seleccionan  $k$  centroides, de las muestras del espacio  $Z_{FC+SC}$  como centros de conglomerados iniciales. Se seleccionan en número óptimo de clusters como  $k=4$ , siguiendo la figura 7.15.
2. Se asigna cada muestra al centroide más cercano, denominado como  $\mu^{(m)}$ ,  $m \in \{1, \dots, k\}$ .
3. Los centroides se mueven al centro del conjunto de muestras que le fueron asignadas.
4. Se repiten los pasos 2 y 3 hasta que las asignaciones de cluster no cambien o se alcance una tolerancia definida por el número máximo de iteraciones. Se aplicaron 10000 iteraciones.

Se cambia la notación de [64]  $j = m$ .

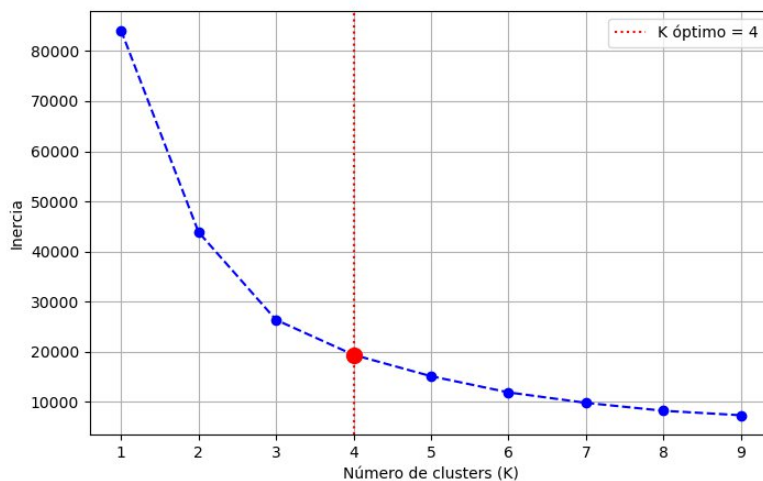


Figura 7.15:

Método gráfico para la selección de  $k$  óptimo en el método de kMeans

El fin de kMeans es agrupar las muestras según la similitud entre sus características. En este método, la similitud puede interpretarse como lo opuesto a la distancia, una

medida común para cuantificar esta distancia es la distancia euclidiana al cuadrado entre dos muestras  $z_i$  y  $z_j$  en un espacio  $s$ -dimensional:

$$d(z_i, z_j)^2 = \sum_{m=1}^s (z_{i,m} - z_{j,m})^2 \|z_i - z_j\|_2^2, \quad (7.24)$$

donde el índice  $m$  se refiere a la dimensión, (columna de características) de las muestras  $z_i$  y  $z_j$  [64]. Cada muestra se asignó a uno de los cuatro clusters. En el contexto de la agrupación en clusters, la etiqueta de una muestra es el índice del clúster al que el algoritmo la asigna. La muestra conserva una copia de kMeans de las etiquetas de las muestras con las que se entrenó [5].

### 7.2.1. Clusters.

Por consiguiente se obtienen cuatro clusters definidos por el método de kMeans, denominados como  $C_0$ , con 40 muestras,  $C_1$  con 48 muestras,  $C_2$  con 39 muestras y  $C_3$  con 34 muestras. Posteriormente se aplica el modelo anatómico funcional a dichos clusters.

Cluster-Grupo	$I_1$ (10–20)	$I_2$ (20–40)	$I_3$ (40–60)	$I_4$ (60–80)	No. Muestras $z_i$
$C_0$	5.0 %	27.5 %	40.0 %	27.5 %	40
$C_1$	43.8 %	41.7 %	10.4 %	4.2 %	48
$C_2$	10.3 %	35.9 %	12.8 %	41.0 %	39
$C_3$	2.9 %	2.9 %	8.8 %	85.3 %	34

Cuadro 7.1: Las 161 muestras están ordenadas de manera ascendente. Tabla de comparación de distribución de los índices que están en los clusters  $C_0, C_1, C_2$  e  $C_3$  con los los cuatro grupos definidos en [55]  $I_1, I_2, I_3$  e  $I_4$ .

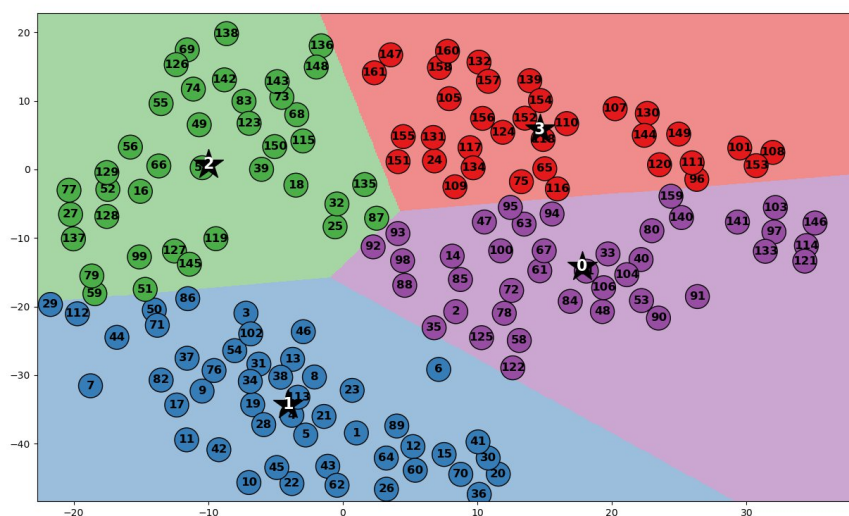


Figura 7.16: La agrupación de los clusters  $C_0, C_1, C_2$  e  $C_3$ , el centroide de cada cluster está asignado por una  $\star$ . KMeans genera estas agrupaciones mediante un proceso iterativo cuyo fin es minimizar la dispersión dentro de cada cluster. Cada centroide  $\star$  se calcula como la media de las muestras que pertenecen al cluster correspondiente. De esta manera, las muestras quedan organizadas en clusters cuyos elementos presentan una alta similitud entre sí y al mismo tiempo, una mayor diferencia respecto a los elementos de los demás clusters.

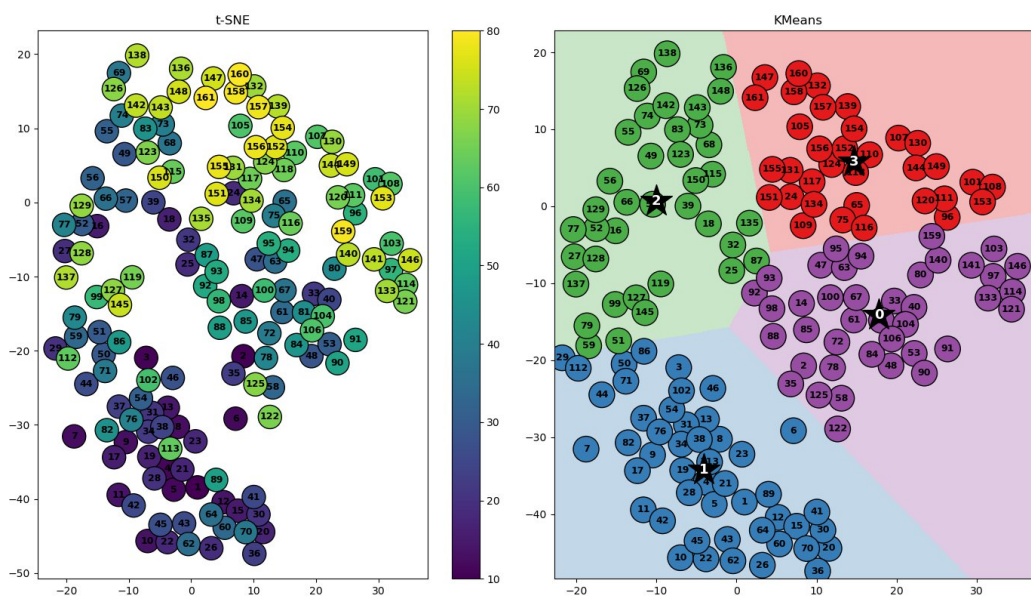
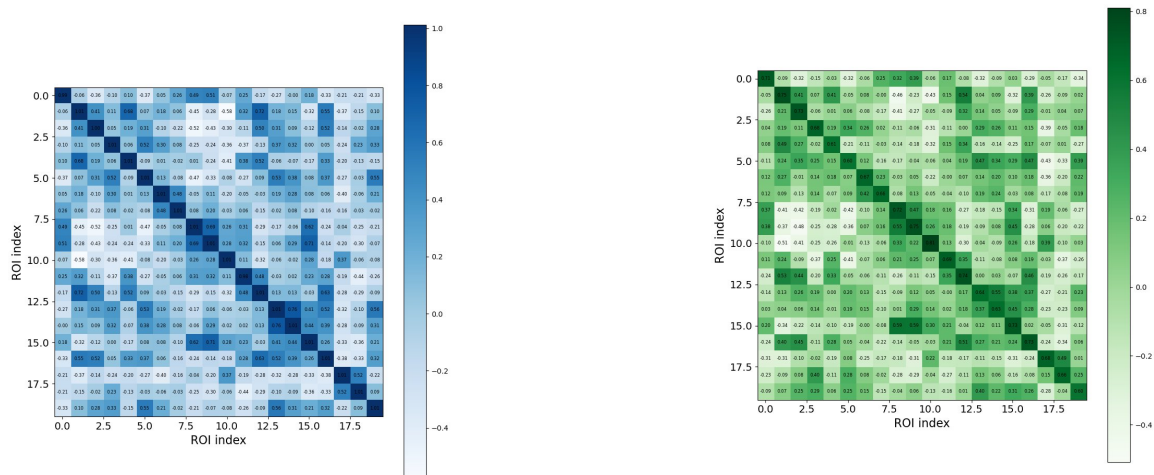


Figura 7.17: Comparación del subespacio de t-SNE con el agrupamiento obtenido por kMeans.

## 7.2.2. Modelo anatómico funcional aplicado a los clusters.

Se sigue las siguientes fases del modelo anatómico funcional a los clusters  $C_0, C_1, C_2$  y  $C_3$ : matrices de covarianza empírica, estimación de parámetros, validación, clasificación y resultados.

Matrices de covarianza empírica  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$ .



a)  $\hat{S}(0)_{ij}$  Matriz de conectividad funcional (FC) sin desfase temporal.

b)  $\hat{S}(1)_{ij}$  Matriz de conectividad funcional (FC) con desfase temporal de uno.

Figura 7.18: De acuerdo a las ecuaciones (7.1) y (7.2), se obtienen las matrices de covarianza empírica sin y con desfase temporal  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$  aplicado a los clusters.

## Estimación de parámetros.

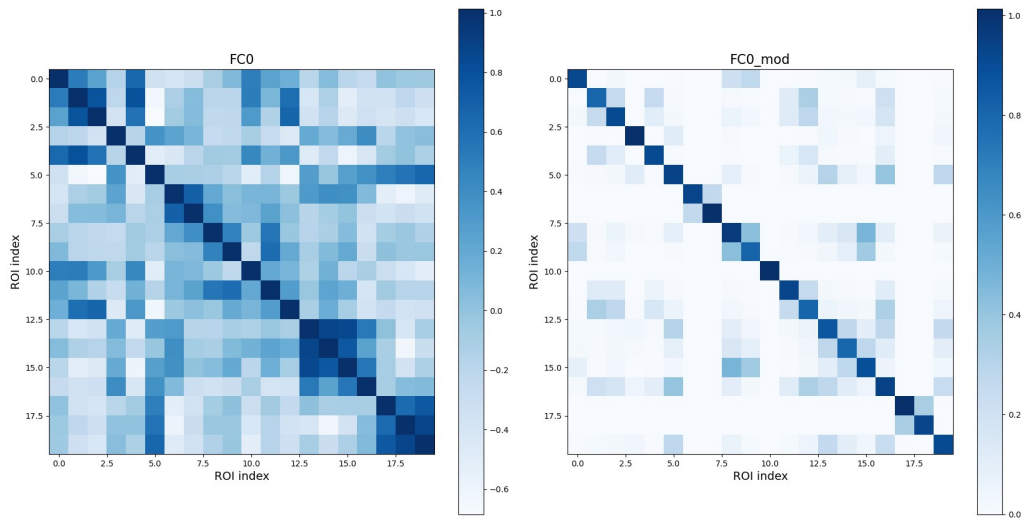


Figura 7.19: El ajuste del modelo anatómico funcional consiste en estimar dos parámetros: la matriz Jacobiana  $C_{ij}$ , que describe la conectividad entre ROIs, y la matriz de covarianza de entrada  $D_{ij}$ , que representa el ruido. A partir de éstos se realiza la optimización (ver 7.6), consiguiendo se obtiene la matriz de covarianza teórica  $S(0)_{ij}$ , la cual permite describir las correlaciones entre los ROIs del modelo.

## Validación del modelo

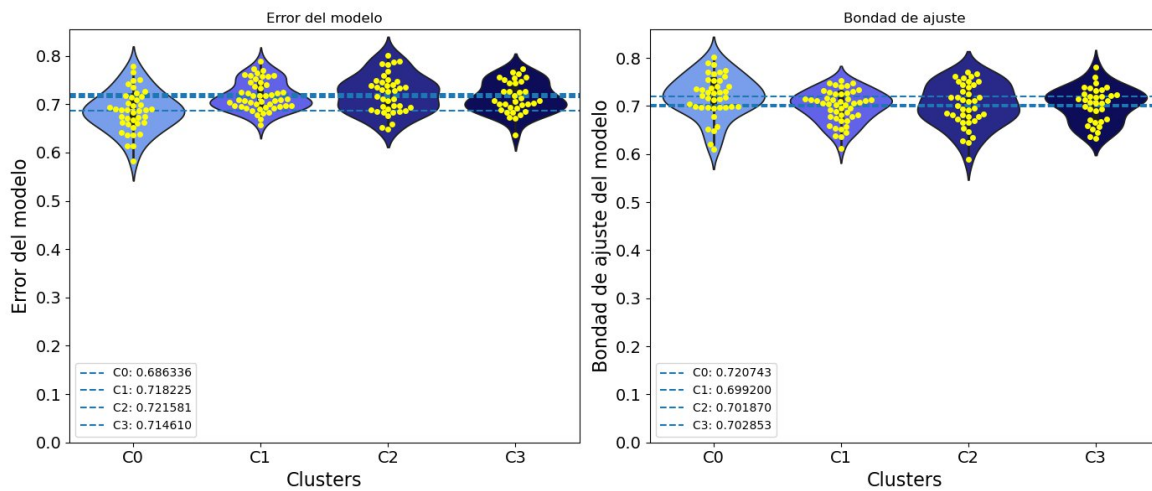
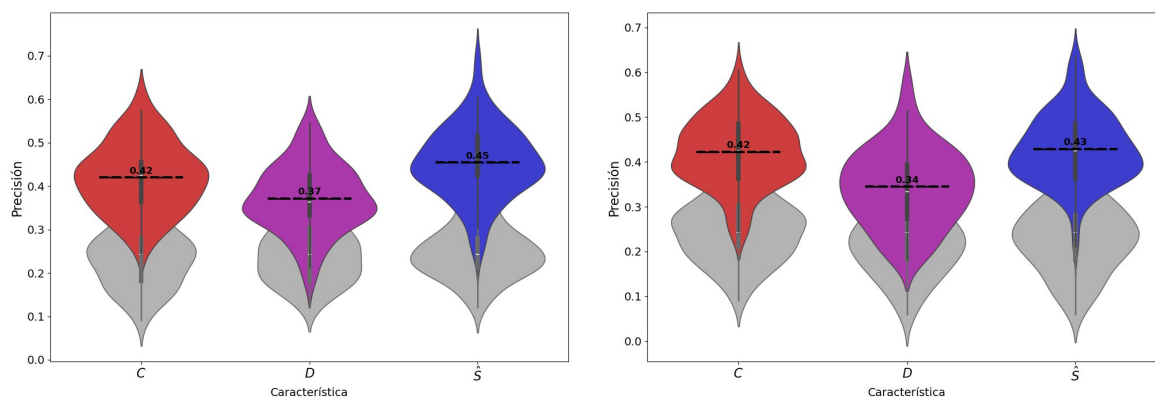


Figura 7.20: El error del modelo y la bondad de ajuste permiten evaluar qué tan bien el modelo reproduce la matriz teórica  $S(0)_{ij}$ , obtenida a partir de los parámetros estimados. El modelo anatómico funcional presenta un buen desempeño al capturar una parte importante de las covarianzas empíricas  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$ , con medias entre 0.68 y 0.72.

## Clasificación



a) Clasificador de regresión logística multinomial.

b) Clasificador 1-NN.

Figura 7.21: La clasificación se realiza para validar que la dinámica estimada contenga información discriminadora real y no basada en el azar. Para este fin, se toman las características  $C$  (conectividad),  $D$  (ruido) y  $\hat{S}$  (covarianza). Se observa que las medias se encuentran por encima del valor obtenido con shuffle, que representa la aleatoriedad, (representado por los violines de color gris), lo que indica que el modelo logra discriminar entre las características reales y el azar. Comparando con la gráfica 7.9, para los clusters se reconocen mejor las características, en particular para la covarianza  $\hat{S}$ .

### 7.3. Aprendizaje de máquina supervisado.

Los biomarcadores pueden definirse como características explicativas que presentan mayor relevancia en el desempeño de modelos de aprendizaje de máquina y aprendizaje profundo, por lo que constituyen indicadores cuantificables. En este contexto, dichos biomarcadores se extraen a partir de características presentes en los datos de conectividad funcional (FC) y conectividad estructural (SC). Con este propósito, en la tercera etapa se desarrollan modelos de aprendizaje de máquina supervisado. Dos enfoques basados en bosques aleatorio: Random Forest Classifier y Random Forest Regressor. Asimismo, se desarrollan cinco tipos de arquitecturas de aprendizaje profundo: redes neuronales densas (Deep Neural Networks, DNN), redes neuronales convolucionales unidimensionales (Convolutional Neural Networks 1D, CNN-1D), redes neuronales convolucionales bidimensionales (Convolutional Neural Networks 2D, CNN-2D), redes neuronales recurrentes (Recurrent Neural Networks, RNN) y un modelo híbrido, con el fin de identificar biomarcadores relevantes, sobre los cuales posteriormente se aplica el modelo anatómico funcional.

De acuerdo al análisis realizado por los métodos de aprendizaje no supervisado: PCA, t-SNE y kMeans, el proceso queda representado como:

$X_{FC+SC}(161, 380)$  a  $Y_{FC+SC}(161, 3)$  a  $Z_{FC+SC}(161, 2)$  se obtiene  $C_n$ .

Los cuatro clusters se utilizan como etiquetas que sirven como salida a los modelos de aprendizaje supervisado. Éstos aprenden a asociar una entrada con una salida conocida a partir de un conjunto de entrenamiento. Dicho conjunto corresponde al subconjunto de datos que se emplea para que el algoritmo identifique patrones y establezca la relación entre las variables de entrada y las etiquetas asignadas [28].

### Random Forest Classifier

Random Forest es un modelo no paramétrico de aprendizaje supervisado basado en un método de ensamblaje. Este se fundamenta en estimadores simples denominados árboles de decisión. Un estimador es una función que, a partir de los datos de entrada, aproxima o predice una variable de salida aprendiendo patrones presentes en los datos. Cada árbol divide iterativamente las muestras a lo largo de un eje según algún criterio cuantitativo; en cada nivel se asigna la etiqueta de la nueva región de acuerdo con el valor mayoritario de las muestras contenidas en ella [34]. Random Forest puede entenderse como la suma de funciones lineales por partes, es decir, se subdivide el espacio de entrada en regiones más pequeñas que se vuelven más manejables [64]. Se implementan dos enfoques del algoritmo Random Forest: uno para clasificación (Random Forest Classifier) y otro para regresión (Random Forest Regressor).

La entrada al modelo Random Forest Classifier es  $X_{FC+SC}$  (ver 5.1 y 5.2) y la salida son los cuatro clusters  $C_0, C_1, C_2$  y  $C_3$ . El estimador utilizado es el RandomForestClassifier, con 20000 números de estimadores o árboles de decisión, con un 80 % de entrenamiento y 20 % de prueba. Para evaluar el rendimiento del modelo se utiliza la matriz de confusión.

### Matriz de confusión.

La matriz de confusión. Es una matriz cuadrada que reporta los conteos de predicciones de verdaderos positivos (TP true positive ), verdaderos negativos (TN true negative), falsos positivos (FP false positive) y falsos negativos (FN false negative) [64].

