



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**APRENDIZAJE AUTOMÁTICO PARA LA CONSTRUCCIÓN DE
CONECTOMA PEDIÁTRICO A PARTIR DE MRI**

TESIS
QUE PARA OPTAR POR EL GRADO DE:
**MAESTRO EN CIENCIA E INGENIERÍA DE LA
COMPUTACIÓN**

PRESENTA:
IB.RAMÓN DANIEL COTA AGUIAR

TURORA:
DRA. NIDIYARE HEVIA MONTIEL
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS APLICADAS Y EN
SISTEMAS

CIUDAD UNIVERSITARIA, CDMX, JUNIO 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICATORIA

A mis padres y hermana, por siempre apoyarme en todo lo que me dispongo a realizar y estar pendiente de mí en todo momento.

A mi novia, por haberme acompañado en este proceso y por darme ánimo en los momentos más difíciles.

A mis abuelos, de los que aprendí a no rendirme y a siempre salir adelante, por más difícil que sea la marea o por más alta que esté la montaña. Los extraño.

A todas las personas con enfermedades neurológicas, la ciencia y la tecnología avanzan, no pierdan la esperanza.

AGRADECIMIENTOS

Agradezco a mi tutora, la Dra. Nidiyare Hevia Montiel, por presentarme este proyecto y aceptar ser mi tutora, gracias por estar al pendiente de mi trabajo y por darme ese empujón de confianza para exponer avances de nuestro trabajo en conferencias, así como en apoyarme durante el tiempo que necesité para la conclusión de este trabajo.

A mis sinodales: Al Dr. Jorge Luis Pérez González, por apoyar mi trabajo desde sus inicios y brindarme su ayuda en momentos de incertidumbre. Además, de sus enseñanzas en el uso