Tanto el error de predicción (prediction error ERR) como la precisión (accuracy ACC) proporcionan información general sobre cuántas muestras se obtienen. El error puede entenderse como la suma de todas las predicciones falsas dividida entre el número total de predicciones, y la precisión se calcula como la suma de las predicciones correctas dividida entre el número total de predicciones, respectivamente:

$$ERR = \frac{FP + FN}{FP + FN + VP + VN}, \quad (7.25)$$

$$ACC = 1 - ERR. \quad (7.26)$$

La tasa de positivos verdaderos y la tasa de positivos falsos se denominan respectivamente TPR y FPR (TPR true positive rate y FPR false positive rate) [64].

$$FPR = \frac{FP}{FP + TN}, \quad (7.27)$$

$$TPR = \frac{TP}{FN + TP}. \quad (7.28)$$

La métrica de precisión (PRE), está definida como:

$$PRE = \frac{TP}{TP + FP}. \quad (7.29)$$

La métrica de recall (REC) es igual a la de  $TPR$ . Y la métrica F1-Score está definida como:

$$F1 = 2 \frac{PRE \cdot REC}{PRE + REC}. \quad (7.30)$$

Siguiendo las ecuaciones (7.28), (7.29) y (7.30) se obtiene precisión, recall y F1-score. Para cada clase, en este contexto cluster, tenemos que en  $C_0$ , detecta 6 de 7 muestras, en  $C_1$ , 9 de cada 9 muestras, en  $C_2$ , 6 de 11 y en  $C_3$ , 3 de 6 muestras. Hay una precisión global de 0.73 (ver tabla 7.22).

	precision	recall	f1-score	support
0	0.6	0.86	0.71	7.0
1	0.64	1.0	0.78	9.0
2	1.0	0.55	0.71	11.0
3	1.0	0.5	0.67	6.0
accuracy	0.73	0.73	0.73	0.73
macro avg	0.81	0.73	0.72	33.0
weighted avg	0.82	0.73	0.72	33.0

Figura 7.22: Tabla en el que se describe el desempeño de clasificación.

La matriz de confusión muestra el desempeño del randomforestclassifier, mostrando la correspondencia entre las clusters reales y predichas. Para el cluster 0, 6 muestras se predicen correctamente, mientras que 1 muestra se clasifica erróneamente como cluster 1. En el cluster 1, 9 muestras se predicen correctamente. Para el cluster 2, 6 muestras se clasifican correctamente, 4 se clasifican erróneamente como cluster 1 y 1 muestra como cluster 0. Finalmente, en el cluster 3, 3 muestras se clasifican correctamente y 3 se clasifican erróneamente como cluster 0. Los elementos de la diagonal principal corresponden a las muestras clasificadas correctamente. En total, de 33 muestras evaluadas, el modelo clasifica correctamente 24 (ver figura 7.23).

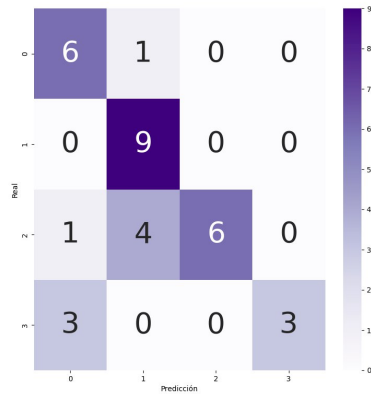


Figura 7.23: Los elementos de la diagonal principal representan el número de muestras clasificadas correctamente, por cluster, mientras que los elementos fuera de la diagonal indican las muestras clasificadas erróneamente.

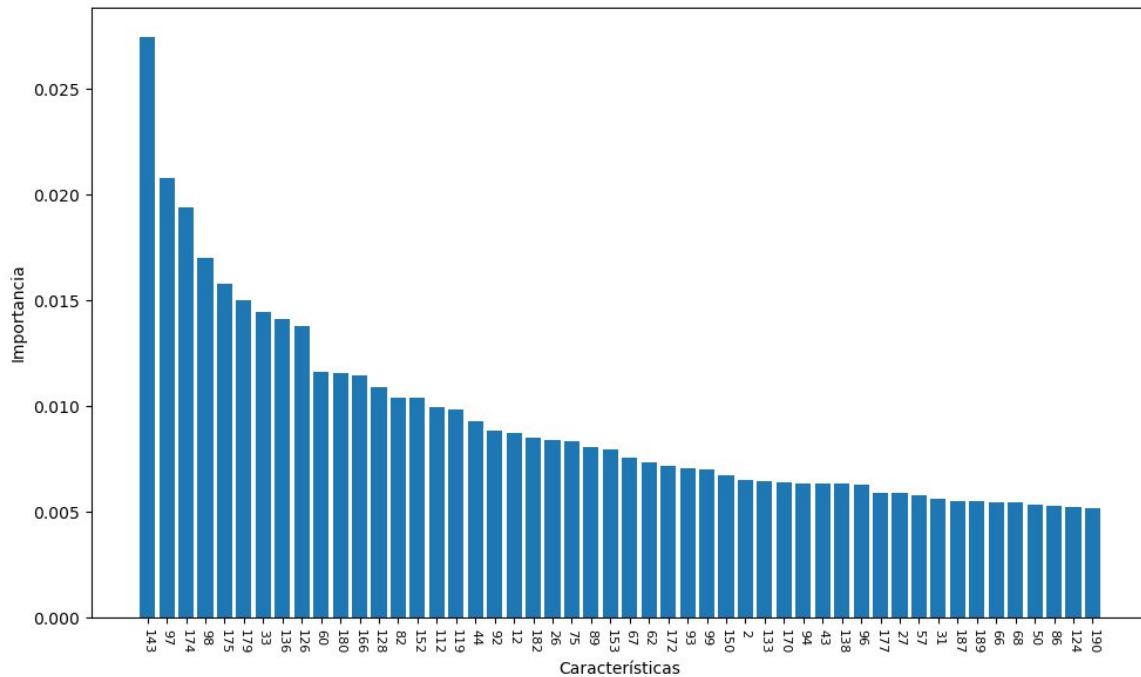


Figura 7.24: En Random Forest la importancia de las características se determina evaluando cuánto contribuye cada variable a la reducción en los nodos de decisión de los árboles. La importancia de cada característica se obtiene promediando dicha reducción a lo largo de todos los árboles [64].

### SHapley Additive exPlanations (SHAP) aplicado a modelos de aprendizaje supervisado.

Es un enfoque de teoría de juegos para explicar el resultado de cualquier cualquier modelo de aprendizaje de máquina. Conecta la asignación óptima de créditos con explicaciones locales utilizando los valores clásicos de Shapley de la teoría de juegos y sus extensiones relacionadas [44]. A partir de la matriz de muestras de entrada  $X_{FC+SC}$  (5.1), la cual se compone de 380 características, 190 funcionales y 190 estructurales.

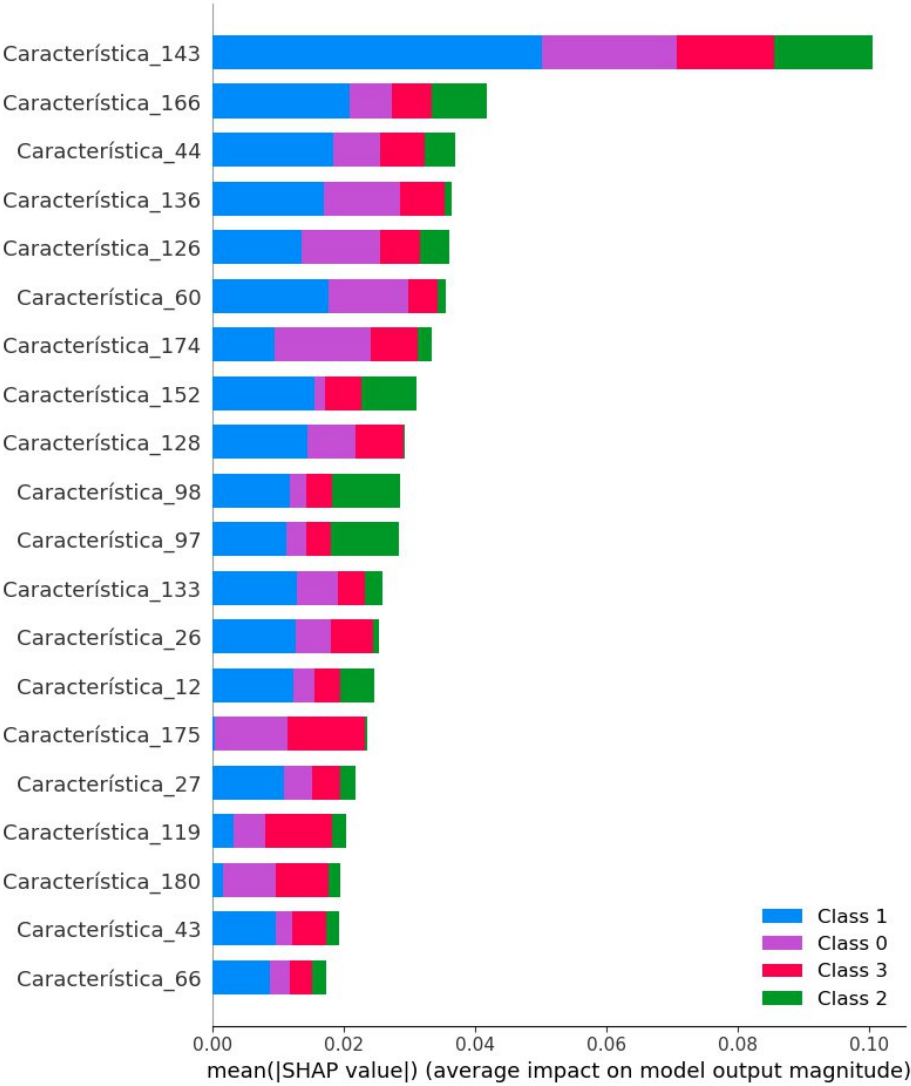
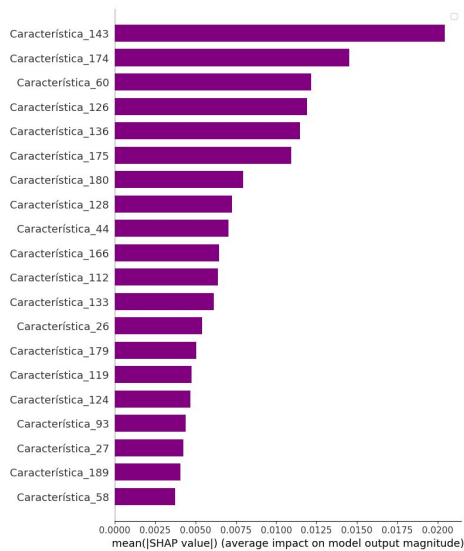
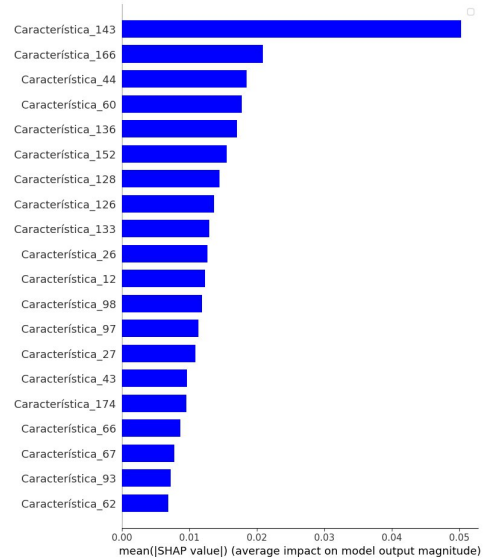


Figura 7.25: Gráfica general de barras de los valores Shapley para los cuatro clusters  $C_n$ . En ella se muestra, en orden descendente, la importancia de las características que contribuyen a la clasificación realizada por el modelo Random Forest Classifier. De acuerdo a la matriz  $X_{FC+SC}$  (5.2), se observa que sólo contribuyen características funcionales, además el cluster  $C_1$ , es predominante, el cual se compone de un 43.8% de sujetos de edad entre 10 y 20 años.



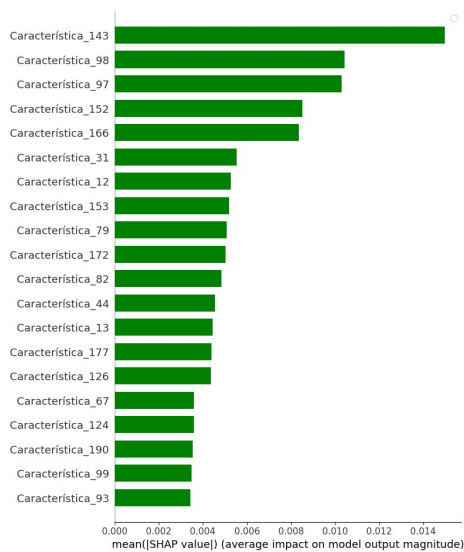
(a) Cluster 0

Los índices pertenecen a las características funcionales son: 174[14,20], 175[15,16], 33[2,17], 143[10,19] y 136[10,12].



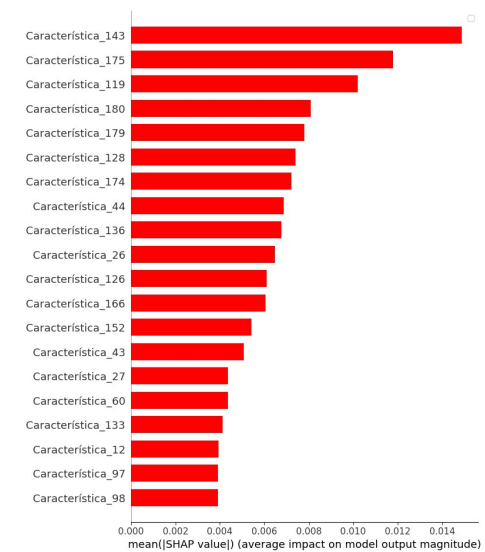
(b) Cluster 1

Los índices pertenecen a la FC las características: 143[10,19], 136[10,12], 97[6,19], 60[4,11] y 44[3,11].



(c) Cluster 2

Los índices pertenecen a la FC las características: 97[6,19], 98[6,20], 143[10,19], 82[5,18] y 33[2,17].



(d) Cluster 3

Los índices pertenecen a la FC las características: 174[14,20], 179[15,20], 175[15,16], 180[16,17] y 119[8,16].

Figura 7.26: ROIs con mayor contribución explicativa en la clasificación realizada por el modelo Random Forest. Donde donde los valores entre corchetes  $[i, j]$  indican los ROIs de la matriz de conectividad funcional (FC) de acuerdo al índice, donde del 1 al 190 son funcionales y del 191 al 380 son estructurales.

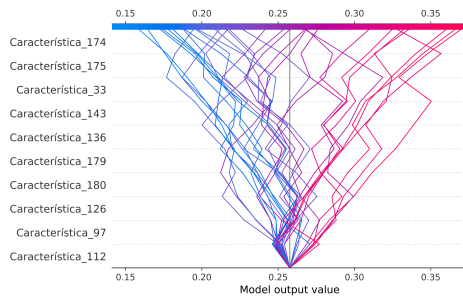
Cluster	Total de sujetos	Edad predominante	Sujetos predicción	$E[C_n]$	Característica	Descripción
<b>C0</b>	40	40% de 60-80 años	10	0.261	174 [01110202]	ROI 14 y ROI 20
<b>C1</b>	48	43.8% de 10-20 años	14	0.311	143 [11112111]	ROI 10 y ROI 19
<b>C2</b>	39	41.08% de 60-80 años	6	0.2178	97 [0222231]	ROI 6 y ROI 19
<b>C3</b>	34	85.3% de 60-80 años	3	0.2092	174 [01110202]	ROI 14 y ROI 20

Figura 7.27: En la tabla se muestra, para cada cluster, el número total de sujetos, la edad predominante, los sujetos predichos por el modelo Random Forest Classifier y el valor de  $\mathbb{E}$ , utilizado como  $\phi_0$  en el cálculo de los valores SHAP, los cuales se representan en las gráficas de decisión correspondientes a cada cluster (ver 7.28). Asimismo, se indica la característica dominante en cada cluster. Por ejemplo, en el cluster 1 la característica predominante es la 143; en este caso, 7 de los 8 sujetos con dicha característica son impulsados hacia el cluster 1. Asimismo a que ROIs pertenecen dichas características.

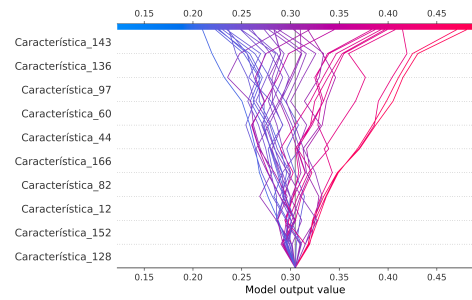
La propiedad de precisión local establece que la predicción puede descomponerse en la suma del valor base y las contribuciones individuales de cada característica [47].

$$f(x) = g(x') = \phi_0 + \sum_{i=1}^M \phi_i x'_i,$$

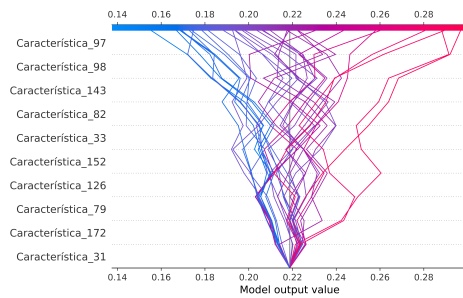
donde:  $f(x)$  representa la predicción del modelo original.  $g(x')$  corresponde al modelo de explicación. Las variables  $x'_i$  representan variables simplificadas, donde  $x'_i = 1$  si la característica  $i$  está presente y  $x'_i = 0$  si está ausente. El término  $\phi_i$  representa la contribución de la característica  $i$  a la predicción del modelo,  $\phi_0$  corresponde al valor base del modelo, es decir, la predicción cuando ninguna característica está presente. En este contexto, dicho valor base corresponde al del modelo Random Forest Classifier, y  $M$  representa el número total de características,  $M = 380$ , de las cuales 190 corresponden a características funcionales y 190 a características estructurales.



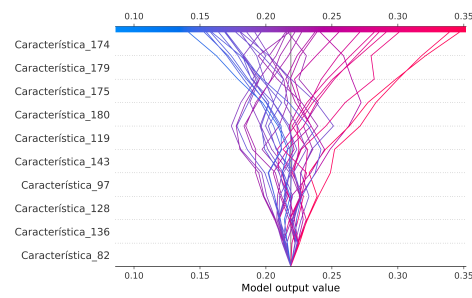
(a) Cluster 0



(b) Cluster 1



(c) Cluster 2



(d) Cluster 3

Figura 7.28: Los valores SHAP muestran cómo cada característica contribuye a la predicción del modelo Random Forest Classifier, desplazando la clasificación desde el valor base —correspondiente a la predicción promedio del modelo antes de considerar las características— hasta la predicción final para una o varias observaciones. En la gráfica, el color rojo indica valores altos de la característica y el azul valores bajos. La dirección de las líneas indica su efecto en la predicción de pertenencia a un cluster: cuando  $\phi_i > 0$  la probabilidad aumenta mientras que cuando  $\phi_i < 0$  disminuye.

## Random Forest Regressor

El segundo método de aprendizaje supervisado es el Random Forest Regressor. El estimador que se utiliza para este fin es el RandomForestRegressor [34]. La entrada al modelo Random Forest Regressor es  $X_{FC+SC}$  (ver 5.1 y 5.2) y la salida es el vector de edades de los 161 sujetos. Con 10000 número de estimadores, con un 80% de entrenamiento y 20% de prueba.

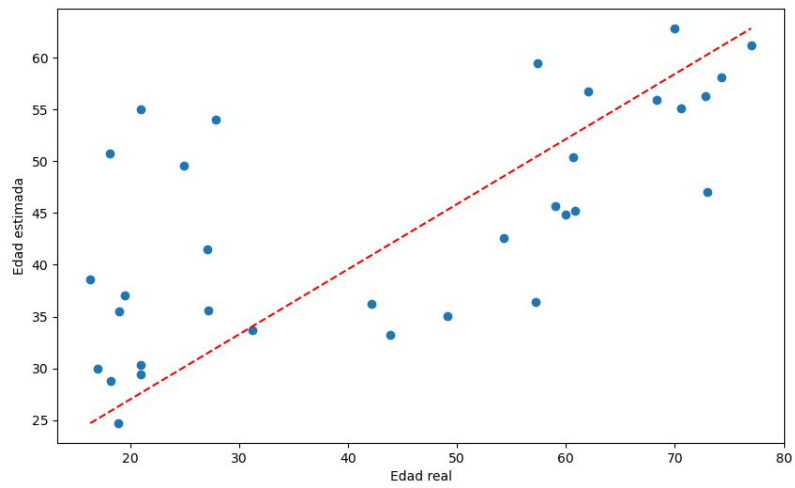


Figura 7.29: Random Forest Regressor

La gráfica muestra mayor cercanía de predicción en las edades de 60 a 80 años, en comparación con edades de 10 a 30 años.

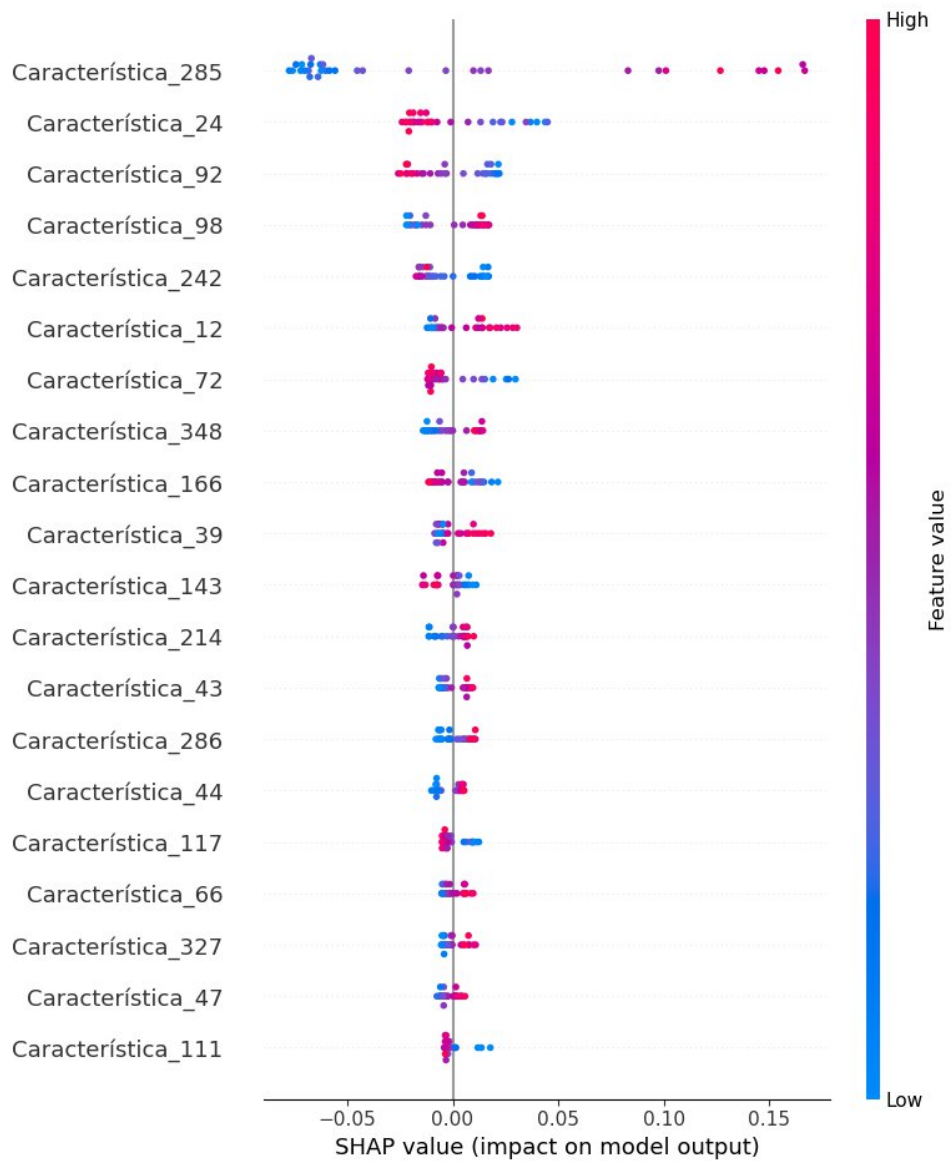


Figura 7.30: De acuerdo a la matriz de muestras de entrada  $X_{FC+SC}$  Los índices corresponden a las características: 285 [16,6], perteneciente a la SC, 24 [2,8] y 92 [6,14] a la FC.

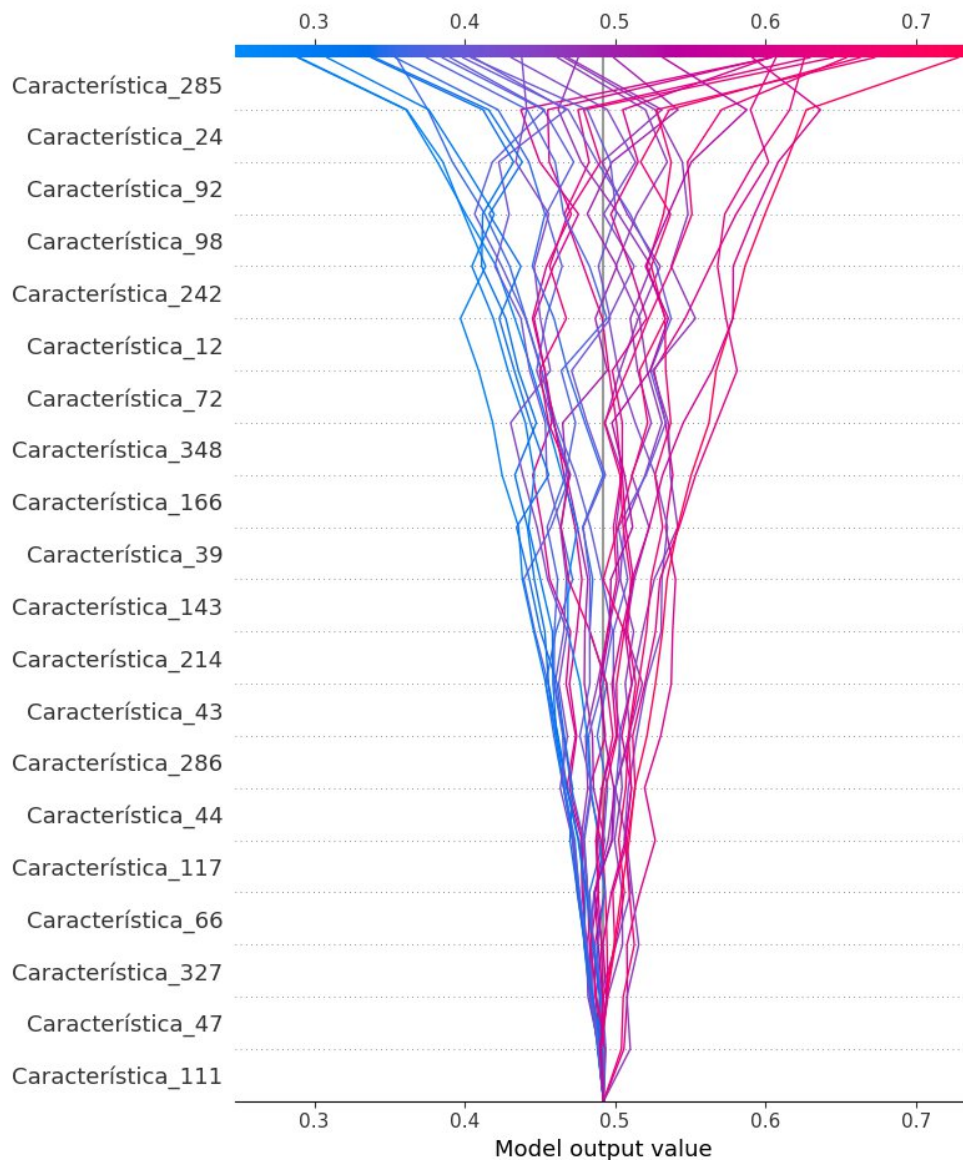


Figura 7.31: Siguiendo una descripción análoga al modelo Random Forest Classifier (ver 7.31), la gráfica de decisión muestra las características predominantes en el modelo Random Forest Regressor. Para éste modelo las características estructurales tienen mayor contribución a la predicción.

Después de entrenar los dos modelos basados en bosques aleatorios Random Forest Classifier y Random Forest Regressor y calcular los valores de Shapley, es decir, las características explicativas que contribuyen a la decisión del modelo, se identificaron los siguientes ROIs predominantes: para el Random Forest Classifier, 6, 10, 14, 19 y 20, todas pertenecen a la conectividad funcional y para el Random Forest Regressor, 6 y 16 perteneciente a la conectividad estructural y 2, 6, 8 y 14 pertenecen a la conectividad funcional. Posteriormente se implementan las cinco arquitecturas de los modelos de aprendizaje profundo.

### 7.3.1. Aprendizaje profundo.

Para los modelos de aprendizaje profundo se emplea Keras, es una interfaz de programación de aplicaciones (API) de aprendizaje profundo `multibackend` escrita en python diseñada para ser simple, flexible y eficiente. Esta biblioteca permite construir, entrenar e implementar modelos utilizando distintos motores de cómputo, como TensorFlow, JAX o PyTorch.

Las estructuras de datos principales en Keras son las capas. Una capa constituye el componente fundamental de una red neuronal en Keras. Se trata de un objeto invocable que define una función de cálculo entre tensores de entrada y de salida, y que además mantiene un estado interno representado por las variables de TensorFlow, es decir, los pesos de la capa[56].

La entrada a los modelos de aprendizaje profundo son las series BOLD (por sus siglas en inglés, Blood Oxygen Level Dependent, dependientes del nivel de oxigenación sanguínea) representada como la matriz (161, 20, 159) donde 161 corresponde a las muestras, 20 a las regiones de interés (ROIs) y 159 a los tiempos de repetición (TR) y como salida, los clusters obtenidos por kMeans (ver 7.16), distribuidos de la siguiente manera:  $C_0$  con 40 muestras,  $C_1$  con 48 muestras,  $C_2$  con 39 muestras y  $C_3$  con 34 muestras.

Para el entrenamiento de redes neuronales, se emplean funciones de no linealidad y funciones de pérdida. Para éste trabajo se emplea la función softmax que permite calcular probabilidades de los clusters  $C_n$ , que son significativas en entornos multiclase (regresión logística multinomial) [71]. La función softmax asigna a una muestra con entrada neta  $z$  una probabilidad  $p(y = i|z)$  de pertenecer a la clase (cluster)  $C_n, i$ , mediante un término de normalización que corresponde a la suma de todas las  $M$  salidas lineales, se describe mediante la siguiente ecuación:

$$p(y = i|z) = f(z) = \frac{e^{z_i}}{\sum_{j=1}^M e^{z_j}}. \quad (7.31)$$

Para el proceso de aprendizaje, se utiliza una función de pérdida, que cuantifica el error entre las predicciones del modelo y las etiquetas reales, siendo el valor que la red intenta minimizar durante el entrenamiento [56]. En este trabajo se emplea la función Categorical Focal Crossentropy, utilizada para clasificación multiclase.

### Redes neuronales densas (Deep Neural Networks, DNN).

Se construyen cinco modelos densos distintos, incrementando progresivamente tanto los parámetros como la complejidad estructural. En cada modelo, la arquitectura se caracteriza por conexiones totalmente conectadas es decir, cada neurona de una capa está conectada con todas las neuronas de la capa precedente. La operación fundamental

de una neurona densa se puede expresar como:

$$y = f(wx + b), \quad (7.32)$$

donde  $x$  representa el vector de entrada,  $W$  la matriz de pesos que determina la influencia de cada entrada,  $b$  es el término de sesgo que permite desplazar la función de activación y  $f$  es la función de activación aplicada a la salida lineal. Las funciones de activación introducen la no linealidad al modelo, logrando así que la red aprenda relaciones entre las variables [28]. En los modelos densos se utilizaron tres funciones de activación: ReLU, tanh y LeakyReLU.

La función ReLU denominada así por su nombre en inglés Rectified Linear Unit se define como:

$$f(z) = \begin{cases} 0, & \text{si } z < 0, \\ z, & \text{si } z > 0, \end{cases} \quad (7.33)$$

donde  $z$  es la entrada de la neurona antes de aplicar la función de activación.

La Función tangente hiperbólico (tanh) se expresa como:

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}. \quad (7.34)$$

donde  $z$  es la entrada ponderada de la neurona. La función tanh normaliza la salida en el rango de  $[-1, 1]$ , lo que centra los datos y mejora la convergencia.

La función LeakyReLU es una variante de ReLU.

$$f(z) = \begin{cases} \alpha z, & \text{si } z < 0, \\ z, & \text{si } z \geq 0, \end{cases} \quad (7.35)$$

donde  $z$  es la entrada de la neurona y  $\alpha$  es un valor pequeño positivo.

	Modelo	Pérdida de entrenamiento	Precisión de entrenamiento	Pérdida de validación	Precisión de validación
0	Dense_one	0.089	0.823	0.263	0.563
1	Dense_two	0.178	0.724	0.529	0.306
2	Dense_three	0.044	0.874	0.108	0.653
3	Dense_four	0.09	0.825	0.261	0.585
4	Dense_five	0.06	0.829	0.165	0.486

Figura 7.32: El modelo la mayor precisión en validación y menor pérdida es el Dense-four con 0.585 en precisión y 0.261 en pérdida. La capacidad del modelo puede discriminar patrones de conectividad funcional asociados a los clusters  $C_n$ .

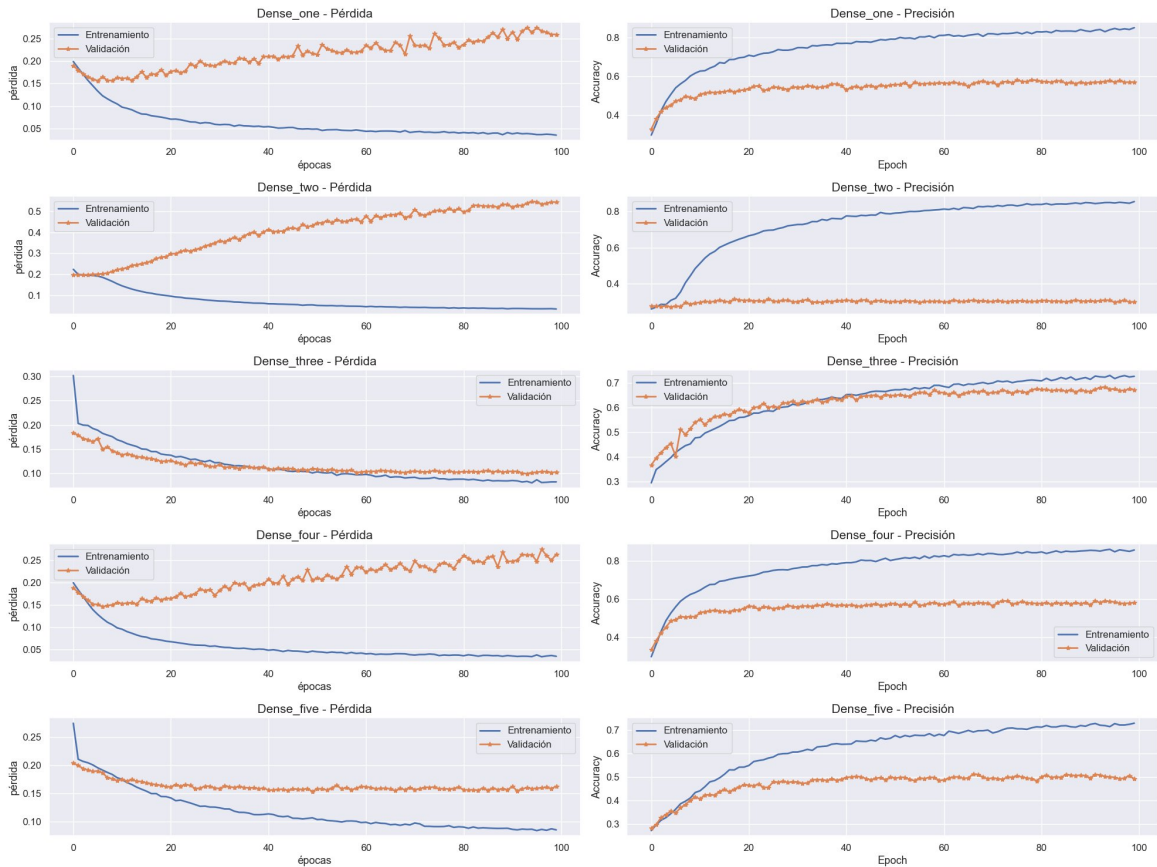


Figura 7.33: Las funciones de activación utilizadas en los distintos modelos fueron las siguientes: Dense-one: ReLU con un aumento progresivo de neuronas con 64, 128, 256 neuronas en cada capa. Dense-two: tanh con neuronas distribuidas en 64 y 128. Dense-three: ReLU con capas concatenadas que incluyen 16, 64, 128 y 56 neuronas. Dense-four: LeakyReLU con un incremento de neuronas en las capas 64, 128, 256 y el Dense-five: ReLU y tanh con neuronas 16, 64, 128 y 256.

## Redes neuronales convolucionales unidimensionales (Convolutional Neural Networks 1D, CNN-1D).

Los modelos convolucionales CNN-1D, así denominados por su nombre en inglés Convolutional Neural Network, en una dimensión, es el modelo que utiliza filtros deslizantes que recorren los datos para extraer patrones locales de manera jerárquica. En el caso de datos 1D, las convoluciones permiten localizar patrones temporales o secuenciales [1], mientras que el convolucional 2D, las convoluciones detectan patrones espaciales, siendo ampliamente utilizadas en procesamiento de imágenes [56].

En este trabajo se desarrollaron tres arquitecturas con diferente tamaño de kernel, es decir ventana, y la función de activación ReLU definida por la ecuación (7.33). La siguiente ecuación (7.36) define la operación de convolución 1D:

$$y(t) = \sum_{k=0}^{k-1} w_k \cdot x(t+k). \quad (7.36)$$

donde  $x(t+k)$  es la entrada,  $w_k$  son los pesos,  $K$  es el tamaño del Kernel, que define cuántos elementos consecutivos de la entrada se consideran para la salida  $y(t)$  [28].

	Modelo	Pérdida de entrenamiento	Precisión de entrenamiento	Pérdida de validación	Precisión de validación
0	Conv1D_one	0.00229	0.993	0.00652	0.97845
1	Conv1D_two	0.00088	0.99825	0.00253	0.99127
2	Conv1D_three	0.00034	0.99891	0.002	0.99521

Figura 7.34: El modelo con mayor precisión en validación y menor pérdida es el *Conv1D-three* con 0.99521 en precisión y 0.002 en pérdida. El modelo captura patrones locales y secuenciales en series temporales, lo que permite identificar características diferenciables de la dinámica de las series BOLD que hacen posible distinguir efectivamente los clusters  $C_n$ .

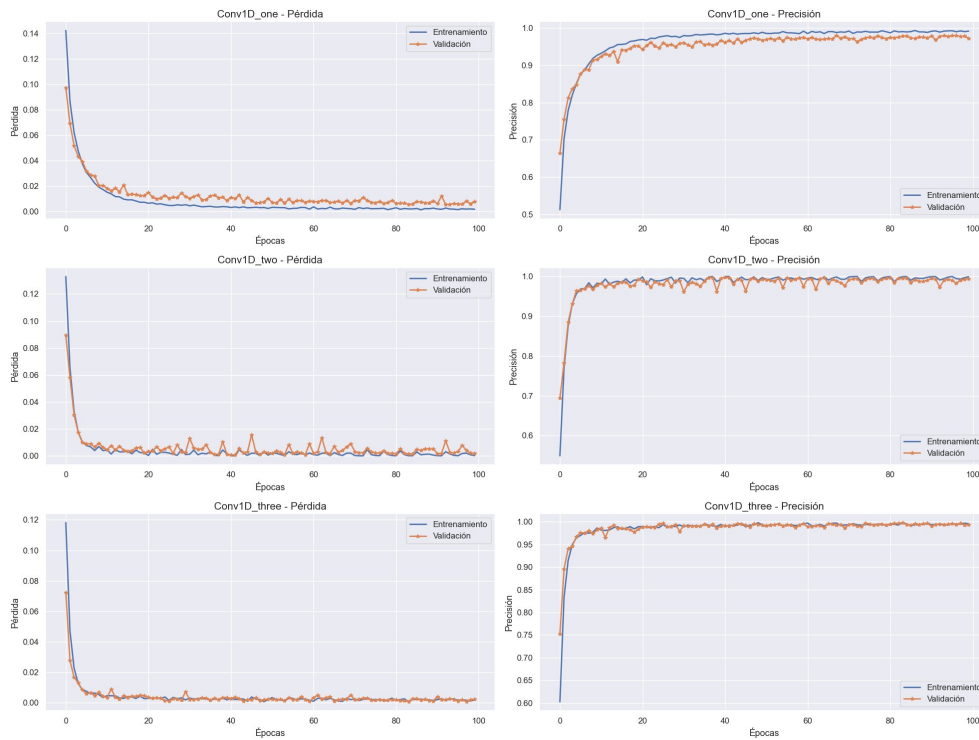


Figura 7.35: La función de activación es la misma para los tres modelos. El tamaño de kernel y número de neuronas es la siguiente: Conv1D-one: 3, 5 con 64 y 128 neuronas, Conv1D-two: 1,3 y 5 con 64,128 y 256 neuronas y Conv1D-three: 2,3,5,3,5,2 con 64, 128 y 256 neuronas, para éste caso se usaron las capas concatenadas.

## Redes neuronales convolucionales bidimensionales (Convolutional Neural Networks 2D, CNN-2D).

Los modelos convolucionales CNN-2D, así denominados por su nombre en inglés Convolutional Neural Network, en una dos dimensiones, es un tipo de red neuronal diseñada para procesar datos con estructura espacial bidimensional, como imágenes. A diferencia de las redes densas, que conectan cada neurona con todas las entradas, las CNN 2D explotan la correlación local entre píxeles o valores adyacentes mediante filtros deslizantes [28]. El tamaño de kernel se define como:

$$Y_{i,j} = \sum_{m=0}^{k_h-1} \sum_{n=0}^{k_w-1} W_{m,n} \cdot X_{i+m,j+n} \quad (7.37)$$

donde  $X$  es la entrada,  $k_h \times k_w$  es el tamaño de kernel,  $W_{m,n}$  son los pesos del filtro y  $(i, j)$  son las posiciones de la entrada. Se desarrollan tres diferentes modelos CNN-2D con diferentes tamaños de kernel y la función de activación ReLU definida en la ecuación (7.33).

	Modelo	Pérdida de entrenamiento	Precisión de entrenamiento	Pérdida de validación	Precisión de validación
0	Conv1D_one	0.00229	0.993	0.00652	0.97845
1	Conv1D_two	0.00088	0.99825	0.00253	0.99127
2	Conv1D_three	0.00034	0.99891	0.002	0.99521

Figura 7.36: El modelo con mayor precisión en validación y menor pérdida es el Conv-two con 0.978 en precisión y 0.007 en pérdida. El modelo captura patrones espaciales y temporales simultáneamente, los diferentes tamaños de kernel permiten detectar patrones en la actividad BOLD.

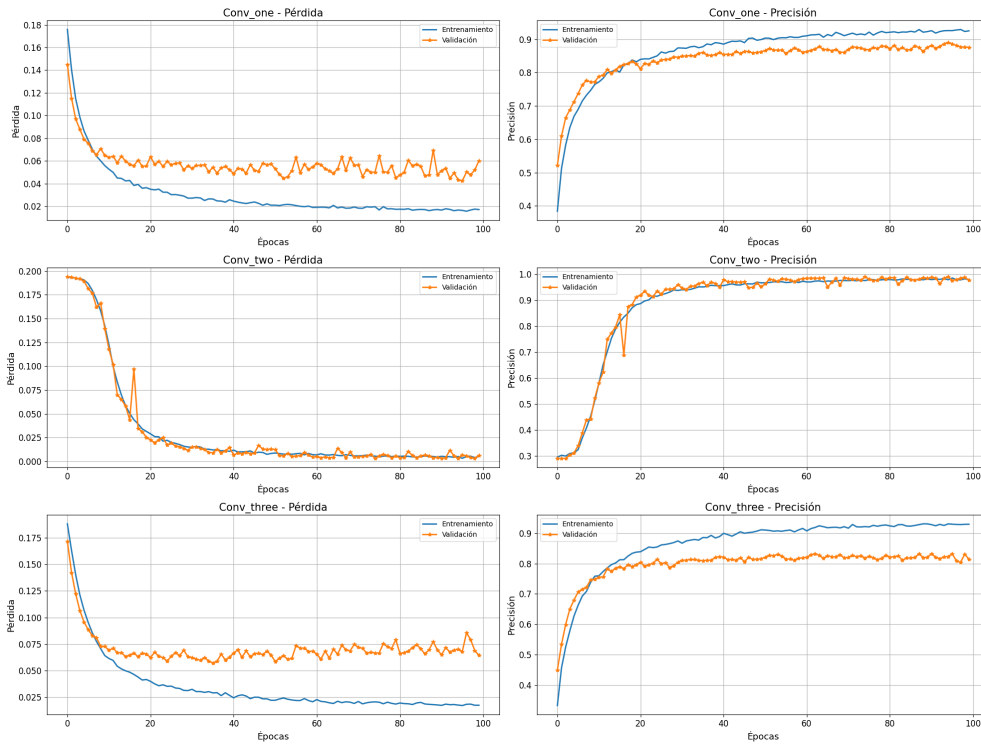


Figura 7.37: La arquitectura de los modelos tienen los diferentes tamaños de kernel: Conv-one [5,3], [3,1] y una capa densa. Conv-two con [3,1], [5,3], [5,7] y varias capas densas concatenadas y Conv-three [3,1], [5,3], [7,5] con varias capas densas concatenadas.

### Redes neuronales recurrentes (Recurrent Neural Networks, RNN).

Los modelos recurrentes (RNN) procesan datos secuencias manteniendo la memoria del pasado en el orden de los datos [56]. Se realizan tres diferentes arquitecturas de modelos con la misma función de activación ReLU, definida por la ecuación (7.33). Y con el uso de la capa LSTM denominada así por su término en inglés Long Short Term Memory, diseñada para procesar secuencias de datos y superar las limitaciones de desvanecimiento del gradiente [71].

	Modelo	Pérdida de entrenamiento	Precisión de entrenamiento	Pérdida de validación	Precisión de validación
0	CNN_one	0.00033	0.99922	0.00185	0.9962
1	CNN_two	0.00058	0.99837	0.00273	0.99465
2	CNN_three	0.00035	0.99873	0.00136	0.99662

Figura 7.38: El modelo con mayor precisión y menor pérdida es el CNN-one, por su arquitectura mantiene la memoria de estados pasados para capturar la dinámica de las series BOLD, lo que permite extraer representaciones espaciales entre ROIs como evolución temporal, lo que permitió diferenciar a los sujetos en cada cluster  $C_n$

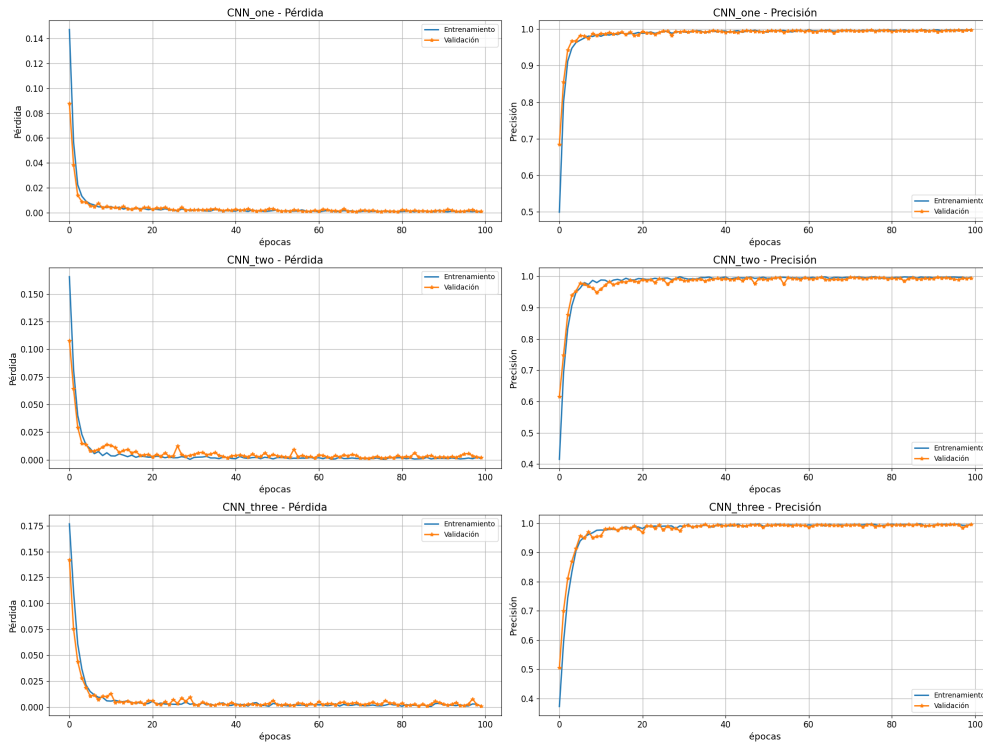
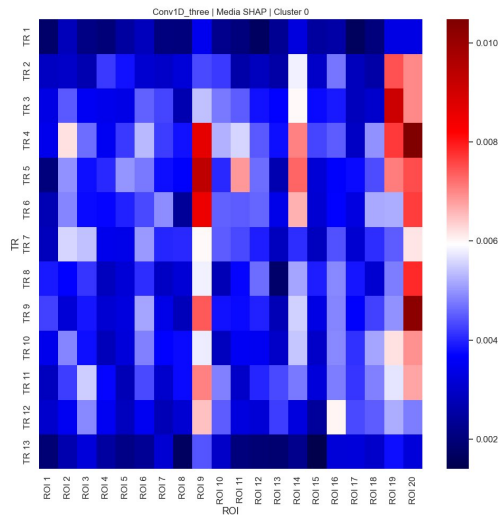


Figura 7.39: La arquitectura de los modelos son las siguientes: CNN-one: Tres capas convoluciones con los tamaños de kernel [2,1],[3,2],[4,3] una capa densa y una LSTM, CNN-two: Tres capas convolucionales 2D con los tamaños de kernel [2,1][3,2],[5,3] concatenadas, dos capas LSTM y dos densas y el CNN-Three: Tres capas convolucionales 2D con los tamaños de kernel [2,1][3,1][5,3] concatenadas, una LSTM, dos convolucionales 1D y dos densas.

### SHapley Additive exPlanations (SHAP) aplicado a CNN-1D y RNN

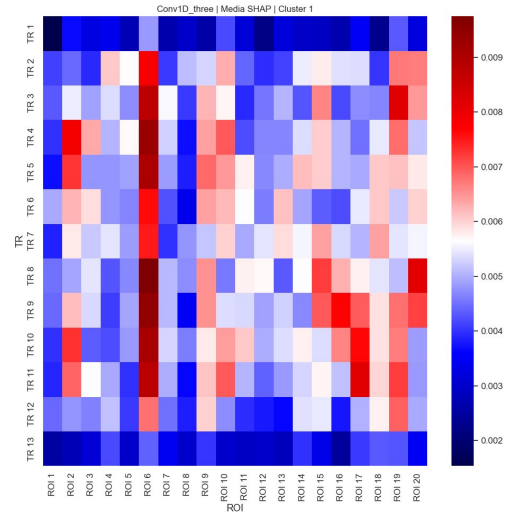
Para interpretar los modelos con mayor precisión y menor pérdida, se seleccionaron Conv1D-three y RNN-one; con este fin, se aplicó SHAP para analizar la importancia de las características de entrada y la matriz de confusión para evaluar la correspondencia entre las predicciones del modelo y las muestras asociadas a cada cluster  $C_n$ .

## SHAP para el modelo convolucional unidimensional. CNN-1D



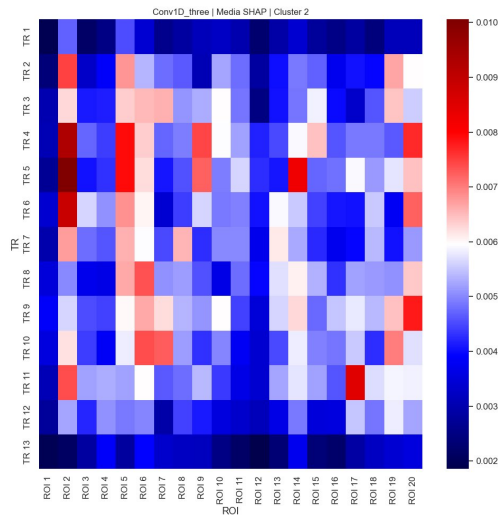
(a) Cluster  $C_0$

ROI 20 con 0.006836, ROI 9 con 0.006359, ROI 19 con 0.005775 y ROI 14 con 0.005116.



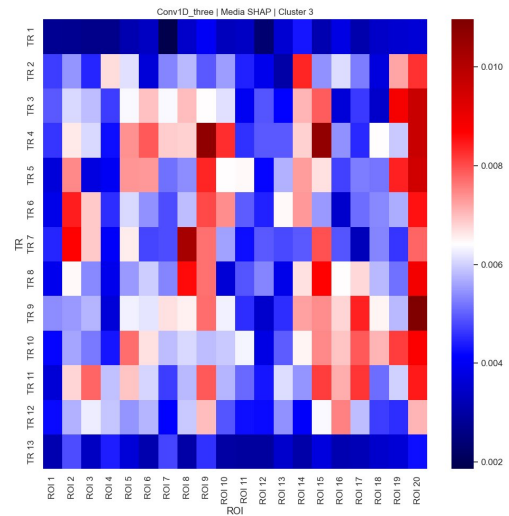
(b) Cluster  $C_1$

ROI 6 con 0.007929, ROI 19 con 0.006137, ROI 9 con 0.005687, ROI 2 con 0.005660 y ROI 10 con 0.005609.



(c) Cluster  $C_2$

ROI 2 con 0.006537. ROI 5 con 0.006027, ROI 6 con 0.005853, ROI 20 con 0.005819 y ROI 14 con 0.005429.

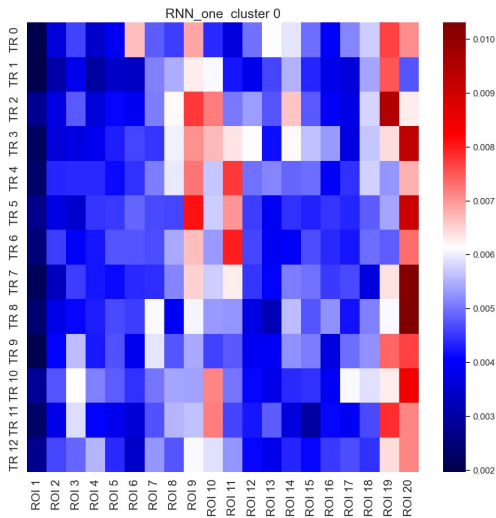


(d) Cluster  $C_3$ .

ROI 20 con 0.008105, ROI 9 con 0.006976, ROI 15 con 0.006844, ROI 8 con 0.006192 y ROI 2 con 0.006182.

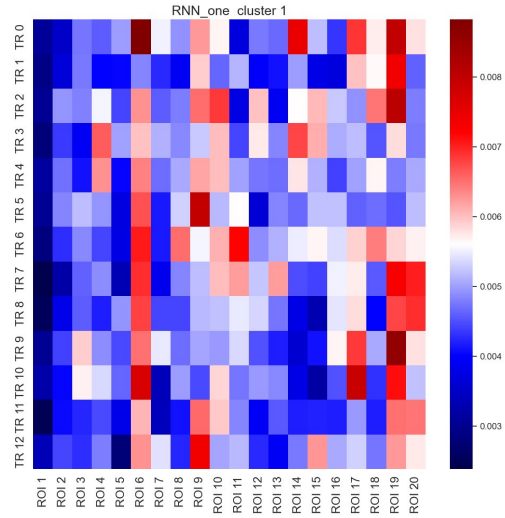
Figura 7.40: Para cada cluster, se calcula el valor absoluto de los valores SHAP y posteriormente se obtiene su media considerando todos los sujetos. En cada matriz resultante, el color rojo indica las características con mayor contribución a la decisión del modelo, mientras que el color azul representa aquellas con una menor influencia.

## SHAP para el modelo recurrente. RNN



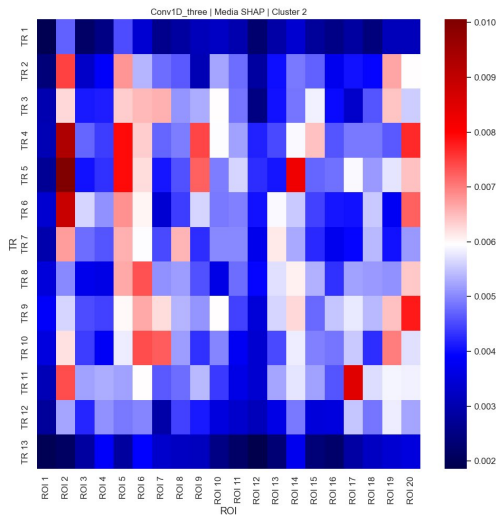
(a) Cluster  $C_0$

ROI 20 con 0.007812, ROI 19 con 0.006702, ROI 9 con 0.006545 y ROI 10 con 0.005921.



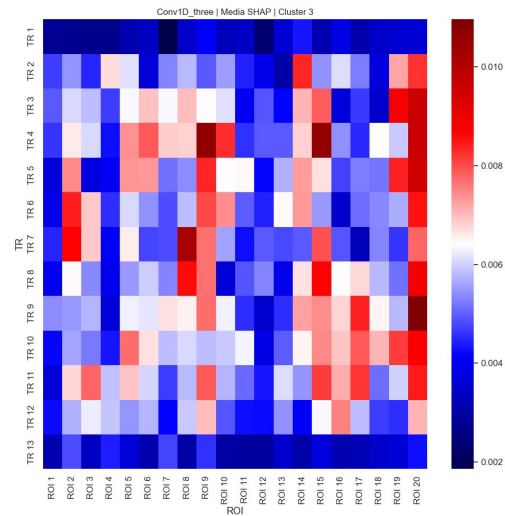
(b) Cluster  $C_1$

ROI 19 con 0.006687, ROI 6 con 0.006644, ROI 9 con 0.005950, ROI 17 con 0.005775 y ROI 10 con 0.005653.



(c) Cluster  $C_2$

ROI 20 con 0.006990, ROI 19 con 0.006208, ROI 6 con 0.005964, ROI 7 con 0.005957 y ROI 18 con 0.005269.

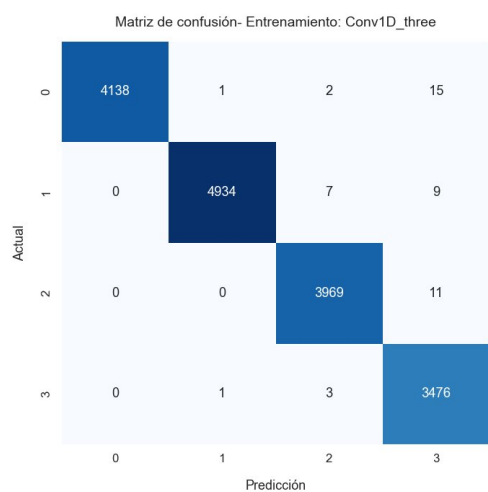


(d) Cluster  $C_3$ .

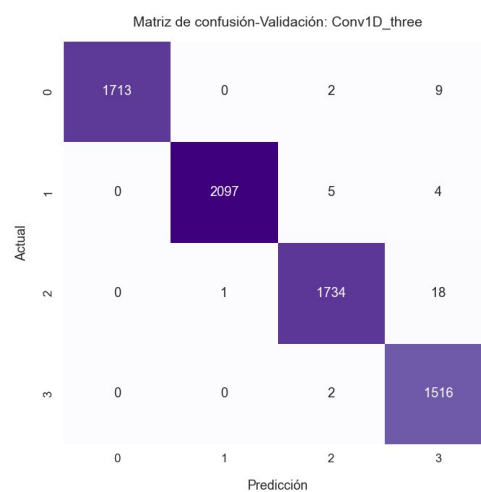
ROI 10 con 0.006163, ROI 18 0.005514, ROI 19 con 0.005494, ROI 15 con 0.005126 y ROI 11 con 0.004823.

Figura 7.41: Para cada cluster, se calcula el valor absoluto de los valores SHAP y posteriormente se obtiene su media considerando todos los sujetos. En cada matriz resultante, el color rojo indica las características con mayor contribución a la decisión del modelo, mientras que el color azul representa aquellas con una menor influencia.

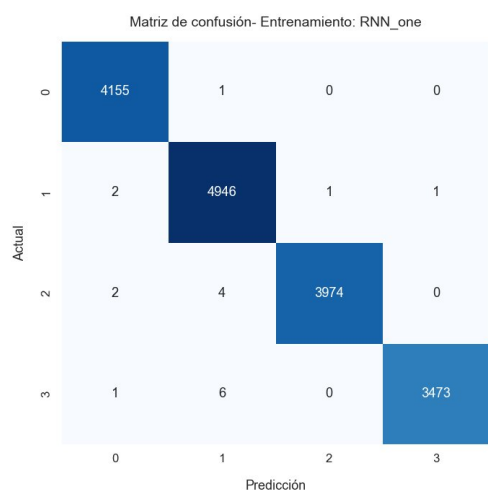
## Matrices de confusión para los modelos de CNN-1D y RNN.



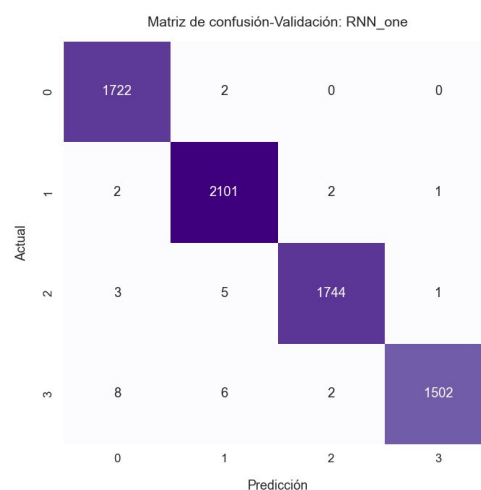
(a) Entrenamiento para el modelo convolucional 1D



(b) Validación para el modelo convolucional 1D



(c) Entrenamiento para el modelo recurrente



(d) Validación para el modelo recurrente

Figura 7.42:

a) La diagonal de la matriz de confusión indica los aciertos del modelo, mientras que las filas representan los clusters reales y las columnas los clusters predichos. Para el cluster 0, el modelo clasifica correctamente 4155 muestras, con un único error asignado al cluster 1. El cluster 1 se clasifica correctamente en 4946 casos, con 2 errores en el cluster 0, 1 en el cluster 2 y 1 en el cluster 3. En el cluster 2, se aciertan 3974 muestras, con 2 errores asignados al cluster 0 y 4 al cluster 1. Finalmente, el cluster 3 presenta 3473 aciertos, con 1 error en el cluster 0 y 6 en el cluster 1. Se aplica la misma lógica para b), c) y d).

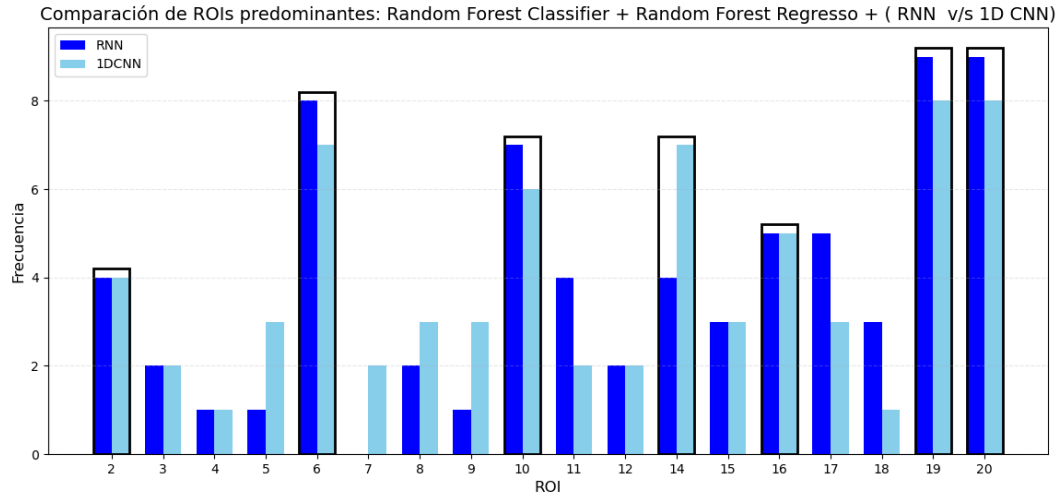


Figura 7.43: Conjunto de 7 ROIs como resultado de la alta relevancia explicativa y capacidad discriminativa frente al resto de las demás ROIs, en los modelos de aprendizaje y aprendizaje profundo.

### Modelo Híbrido.

Finalmente se aplica un modelo convolucional unidimensional con entrada las series BOLD con dimensión (23667,13,20) y dos salidas: una de regresión, correspondiente a la tasa de producción de entropía  $\Phi$  y otra de clasificación asociada a los clusters. El desempeño del modelo se evalúa mediante las correlaciones de Pearson con 0.9318, Spearman con 0.9362 y Kendall  $\tau$  con 0.81031, para la salida de regresión, mientras que para la salida de clasificación se utilizó la matriz de confusión, para determinar la precisión en la asignación de clusters  $C_n$ .

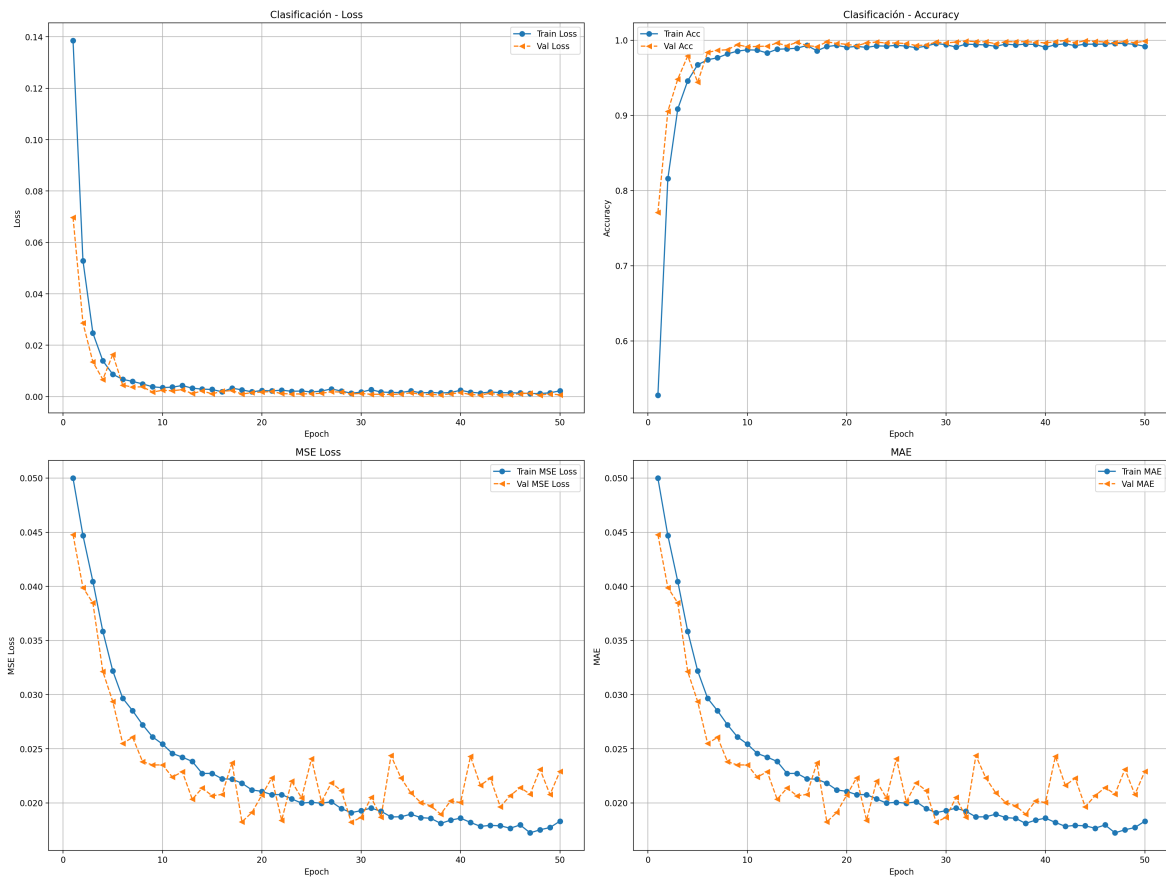


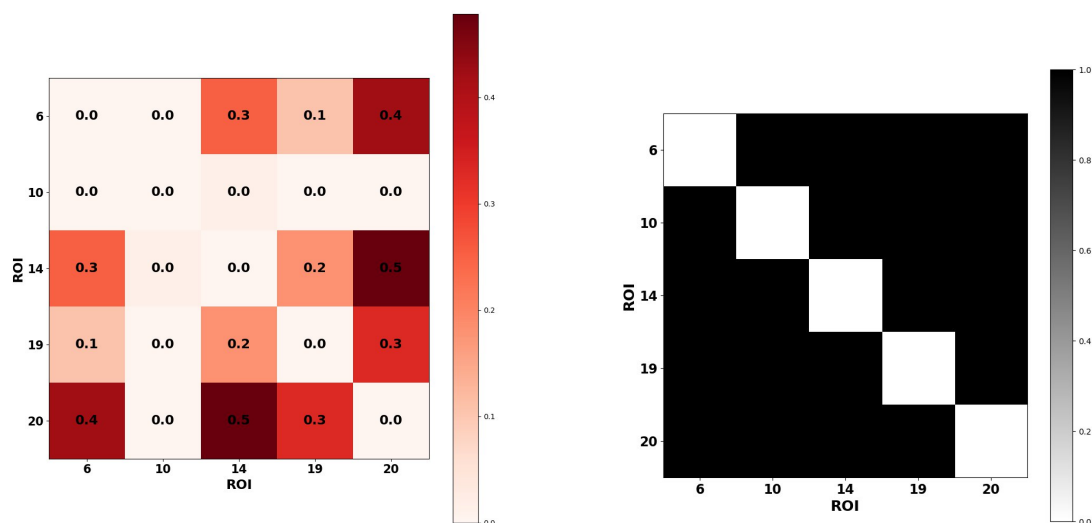
Figura 7.44: En la de clasificación, el modelo alcanzó una precisión en validación de 0.9985, con una pérdida en validación de 0.000885, lo que evidencia una alta capacidad de discriminación.

Para la regresión, el desempeño se evaluó mediante las métricas de Error Absoluto Medio (MAE) y Error Cuadrático Medio (MSE). En el conjunto de validación se obtuvieron valores de  $MAE = 0.022$  y  $MSE = 0.022395$ , lo que indica errores promedio bajos y una distribución homogénea de las predicciones.

### 7.3.2. Biomarcadores.

De acuerdo a la métricas de desempeño de los modelos de aprendizaje supervisado y aprendizaje profundo, es decir SHAP y la matriz de confusión. Se determinaron biomarcadores predominantes, las regiones de interés (ROIs) 2, 6, 10, 14, 16, 19 y 20. Posteriormente, se aplica el modelo anatómico funcional considerando tres variantes: *(i)* utilizando los siete biomarcadores, *(ii)* empleando únicamente los cinco biomarcadores predominantes 6, 10, 14, 19 y 20, y *(iii)* considerando las quince ROIs restantes que no pertenecen al conjunto de biomarcadores. Este análisis se realiza tanto para los grupos, como para los clusters.

### 7.3.3. Modelo anatómico funcional aplicado a biomarcadores.

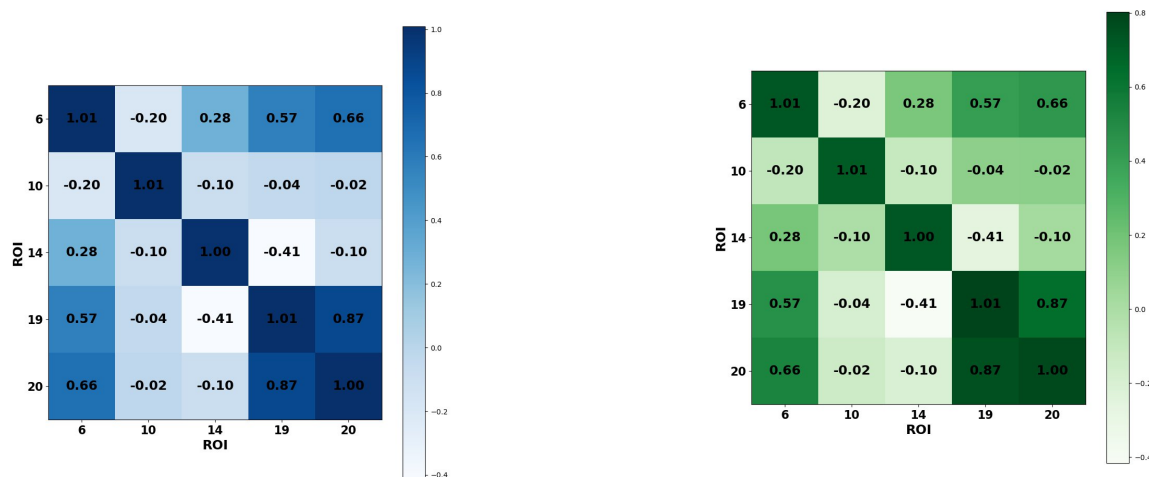


a) Media de la matriz de conectividad estructural (SC) de los 161 sujetos, la cual forma un esqueleto.

b) Máscara booleana con una densidad de conexiones del 100%.

Figura 7.45: Matriz de conectividad estructural (SC) de 5 ROIs y su correspondiente máscara booleana.

Matrices de covarianza empírica  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$ .



a)  $\hat{S}(0)_{ij}$  Matriz de conectividad funcional (FC) sin desfase temporal.

b)  $\hat{S}(1)_{ij}$  Matriz de conectividad funcional (FC) con desfase temporal de uno.

Figura 7.46: De acuerdo a las ecuaciones (7.1) y (7.2), se obtienen las matrices de covarianza empírica sin y con desfase temporal  $\hat{S}(0)_{ij}$  y  $\hat{S}(1)_{ij}$  aplicado a los biomarcadores.

## Estimación de parámetros.

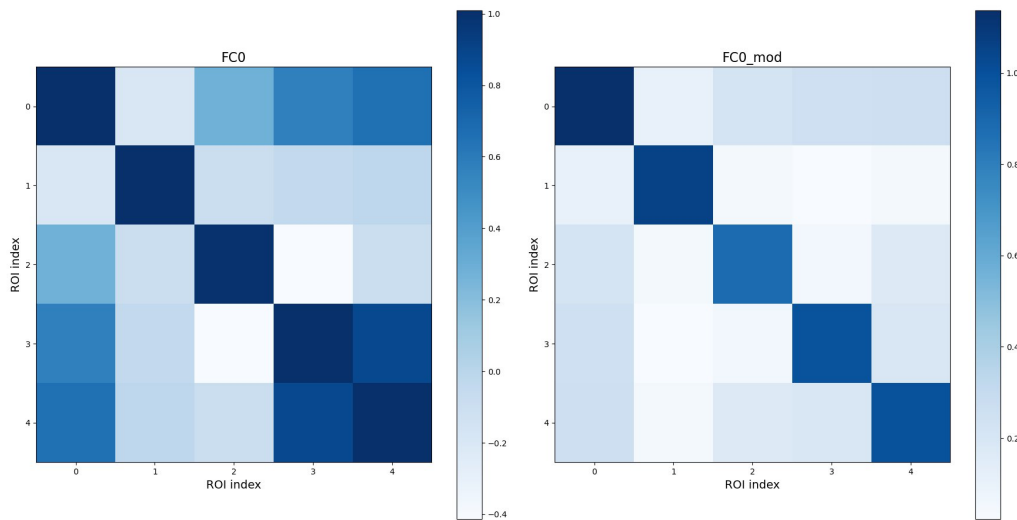
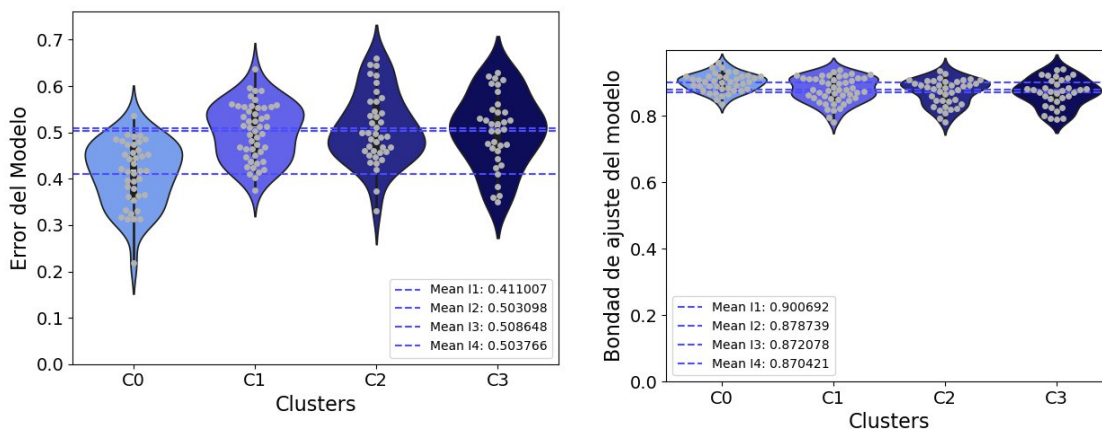


Figura 7.47: El ajuste del modelo anatómico funcional consiste en estimar dos parámetros: la matriz Jacobiana  $C_{ij}$ , que describe la conectividad entre ROIs, y la matriz de covarianza de entrada  $D_{ij}$ , que representa el ruido. A partir de éstos se realiza la optimización (ver 7.6), consiguiendo se obtiene la matriz de covarianza teórica  $S(0)_{ij}$ , la cual permite describir las correlaciones entre los ROIs del modelo.

## Validación del modelo

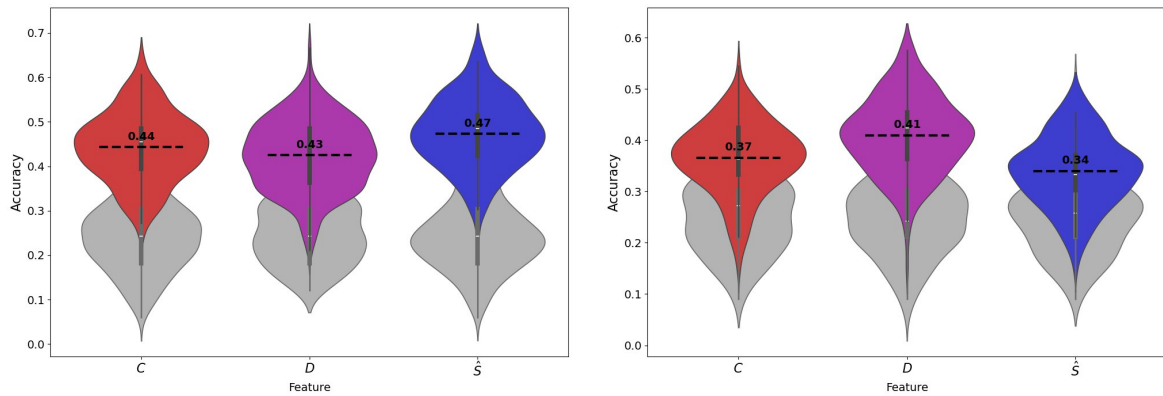


a) Error del modelo para biomarcadores.

b) Bondad de ajuste del modelo para biomarcadores.

Figura 7.48: El error del modelo y la bondad de ajuste permiten evaluar qué tan bien el modelo reproduce la matriz teórica  $S(0)_{ij}$ , obtenida a partir de los parámetros estimados. En bondad de ajuste presenta medias entre 0.86 a 0.89.

## Clasificación



a) Clasificador de regresión logística multinomial.

b) Clasificador 1-NN.

Figura 7.49: La clasificación se realiza para validar que la dinámica estimada contenga información discriminadora real y no basada en el azar. Para este fin, se toman las características  $C$  (conectividad),  $D$  (ruido) y  $\hat{S}$  (covarianza). Se observa que las medias se encuentran por encima del valor obtenido con shuffle, que representa la aleatoriedad, (representado por los violines de color gris), lo que indica que el modelo logra discriminar entre las características reales y el azar.

La interpretación de resultados de las tres etapas, se presentan en el siguiente capítulo.

# 8

## Resultados y Discusión.

Los resultados obtenidos son los siguientes. En la primera etapa, las medias de la tasa de producción de entropía  $\Phi$  para cada grupo son: para  $I_1$ , conformado por 28 sujetos de 10 a 20 años, 0.571989; para  $I_2$ , con 46 sujetos, 0.606820; para  $I_3$ , integrado por 29 sujetos de 40 a 60 años, 0.755027; y para  $I_4$ , con 58 sujetos de 60 a 80 años, 0.757147. Se observa una tendencia de crecimiento casi lineal conforme el aumento de edad, sin embargo, el grupo  $I_3$  tiene un valor atípico con  $\Phi = 3.527796$ , mientras que el grupo  $I_4$  tiene tres valores atípicos, siendo el mayor con  $\Phi = 6.19$ , por lo que al excluir dichos valores, se mantienen la medias de los grupos prácticamente constantes.

Este comportamiento sugiere que, dentro de la población estudiada, el cerebro sano en estado de reposo mantiene niveles similares de tasa de producción de entropía a lo largo de la vida. En particular, no se observan diferencias significativas entre sujetos jóvenes (a partir de los 10 años) y envejecidos (hasta los 80 años) en términos de  $\Phi$ . Esta ausencia de diferencias podría asociarse a mecanismos de compensación o reorganización a nivel global del cerebro. Debido a que el estado de reposo refleja la dinámica intrínseca del sistema, los resultados indican que  $\Phi$  no constituye una medida suficientemente sensible para detectar cambios funcionales y estructurales asociados al envejecimiento en sujetos sanos. Sin embargo esta consistencia de las medias de  $\Phi$  en sujetos sanos sugiere que esta medida puede servir como referencia basal, de modo que desviaciones respecto a este patrón podrían indicar alteraciones patológicas. Comparando el valor de un sujeto con este rango, desviaciones significativas pueden asociarse a enfermedades neurodegenerativas como Alzheimer o Parkinson.

En la segunda etapa se aplicó el modelo anatómico funcional a los clusters. Las medias de  $\Phi$  son: para  $C_0$  con 40 sujetos, 0.768370; para  $C_1$  con 48 sujetos, 0.551346, para  $C_2$  con 39 sujetos 0.672643 y para  $C_3$  con 34 sujetos 0.774718.

Los resultados muestran que los clusters  $C_n$  presentan una correspondencia parcial con los grupos definidos según la edad biológica  $I_n$ . En particular, el cluster  $C_1$  muestra una correspondencia con el grupo  $I_1$ , conformado por sujetos de 10 a 20 años, ya que el 43,8% de los sujetos pertenecen a esta edad. De manera similar, el cluster  $C_3$  presenta una coincidencia notable con el grupo  $I_4$ , integrado por sujetos de 60 a 80 años, donde

el 85,3% de los sujetos corresponde a esta edad.

La distribución de los sujetos entre los clusters es relativamente homogénea. Estos se obtienen a partir de las propiedades intrínsecas de los datos, que integran la conectividad funcional (FC) y estructural (SC). En este sentido, los clusters capturan patrones funcionales y estructurales que permiten caracterizar la edad de los sujetos, la correspondencia del  $C_3$  con el grupo  $I_4$  sugiere que, las características propias de este rango de edad están claramente representadas en dicho cluster.

Por otro lado, al analizar las características predominantes en el modelo de Random Forest Classifier, se observa que los ROIs más relevantes, correspondientes a conectividad funcional, se concentran principalmente en  $C_1$ . Esto sugiere que las propiedades funcionales correspondientes a edades tempranas son más discriminativas, lo que podría explicar su contribución a la separación entre los sujetos, mientras que el modelo de Random Forest Regressor resalta características estructurales, lo que sugiere que la conectividad estructural estaría asociada a cambios más graduales y continuos a lo largo del envejecimiento.

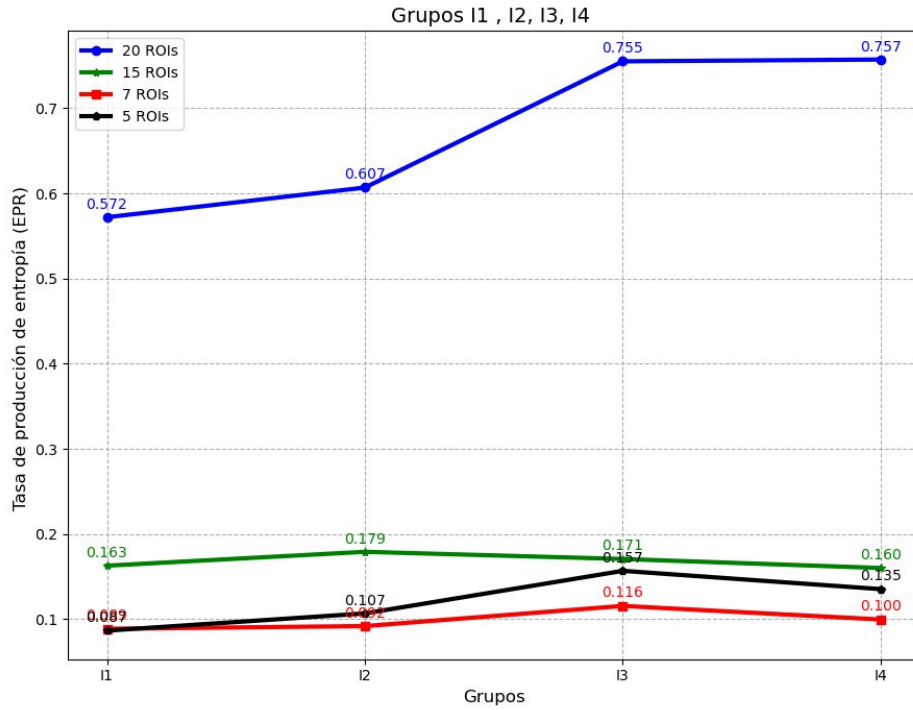
En la tercera etapa se aplicó el modelo anatómico funcional a los biomarcadores, tanto a los grupos como a los clusters. Las medias de  $\Phi$  por grupos son:

- 15 ROIs:  $I_1 = 0.162869$ ,  $I_2 = 0.179146$ ,  $I_3 = 0.170701$ ,  $I_4 = 0.160134$
- 7 ROIs:  $I_1 = 0.088746$ ,  $I_2 = 0.091937$ ,  $I_3 = 0.115599$ ,  $I_4 = 0.099675$
- 5 ROIs:  $I_1 = 0.086604$ ,  $I_2 = 0.106800$ ,  $I_3 = 0.156799$ ,  $I_4 = 0.135345$

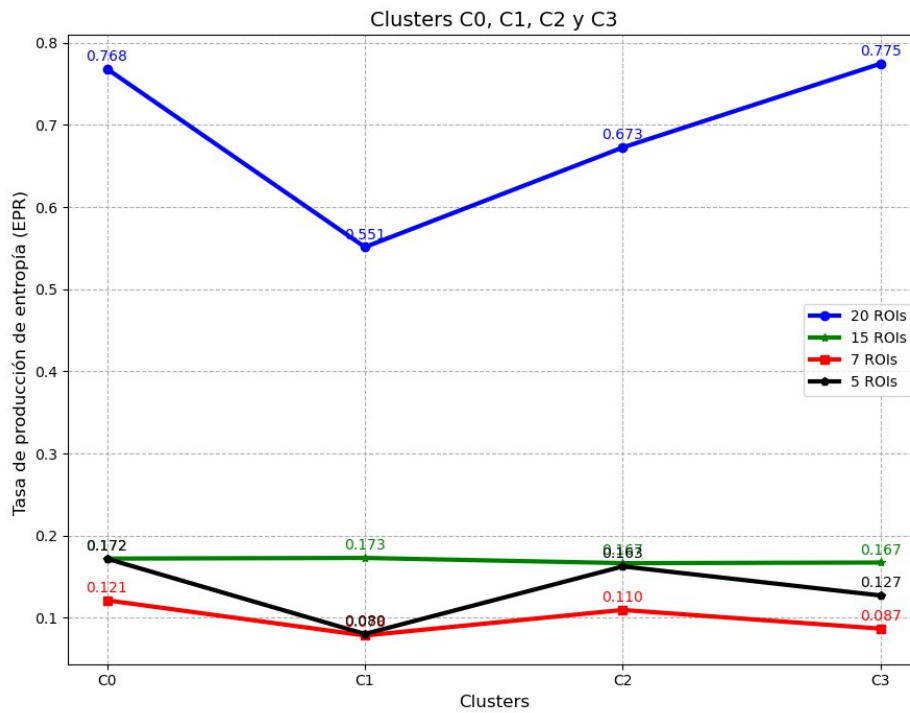
Las medias de  $\Phi$  por cluster son:

- 15 ROIs:  $C_0 = 0.172144$ ,  $C_1 = 0.172912$ ,  $C_2 = 0.166594$ ,  $C_3 = 0.167325$
- 7 ROIs:  $C_0 = 0.121330$ ,  $C_1 = 0.078488$ ,  $C_2 = 0.109593$ ,  $C_3 = 0.086846$
- 5 ROIs:  $C_0 = 0.171899$ ,  $C_1 = 0.080379$ ,  $C_2 = 0.162606$ ,  $C_3 = 0.127215$

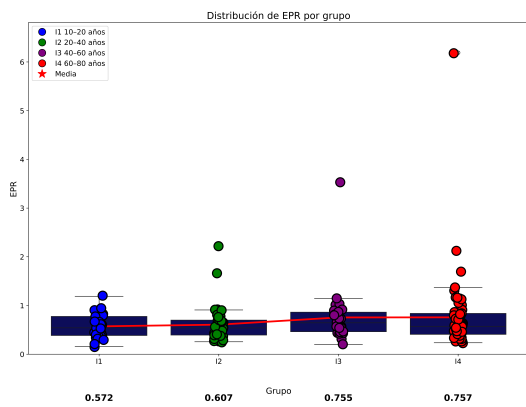
El análisis de  $\Phi$  en los distintos subconjuntos de ROIs no muestra un patrón consistente entre las medias como diferenciación entre agrupaciones. Mientras que para 20 y 5 ROIs se observan diferencias entre agrupaciones, mientras que en 7 y 5 no son tan evidentes, en 15 es casi constante. Esto indica que  $\Phi$  depende tanto del número como de la selección de ROIs, lo que limita su capacidad para caracterizar de manera robusta los cambios asociados al envejecimiento cerebral. En este sentido, sería pertinente evaluar esta medida utilizando otros datos y parcelaciones con un mayor número de regiones, como el Automated Anatomical Labeling (AAL), que considera 90 ROIs.



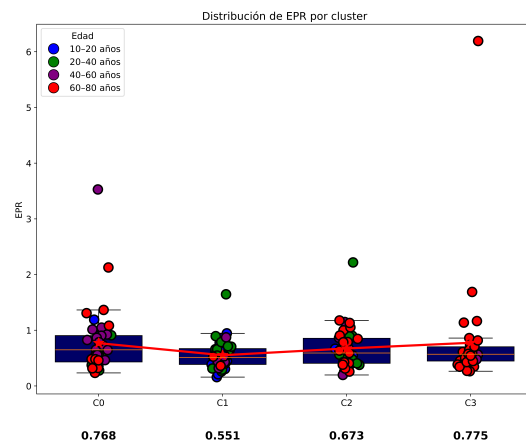
(a) Comportamiento de las medias de la tasa de producción de entropía de los grupos.



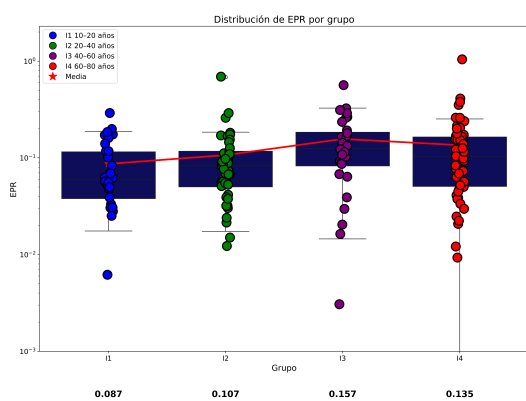
(b) Comportamiento de las medias de la tasa de producción de entropía de los clusters.



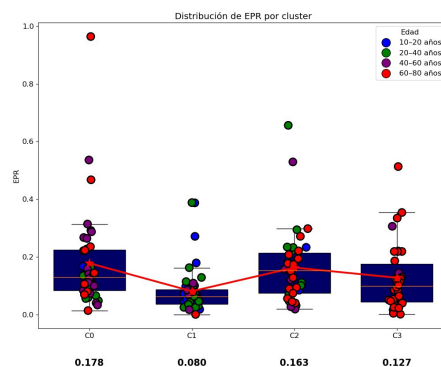
(a) Medias de  $\Phi$  para los grupos con 20 ROIs



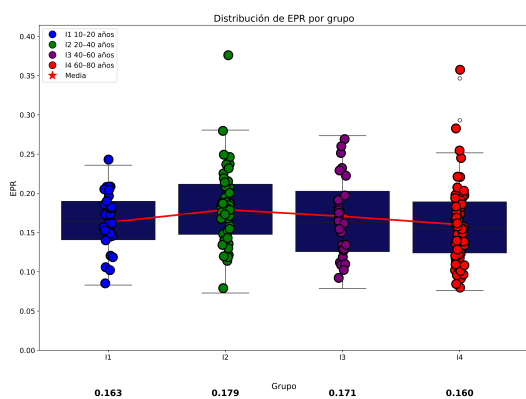
(b) Medias de  $\Phi$  para los clusters con 20 ROIs



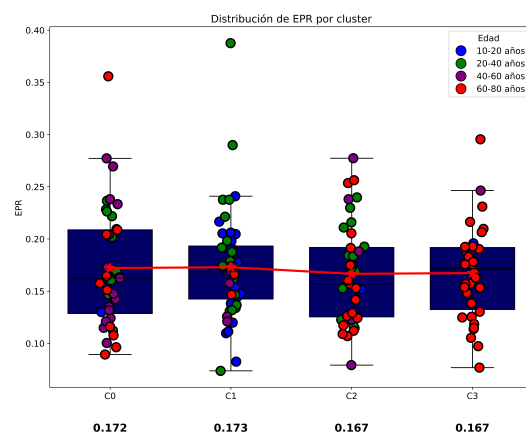
(c) Medias de  $\Phi$  para los grupos con biomarcadores



(d) Medias de  $\Phi$  para los clusters con biomarcadores



(e) Medias de  $\Phi$  para los grupos sin los biomarcadores



(f) Medias de  $\Phi$  para los clusters sin los biomarcadores

Figura 8.1: Tasa de producción de entropía de los grupos  $I_n$ , clusters  $C_n$  y con los biomarcadores.

## Conclusiones

En neurociencia, la tasa de producción de entropía  $\Phi$ , se ha empleado para investigar la dinámica cerebral y comprender cómo los sistemas neuronales transitan entre diferentes estados en diversas condiciones y a diferentes escalas. Una alta entropía refiere un estado desorganizado de la actividad cerebral, observado en algunos trastornos como convulsiones epilépticas mientras que una baja entropía refiere a un estado altamente predecible de la actividad cerebral, observado en afecciones como anestesia general, sueño o trastornos de conciencia [74]. El sistema biológico, como es el cerebro, dicha magnitud se ha relacionado con la dinámica neuronal que mantiene al sistema en un estado estacionario lejos del equilibrio.

El análisis de 20, 15, 7 y 5 ROIs permite evaluar la irreversibilidad y la tasa de producción de entropía. Con 20 ROIs se preserva un mayor número de grados de libertad [3, 86], lo que facilita capturar interacciones y flujos irreversibles que caracterizan la dinámica cerebral envejecida. En contraste, el análisis con los biomarcadores integra dinámicas más locales, lo que se refleja en una disminución de la tasa de producción de entropía.

Los biomarcadores identificados corresponden a los ROIs 2, 6, 10, 14, 16, 19 y 20, observados en los subconjuntos de 7 y 5 ROIs. En estudios previos éstos presentan la particularidad de que los ROIs 10 y 16 se solapan con la red por defecto (Default Mode Network, DMN), éste último contiene los ROIs 2, 6 y 14, (ver 9). El ROI 16 abarca el giro medio temporal, mostrando correspondencia en el atlas estructural Automated Anatomical Labeling (AAL) [29]. En [7] muestra que a lo largo de la vida la dinámica cerebral en reposo se reorganiza hacia patrones crecientemente redundantes - múltiples regiones que aportan información solapada- con una contribución predominante de la red DMN. El ROI 16 se vuelve sinérgico - es decir que emerge sólo en la interacción- en edad avanzada como en  $I_4$  [55] y  $C_3$ . Asimismo, la actividad intrínseca de la DMN está relacionada con el envejecimiento [80]. Las descripciones anatómicas de los ROIs se encuentran en 9.

Estos resultados sugieren que los biomarcadores identificados capturan patrones relevantes de la dinámica neuronal asociados a la edad, y permiten establecer correspondencias entre la clasificación por edad biológica de los sujetos y en la basada en sus propiedades funcionales y estructurales de los sujetos. Por lo tanto, se cumplen los dos objetivos planteados en este tesis.

# Anexos

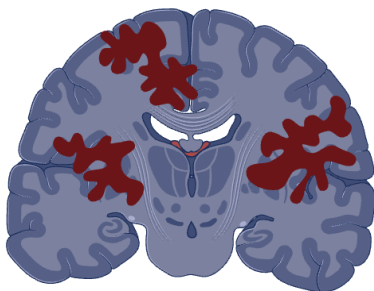
**Descripción anatómica de parcelación módulos de estructura funcional (SFM)** Descripción anatómica de los 20 ROIs que conforma la parcelación estructura funcional, en corte frontal o coronal y corte plano horizontal. Parcelación determinada por [29].



## ROI 1

$7,26 \text{ cm}^3$

Circunvolución cingulada posterior: área posterior de la circunvolución cingulada o circunvolución callosa. Ubicada por encima del cuerpo calloso, se extiende desde el lóbulo frontal hasta el uncus temporal y asciende hasta el esplenio. Pertenecer a la red de modo predeterminado Default mode network DMN.



## ROI 2

$104,36 \text{ cm}^3$

Putamen: estructura redonda ubicada en la base del telencéfalo. Parte de los ganglios basales. Cíngulo anterior: región frontal del giro cingulado. Porción rostral del giro frontal medio: extremo anteroinferior del giro frontal medio. Giro parietal superior: posterior al giro poscentral. Giro supramarginal: rodea el extremo de la cisura de Sylvian. Ínsula: pared medial de la cisura de Sylvian. Giro parietal inferior: detrás del giro poscentral. Giro precentral: giro frontal motor. Giro frontal superior: giro parasagital anterosuperior.



### ROI 3

221,18  $cm^3$

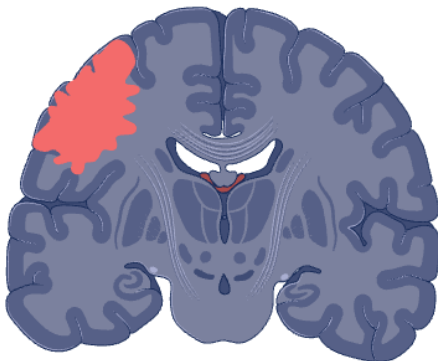
Lóbulo paracentral: circunvolución medial que conecta la circunvolución precentral y poscentral. Circunvolución precentral (ROI 2) Circunvolución poscentral: circunvolución parietal situada entre la cisura de Rolando y el surco poscentral, cuya función principal es sensorial. Precúneo: lóbulo cerebral cuadrado situado antes del surco parietooccipital y detrás del lóbulo paracentral, en la superficie medial del hemisferio cerebral. Circunvolución frontal superior (ROI 2). Circunvolución parietal superior ((ROI 2) Circunvolución temporal superior: circunvolución temporal en la superficie lateral del lóbulo temporal. Se encuentra debajo de la cisura de Silvio y por encima del surco temporal superior. Pertenece al neocórtex temporal. Circunvolución supramarginal (ROI 2) Ínsula (ROI 2).



### ROI 4

91,48  $cm^3$

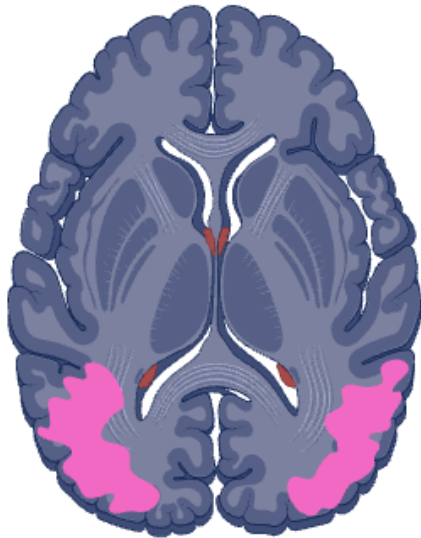
Cúneo: circunvolución occipital entre el surco parietooccipital y el surco calcarino en la superficie medial del lóbulo occipital. Cúneo occipital lateral: superficie lateral externa del lóbulo occipital, cerca del lóbulo occipital, que divide la circunvolución occipital externa. Cúneo lingual: extensión occipital de la circunvolución parahipocámpica en la superficie medial del lóbulo occipital. Corteza pericalcarina: área occipital que rodea el surco calcarino y cuya función está relacionada con la percepción visual. Precúneo (ROI 3).



### ROI 5

37,02  $cm^3$

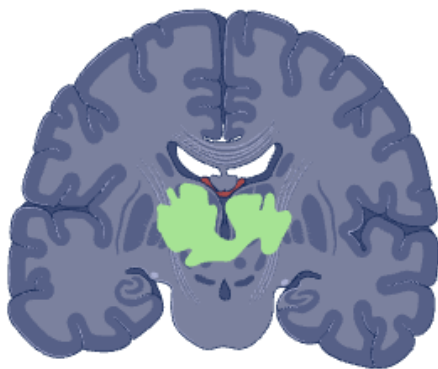
Giro frontal medial: giro frontal en la superficie lateral, debajo del giro frontal superior. Giro precentral (ROI 2). Porción rostral del giro frontal medio (ROI 2).



### ROI 6

159,33  $cm^3$

Cerebelo: parte posterior del rombencéfalo, compuesta por los dos hemisferios y el vermis central. Se ubica debajo del lóbulo occipital. Giro fusiforme: giro temporal en la superficie inferior, entre el giro temporal inferior y el giro parahipocampal. Tiene dos áreas: el giro occipitotemporal medial y el giro occipitotemporal lateral. Giro temporal inferior: giro inferior ubicado en la superficie lateral del lóbulo temporal, debajo del surco temporal inferior. Surco occipital lateral (ROI 4). Giro parietal superior (ROI 2).



### ROI 7

22,30  $cm^3$

Tálamo: Estructura simétrica media del diencefalo con múltiples conexiones aferentes y eferentes, situada alrededor del tercer ventrículo. Núcleo caudado (estructura simétrica): Una de las estructuras básicas de los ganglios basales pertenecientes al cuerpo estriado. Se ubica en la superficie lateral de los ventrículos laterales que rodean el tálamo. Putamen (ROI 2). Pálido: Estructura simétrica dentro de los ganglios basales. Región diencefálica medial del núcleo lenticular. Núcleo accumbens: Estructura simétrica ubicada en la región ventral del cuerpo estriado, por lo que pertenece a los ganglios basales.



### ROI 8

3,29  $cm^3$

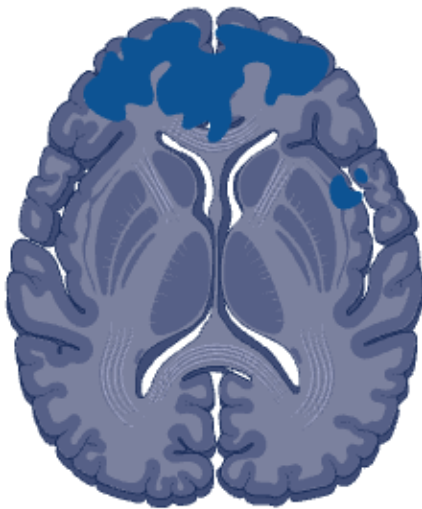
Núcleo caudado (ROI 7)  
Putamen (ROI 2)



### ROI 9

163,67  $cm^3$

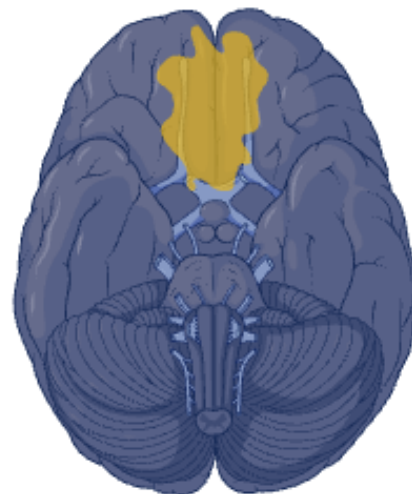
Cerebelo (ROI 6). Frontal medio caudal: circunvolución frontal en la superficie lateral, ubicada por debajo y lateral a la circunvolución frontal superior. Esta región se refiere a su parte más caudal. Istmo cingulado: estrechamiento de la intersección entre la circunvolución cingulada y la circunvolución hipocampal. Se ubica detrás y por debajo del esplenio del cuerpo caloso. Cíngulo posterior (ROI 1). Precúneo (ROI 3). Cíngulo parietal inferior (ROI 2). Pars rostral de la circunvolución frontal media (ROI 2). Cíngulo frontal superior (ROI 2).



### ROI 10

103,55  $cm^3$

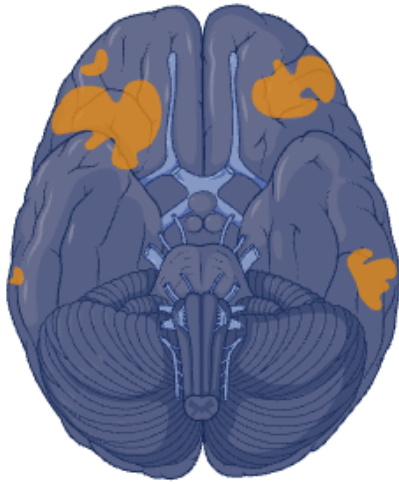
Circunvolución cingulada anterior (ROI 2). Circunvolución parietal inferior (ROI 2). Circunvolución orbitaria: circunvolución frontobasal lateral a la circunvolución recta. Pars opercularis: porción opercular de la circunvolución frontal inferior. Pars orbitalis: porción orbital de la circunvolución frontal inferior. Pars triangularis: porción inferior de la circunvolución frontal inferior. Circunvolución anterior (ROI 2). Pars rostral de la circunvolución frontal media (ROI 2). Circunvolución frontal superior (ROI 2).



### ROI 11

31,08  $cm^3$

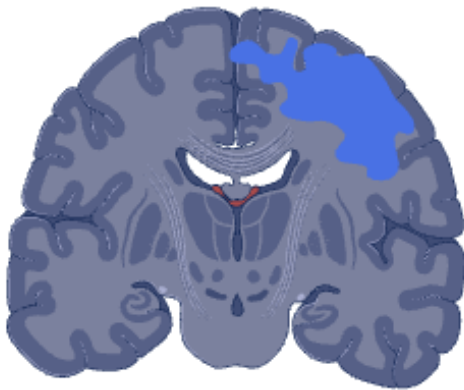
Núcleo caudado (ROI 7). Núcleo accumbens (ROI 7). Giro orbitario frontal lateral: giro orbitario externo, ubicado frontobasal y lateral al giro orbitofrontal medial. Giro orbitario (ROI 10). Cíngulo anterior (ROI 10).



### ROI 12

33,24  $cm^3$

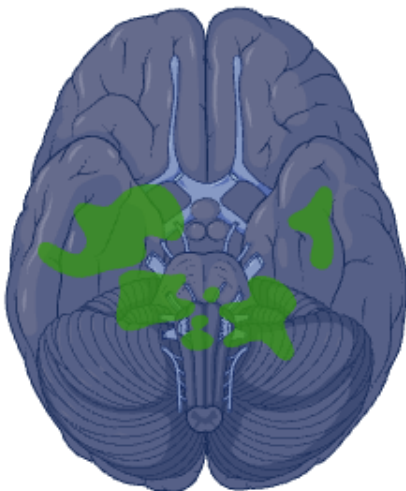
Giro parietal inferior (ROI 2). Giro temporal inferior (ROI 6). Giro orbital frontal lateral (ROI 11). Pars orbitalis (ROI 10). Pars triangularis (ROI 10). Pars rostral del giro frontal medio (ROI 2). Giro frontal superior (ROI 2). Núcleo caudado y cingulado anterior (ROI 2 y 7).



### ROI 13

24,46  $cm^3$

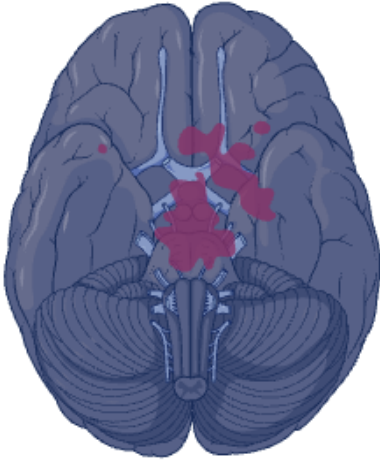
Giro frontal medio: parte caudal del giro frontal medio. Pars opercularis (ROI 10). Giro precentral (ROI 2). Giro frontal superior ROI 2).



### ROI 14

92,75  $cm^3$

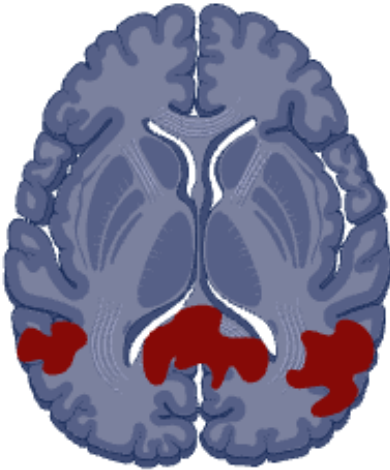
Tálamo (ROI 7). Hipocampo: Estructura simétrica de sustancia gris, ubicada en la región mesiotemporal, en la base del asta temporal. Amígdala: Núcleos grises ubicados en el uncus temporal, por encima del asta ventricular temporal. Pertenece al rinencéfalo. Putamen (ROI 2). Diencefalo ventral: Múltiples estructuras que contienen el hipotálamo, el tubérculo mamilar, el núcleo subtalámico, la sustancia negra, el núcleo rojo, el cuerpo geniculado, el tracto óptico y los pedúnculos cerebrales. Orillas del surco temporal superior: Estructura del lóbulo temporal entre la circunvolución temporal superior y la circunvolución temporal media. Circunvolución parahipocámpica: Circunvolución ubicada debajo del surco hipocámpal en la región mesial temporal. Giro temporal superior (ROI 3). Ínsula (ROI 2). Giro temporal medio: giro ubicado en la superficie lateral del lóbulo temporal, entre los surcos temporales inferior y superior. Polo temporal: extremo anterior del lóbulo temporal.



### ROI 15

42,96  $cm^3$

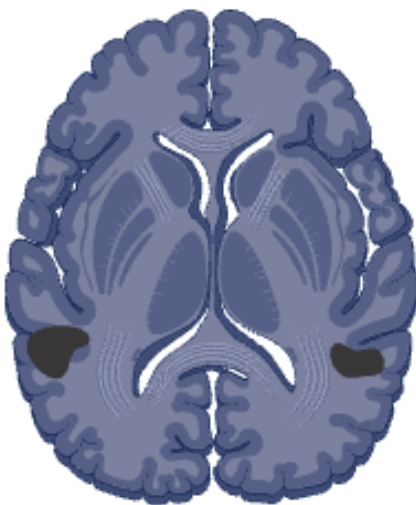
Tálamo (ROI 7). Putamen (ROI 2). Pálido (ROI 7). Tronco encefálico: consta de tres partes: mielencéfalo, protuberancia (metencéfalo) y mesencéfalo. Es la principal vía de comunicación entre el encéfalo, la médula espinal y los nervios periféricos. Hipocampo (ROI 14). Amígdala (ROI 14). Núcleo accumbens (ROI 7). Diencéfalo ventral (ROI 14). Giro orbitario (ROI 10). Ínsula (ROI 2).



### ROI 16

65,58  $cm^3$

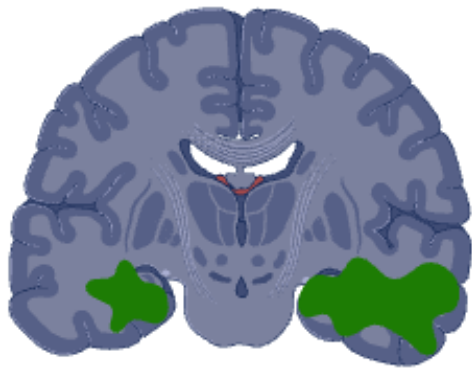
Cerebelo (ROI 6). Riberas del surco temporal superior (ROI 14). Circunvolución parietal inferior (ROI 2). Istmo cingulado (ROI 9). Circunvolución temporal media (ROI 14). Precúneo (ROI 3). Circunvolución temporal superior (ROI 3).



### ROI 17

5,29  $cm^3$

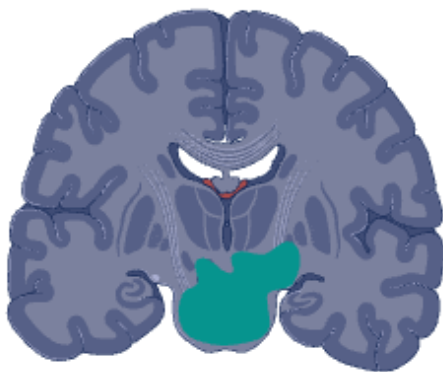
Orillas del surco temporal superior (ROI 14). Giro temporal medio (ROI 14).



### ROI 18

74,39  $cm^3$

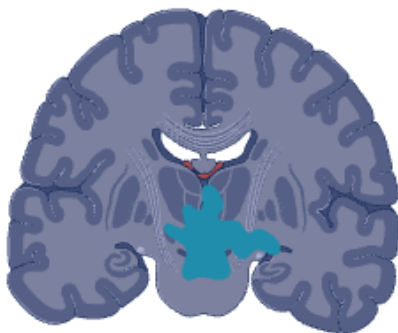
Hipocampo (ROI 14). Amígdala (ROI 14). Corteza entorrinal: área del lóbulo temporal medio ubicada entre el hipocampo y el neocórtex temporal. Giro fusiforme (ROI 6). Giro temporal inferior (ROI 6). Giro temporal medio (ROI 14). Giro parahipocámpico (ROI 14). Polo temporal (ROI 14).



### ROI 19

28,54  $cm^3$

Cerebelo (ROI 6). Tronco encefálico (ROI 15).

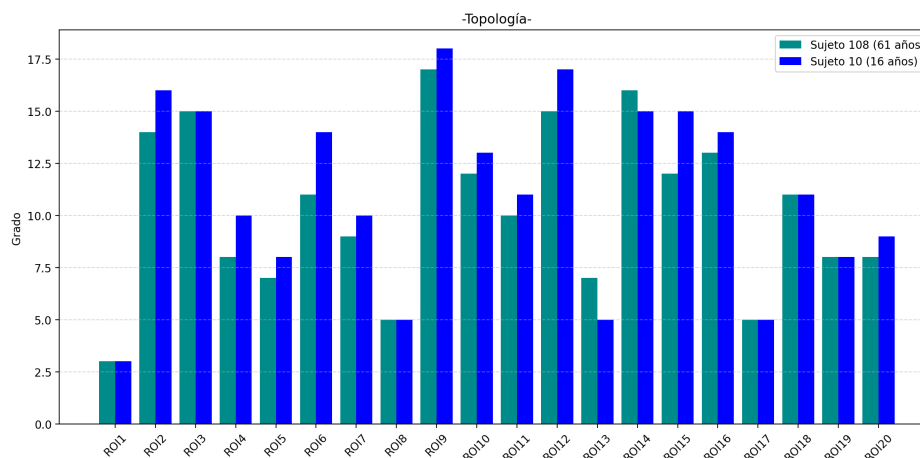


### ROI 20

34,91  $cm^3$

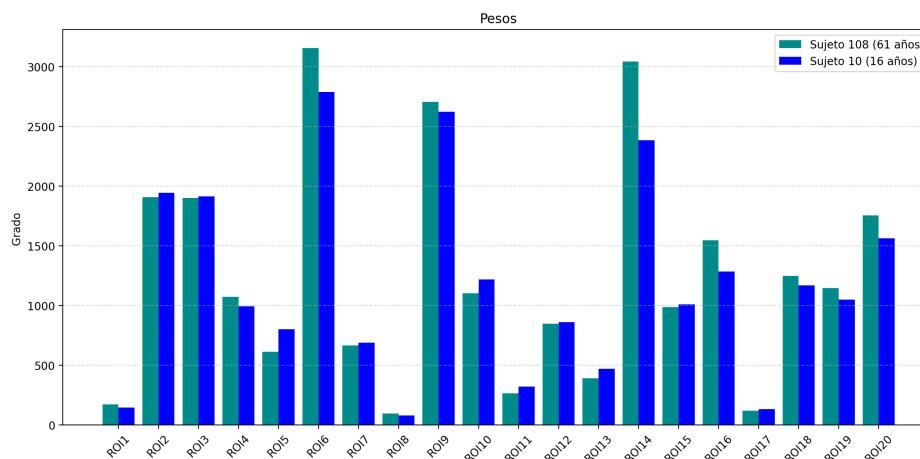
Cerebelo (ROI 6). Giro parahipocampal (ROI 14).

## Características topológicas de la SC

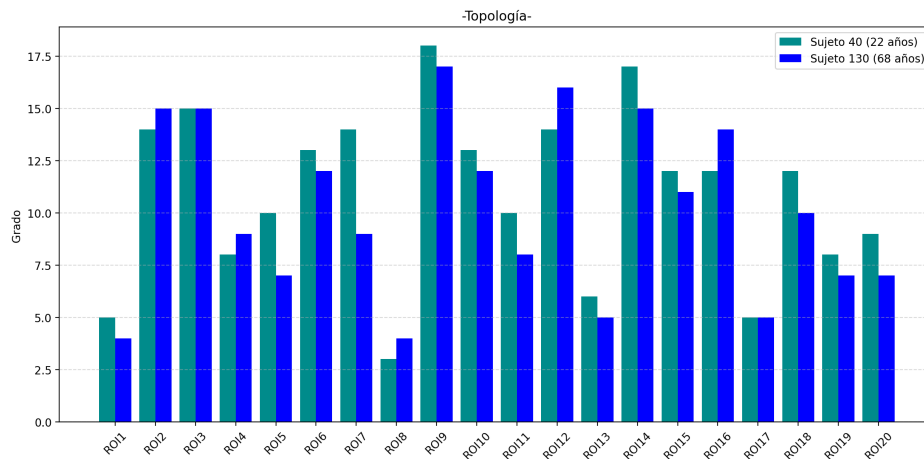


### Conexiones de pesos de los 20 ROIs.

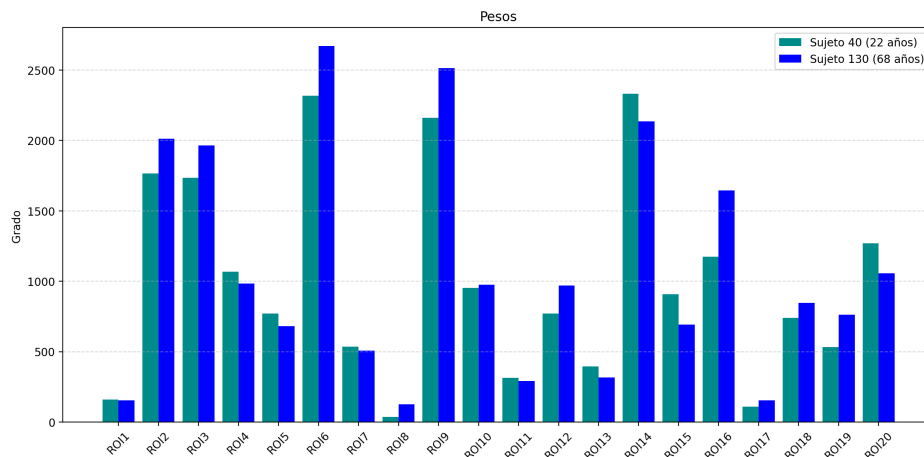
Se muestra una características de la SC. El grado binario cuantifica el número de regiones de interés (ROIs) anatómicamente conectadas con la ROI  $i$ . Esta métrica considera la presencia o ausencia de conexiones, independientemente de su peso, y proporciona un indicador cualitativo de la conectividad estructural. Se realiza una comparación entre un sujeto de 16 años y otro de 61 años. La gráfica evidencia una mayor prevalencia en las ROIs 2,3,9,12,14 y 16.



**Conexiones de pesos de las conexiones de los 20 ROIs** El grado ponderado cuantifica la fuerza nodal de cada una de las 20 ROIs y se define como la suma de los pesos de las conexiones incidentes en cada nodo  $i$ . Esta métrica nos indica la intensidad de las conexiones anatómicas y proporciona información cuantitativa sobre la conectividad estructural. Se realiza una comparación entre un sujeto de 16 años y otro de 61 años. La gráfica evidencia una mayor prevalencia en las ROIs 2, 3, 6, 9, 14 y 20.

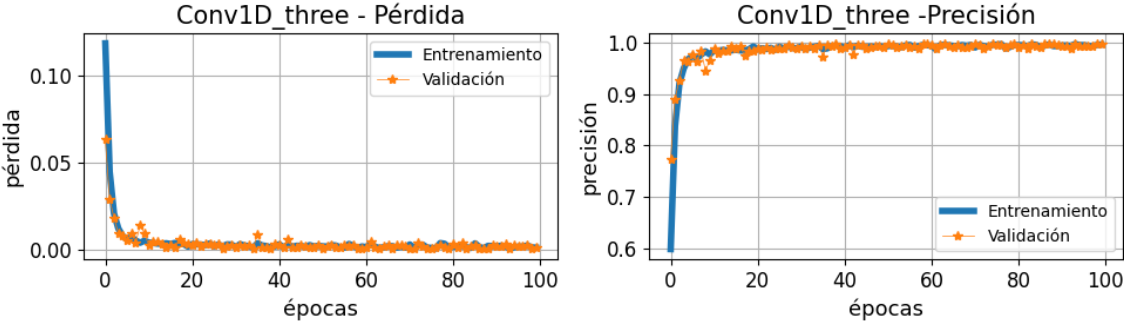


**Conexiones binarias de los 20 ROIs** Se realiza una comparación entre un sujeto de 22 años y otro de 38 años. La gráfica evidencia una mayor prevalencia en las ROIs 3,9, y 14 predominando el sujeto de 22 años.

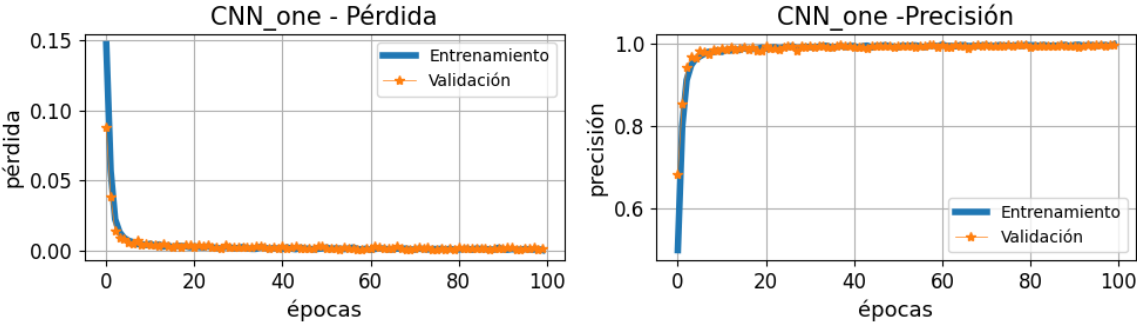


**Conexiones de pesos de las conexiones de los 7 ROIs** Se realiza una comparación entre un sujeto de 22 años y otro de 68 años. La gráfica evidencia una mayor prevalencia en las ROIs 6,9,14,16 y 20.

# Modelos de aprendizaje profundo.



Modelo convolucional 1D



Modelo recurrente

# Bibliografía

- [1] «1D convolutional neural networks and applications: A survey». En: *Mechanical Systems and Signal Processing* 151 (2021), pág. 107398. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymsp.2020.107398>.
- [2] Ince Robin A.A. et al. «A statistical framework for neuroimaging data analysis based on mutual information estimated via a Gaussian copula». En: *Human Brain Mapping* 38:1541-1573 (2017). DOI: 10.1002/hbm.23471.
- [3] Busiello D M et al. «Entropy production for coarse-grained dynamic». En: *New Journal of physics* (2019). DOI: 10.1088/1367-2630/ab29c0.
- [4] Erramuzpe Asier et al. «Identification of redundant and synergetic circuits in triplets of electrophysiological data». En: *Journal of Neural Engineering* 12 066007 (11p) (9 October 2015). DOI: 10.1088/1741-2560/12/6/066007.
- [5] Géron Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly, 2019.
- [6] Buxton Richard B. «The physics of functional magnetic resonance imaging (fMRI)». En: *Rep Prog Phys* (2015). DOI: 10.1088/0034-4885/76/9/096601.
- [7] Camino Pontes Borja et al. «Interaction Information Along Lifespan of the Resting Brain Dynamics Reveals a Major Redundant Role of the Default Mode Network». En: *Entropy (Basel)*. 2018;20(10):742 (28 september 2018). DOI: <https://doi.org/10.3390/e20100742>.
- [8] Hansen Enrique C.A. et al. «Functional connectivity dynamics: Modeling the switching behavior of the resting state». En: *NeuroImage* 105 (2015), págs. 525-535. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2014.11.001>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811914009033>.
- [9] Joana Cabral, Morten L. Kringelbach y Gustavo Deco. «Exploring the network dynamics underlying brain activity during rest». En: *Progress in Neurobiology* 114 (2014), págs. 102-131. ISSN: 0301-0082. DOI: <https://doi.org/10.1016/j.pneurobio.2013.12.005>. URL: <https://www.sciencedirect.com/science/article/pii/S0301008213001457>.
- [10] Godrèche Claude y Jean-Marc Luck. «Characterising the nonequilibrium stationary states of Ornstein-Uhlenbeck processes». En: *J. Phys.A:Math.Theor.* 52 035002 (2019). DOI: 10.1088/1751-8121/aaf190.

- [11] W.Lynn Cristopher, Papadopoulos Lia y S.Bassett Danielle. «Broken detailed balance and entropy production in the human brain». En: *Biophysics and computational biology Vol. 118 / No. 47* (November 17,2021). DOI: 10.1073/pnas.2109889118.
- [12] Frank Till D. *Nonlinear Fokker-Planck Equations. Fundamentals and Applications*. Springer, 2005.
- [13] Greicius M. D. et al. «Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI». En: *Proceedings of the National Academy of Sciences of the United States of America*, 101(13), 4637-4642 (2004). DOI: <https://doi.org/10.1073/pnas.0308627101>.
- [14] G. Deco, Jirsa V. K. y A. R McIntosh. «Emerging concepts for the dynamical organization of resting-state activity in the brain. Nature reviews». En: *Neuroscience*, 12(1), 43-56. (2011). DOI: <https://doi.org/10.1038/nrn2961>.
- [15] Stephan Klaas Enno et al. «Comparing hemodynamic models with DCM». En: *NeuroImage* 38.3 (2007), págs. 387-401. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2007.07.040>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811907006489>.
- [16] Acar Elif F y Sun Lei. «A Generalized Kruskal–Wallis Test Incorporating Group Uncertainty with Application to Genetic Association Studies». En: *Statistics* (2012). DOI: <https://doi.org/10.48550/arXiv.1205.0534>.
- [17] Rosas Fernando et al. «Quantifying high-order interdependencias via multivariate extensions of the mutual information». En: *Phys. Rev. E* 100 (3 March1, 2019), pág. 032305. DOI: <https://doi.org/10.1103/PhysRevE.100.032305>.
- [18] De Borja Rodríguez Ortiz Franciso. *Procesos de estabilización, sincronización y aprendizaje en redes neuronales estocásticas. Memoria de tesis doctoral. Universidad Autónoma de Madrid. Escuela Técnica Superior de Informática. Departamento de Ingeniería Informática*. 1999.
- [19] Rayman Sephira G y L Posto Kathleen. «MRI Biomarkers of motor and non-motor symptoms in Parkinson's disease». En: *Parkinsonism Related Disorders Volumen 73* (April 2020, Pages 85-93). DOI: <https://doi.org/10.1016/j.parkreldis.2019.10.002>.
- [20] E.P. Box George et al. *Time series analysis. Forecasting and Control*. Copyright by John Wiley Sons, Inc. All rights reserved, 2015015492, 2016.
- [21] Cofré Rodrigo Gilson Matthieu Tagliazucchi Enzo. «Entropy production of multivariate Ornstein-Uhlenbeck processes correlates with consciousness levels in the human brain». En: *Physical Review E* 107, 024121 (2023). DOI: <https://doi.org/10.1103/PhysRevE.107.024121>.
- [22] Deco Gustavo y Jirsa K Viktor. «Ongoing cortical activity at rest: Criticality, Multistability ad Ghost Attractors». En: *Journal of Neuroscience*. (2012). DOI: 10.1523/JNEUROSCI.2523-11.2012.

- [23] Deco Gustavo et al. «How local excitation-inhibition ratio impacts the whole brain dynamics». En: *The Journal of Neuroscience* 31(23):7886-7898 (2014). DOI: <https://doi.org/10.1523/JNEUROSCI.5068-13.2014>.
- [24] Deco Gustavo et al. «Whole-Brain Multimodal Neuroimaging Model Using Serotonin Receptor Maps Explains Non-linear Functional Effects of LSD». En: *Current Biology* 28.19 (2018), 3065-3074.e6. ISSN: 0960-9822. DOI: <https://doi.org/10.1016/j.cub.2018.07.083>. URL: <https://www.sciencedirect.com/science/article/pii/S0960982218310455>.
- [25] Deco Gustavo et al. «Resting-State functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations.» En: *Journal of Neuroscience* 33 (27) 11239-11252 (3 July 2013). DOI: [doi.org/10.1523/JNEUROSCI.1091-13.2013](https://doi.org/10.1523/JNEUROSCI.1091-13.2013).
- [26] Grandy Thomas H et al. «On the estimation of brain signal entropy from sparse neuroimaging data». En: *Scientific reports* (2016). DOI: <https://doi.org/10.1038/srep23073>.
- [27] Te Sun Han. «Nonnegative entropy measures of multivariate symmetric correlations». En: *Information and Control* 36.2 (1978), págs. 133-156. ISSN: 0019-9958. DOI: [https://doi.org/10.1016/S0019-9958\(78\)90275-9](https://doi.org/10.1016/S0019-9958(78)90275-9). URL: <https://www.sciencedirect.com/science/article/pii/S0019995878902759>.
- [28] Goodfellow Ian, Bengio Yoshua y Courville Aaron. *Deep Learning*. The MIT Press, November 2016.
- [29] Diez Ibai et al. «A novel brain partition highlights the modular skeleton shared by structure and function». En: *Scientific Reports* (Junio 2015). DOI: [10.1038/srep10532](https://doi.org/10.1038/srep10532).
- [30] Robin Ince. *Cúpula Gaussiana: Gaussian-Copula Mutual Information*. <https://github.com/robince/gcmi/>. Repositorio en GitHub. 2020.
- [31] Zamora-López Gorka Insabato Andrea y Gilson Matthieu. *Analysis of fMRI/BOLD signals: connectivity estimation, task/subject identification, network analysis*. [https://github.com/mb-BCA/notebooks\\_review2019?tab=readme-ov-file](https://github.com/mb-BCA/notebooks_review2019?tab=readme-ov-file). 2019.
- [32] Basser Peter J et al. «In vivo fiber tractography using DT-MRI data». En: *Magnetic Resonance in Medicine* 44:625-632 (2000). DOI: [https://doi.org/10.1002/1522-2594\(200010\)44:4<625::AID-MRM17>3.0.CO;2-0](https://doi.org/10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-0).
- [33] Neuman M. E. J. «Fast algorithm for detecting community structure in networks». En: *Phys. Rev. E* 69, 066133 (2004). DOI: <https://doi.org/10.1103/PhysRevE.69.066133>.
- [34] VanderPlas Jake. *Python data science handbook, essential tools for working with data*. O'Reilly, 2017. ISBN: 978-1-491-91205-8.

- [35] Hlinka Jaroslav et al. «Functional connectivity in resting-state fMRI: Is linear correlation sufficient?» En: *NeuroImage* 54.3 (2011), págs. 2218-2225. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2010.08.042>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811910011249>.
- [36] Ren Jie et al. «Noise Bridges Dynamical Correlation and Topology in Coupled Oscillator Networks». En: *Phys. Rev. Lett.* 104 (5 feb. de 2010), pág. 058701. DOI: 10.1103/PhysRevLett.104.058701. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.104.058701>.
- [37] Friston K.J. et al. «Nonlinear Responses in fMRI: The Balloon Model, Volterra Kernels, and Other Hemodynamics». En: *NeuroImage* 12.4 (2000), págs. 466-477. ISSN: 1053-8119. DOI: <https://doi.org/10.1006/nimg.2000.0630>. URL: <https://www.sciencedirect.com/science/article/pii/S105381190090630X>.
- [38] Friston J Karl. «Functional and effective connectivity: A review». En: *Brain connectivity Vol 1. No.1* (1 June 2011). DOI: 10.1089/brain.2011.0008.
- [39] Alexander Andrew L. et al. «Diffusion Tensor Imaging of the Brain». En: *Neurotherapeutics* 4.3 (2007). Advances in Neuroimaging/Neuroethics, págs. 316-329. ISSN: 1878-7479. DOI: <https://doi.org/10.1016/j.nurt.2007.05.011>. URL: <https://www.sciencedirect.com/science/article/pii/S1878747923006530>.
- [40] Rincón Luis. *Introducción a los procesos estocásticos*. Departamento de Matemáticas, Facultad de Ciencias UNAM Circuito Exterior de CU 04510 CDMX, 2012.
- [41] Rincón Luis. *Introducción a la probabilidad*. Departamento de Matemáticas, Facultad de Ciencias, UNAM, 2013.
- [42] Rincon Luis. *Curso intermedio de probabilidad*. Facultad de Ciencias UNAM, 2007.
- [43] Rincón Luis y De Jesús Romo Veronica. *Una introducción a las series de tiempo*. Facultad de Ciencias. México 04510, Ciudad de México editoriales@ciencias.unam.mx tienda.ciencias.unam.mx, 2024.
- [44] Scott Lundberg. *SHAP Documentation*. <https://shap.readthedocs.io/en/latest/>.
- [45] Cover Thomas M y Thomas Joy A. *Elements of Information Theory*. Second Edition. A Wile-Interscience Publication John Wiley Sons Inc, 1991.
- [46] Gilson M et al. «Framework based on communicability and flow to analyze complex network dynamics». En: *Phys. Rev. E* 97, 052301 (2018). DOI: DOI:<https://doi.org/10.1103/PhysRevE.97.052301>.
- [47] Lundberg Scott M y Lee Su-In. «A unified approach to interpreting model predictions.» En: *Computer Science.Artificial Intelligence* (2017). DOI: <https://doi.org/10.48550/arXiv.1705.07874>.

- [48] Mediano Pedro A. M. et al. *FastDMF: Fast simulator of the Dynamic Mean Field model of brain dynamics*. <https://gitlab.com/concog/fastdmf>. Repositorio GitLab; versión inicial publicada en 2021. 2022.
- [49] Van der Maaten Laurens. «Accelerating t-SNE using Tree-Based Algorithms». En: *Journal of Machine Learning Research* (2014).
- [50] Van der Maaten Laurens e Hinton Geoffrey. «Visualizing data using t-SNE». En: *Journal of machine learning 2579-2605* (2008).
- [51] Van der Maaten Laurens e Hinton Geoffrey. «Visualizing non-metric similarities in multiple maps». En: *Journal of machine learning* (2011). DOI: DOI10.1007/s10994-011-5273-4.
- [52] Timme Marc y Casadiego Jose. «Revealing networks from dynamics: an introduction». En: *Journal of Physics A: Mathematical and Theoretical* (11 August 2014). DOI: 10.1088/1751-8113/47/34/343001.
- [53] Gatica Marilyn. *High order interactions*. url: <https://github.com/KGatica/High-Order-interactions>. Repositorio en GitHub. 2021.
- [54] Gatica Marilyn et al. «High-order functional interactions in ageing explained via alterations in the connectome in a whole-brain model». En: *Plos computational biology* (2021, 15 september). DOI: 10.1101/2021.09.15.460435.
- [55] Gatica Marilyn et al. «High-order interdependencies in the aging brain». En: *Brain connectivity, vol. 11 No.9* (10 November 2021). DOI: 10.1089/brain.2020.0982.
- [56] Watson Matthew et al. *KerasHub*. <https://github.com/keras-team/keras-hub>. 2024.
- [57] Gilson Matthieu. *Light package for simulation and connectivity estimation using the multivariate Ornstein-Uhlenbeck process (MOU)*. <https://github.com/mb-BCA/pyMOU/tree/master>. 2020.
- [58] Gilson Matthieu et al. «Model-based whole-brain effective connectivity to study distributed cognition in health and disease». En: *Network Neuroscience* (2019). DOI: [https://doi.org/10.1162/netn\\_a\\_00117](https://doi.org/10.1162/netn_a_00117).
- [59] Gilson Matthieu et al. «Model-based whole-brain effective connectivity to study distributed cognition in health cognition in health and disease. Supplementary material». En: *Network Neuroscience, 4(2), 338–373*. (2020). DOI: [https://doi.org/10.1162/netn\\_a\\_00117](https://doi.org/10.1162/netn_a_00117).
- [60] Gilson Matthieu et al. «Estimation of directed effective connectivity from fMRI functional connectivity hints at asymmetries of cortical connectome». En: *PLoS Comput Biol 12(3): e1004762* (March 16, 2016). DOI: <https://doi.org/10.1371/journal.pcbi.1004762>.
- [61] William J. McGill. «Multivariate Information Transmission». En: *Psychometrika* 19.2 (1954), págs. 97-116. DOI: 10.1007/BF02289159.

- [62] Y Chan Micaela et al. «Decreased segregation of brain systems across the healthy adult lifespan». En: *Proceedings of the National Academy of Science* 111(46):E4997-E5006 (2014). DOI: <https://doi.org/10.1073/pnas.1415122111>.
- [63] Steele J Michael. *Stochastic Calculus and Financial Applications*. Springer, 2001.
- [64] Raschka Sebastian Mirjalili Vahid. *Phyton Machine Learning. Machine Learning and Deep Learning with Phyton, scikit-learn, and TensorFlow*. Packt, Second edition: september 2017. ISBN: 978-1-78712-593-3.
- [65] Blockley P Nicolás et al. «A review of calibrated blood oxygenation level-dependent (BOLD) methods for the measurement of task-induced changes in brain oxygen metabolism». En: *NMR in biomedicine*, 26(8), 987–1003 (2013). DOI: <https://doi.org/10.1002/nbm.2847>.
- [66] Sporns Olaf, Tononi Giulio y Kötter Rolf. «The human connectome: a structural description of the human brain.» En: *Plot computational biology* (September 30, 2005). DOI: <https://doi.org/10.1371/journal.pcbi.0010042>.
- [67] Dobrow Robert P. *Introduction to Stochastic Processes with R*. Wiley, 2016.
- [68] Dayan Peter y Abbott L.F. *Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems*. The MIT Press Cambridge, Massachusetts-London, England, 2005.
- [69] Fox T Peter et al. «Nanoxidative glucose consumption during focal physiologic neural activity». En: *Science* 241:462-464 (1988). DOI: 10.1126/science.3260686..
- [70] Jezzard Peter y Turner Robert. «Magnetic resonance imaging methods for study of human brain function and their application at high magnetic field». En: *Computerized Medical Imaging and Graphics* 20.6 (1996). Computer-Assisted Imaging of Embryonic and Fetal Development, págs. 467-481. ISSN: 0895-6111. DOI: [https://doi.org/10.1016/S0895-6111\(96\)00044-4](https://doi.org/10.1016/S0895-6111(96)00044-4). URL: <https://www.sciencedirect.com/science/article/pii/S0895611196000444>.
- [71] Ramachandran Prajit, Zoph Barret y V Le Quoc. «Searching for activation functions». En: *ICLR 2018 (International Conference on Learning Representations)* (2017).
- [72] Hong Qian. «Mathematical formalis for isothermal linear irreversibility». En: *Mathematical Physics (math-ph); Statistical Mechanics (cond-mat.stat-mech)* (27 jun 2001). DOI: <https://doi.org/10.1098/rspa.2001.0811>.
- [73] Kandel Eric R et al. *Principles of neural science*. Graw Hill, 2021.
- [74] Cofré Rodrigo y Destexhe Alain. «Entropy and Complexity Tools Across Scales in Neuroscience: A Review». En: *Entropy* (2025). DOI: <https://doi.org/10.3390/e27020115>.

- [75] Cofré Rodrigo et al. «Whole Brain Models to Explore Altered State of Consciousness from the Bottom Up». En: *Brain Science* (2020). DOI: [10.3390/brainsci10090626](https://doi.org/10.3390/brainsci10090626).
- [76] Quián Rodrigo y Panzeri Stefano. «Extracting information from neural populations: information theory and decoding approaches». En: *Nature Reviews Neuroscience, volume 10* (April 2009). DOI: [10.1038/nrn2578](https://doi.org/10.1038/nrn2578).
- [77] Rodríguez Roger y C Tuckwell Henry. «Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks.» En: *Phys. Rev. E* 54, 5585 (1 November 1996). DOI: <https://doi.org/10.1103/PhysRevE.54.5585>.
- [78] Brette Romain y Gerstner Wulfram. «Adaptive exponential integrate and fire model as an effective description of neuronal activity». En: *Journal of Neurophysiology Vol 94 No. 5* (2005). DOI: <https://doi.org/10.1152/jn.00686.2005>.
- [79] Herzog Rubén et al. «A mechanistic model of the neural entropy increase elicited by psychedelic drugs». En: *Scientific Reports* (2020). DOI: [10.1038/s41598-020-74060-6](https://doi.org/10.1038/s41598-020-74060-6).
- [80] Damoiseaux J S et al. «Reduce resting-state brain activity in the "default network" in normal aging.» En: *Cereb Cortex* 18(8):1856-64 (2008). DOI: [10.1093/cercor/bhm207](https://doi.org/10.1093/cercor/bhm207).
- [81] Ogawa S et al. «Brain magnetic resonance imaging with contrast dependent on blood oxygenation». En: *Biophysics Vol 87 pp 9868-9872* (1990). DOI: [10.1073/pnas.87.24.9868](https://doi.org/10.1073/pnas.87.24.9868).
- [82] Watanabe Satoshi. «Information theoretical analysis of multivariate correlation». En: *IBM Journal of Research and Development* (2010).
- [83] Olaf Sporns et al. «Organization, development and function of complex brain networks». En: *Trends in Cognitive Sciences* 8.9 (2004), págs. 418-425. ISSN: 1364-6613. DOI: <https://doi.org/10.1016/j.tics.2004.07.008>.
- [84] Frässle Stefan et al. «Whole-brain estimates of directed connectivity for human connectomics». En: *NeuroImage Volume 225, 117491* (15 January 2021). DOI: <https://doi.org/10.1016/j.neuroimage.2020.117491>.
- [85] Smith W Steven. *The Scientist and Engineer's Guide to Digital Signal Processing*. Second edition California Technical Publishing San Diego California, 1999.
- [86] Seifert Udo. «Universal bounds on entropy production from fluctuating coarse-grained trajectories». En: *Statistical Mechanics* (2025). DOI: <https://doi.org/10.48550/arXiv.2512.07772> Focus to learn more.
- [87] Pallarés Vicente et al. «Extracting orthogonal subject- and condition-specific signatures from fMRI data using whole-brain effective connectivity». En: *NeuroImage* 178 (2018), págs. 238-254. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2018.04.070>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811918303902>.

- [88] Gardiner C. W. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Second edition. Springer, 2012.
- [89] W. Seeley William et al. «Neurodegenerative Diseases Target Large-Scale Human Brain Networks». En: *Neuron*, Volume 62, Issue 1 (2009). DOI: <https://doi.org/10.1016/j.neuron.2009.03.024>.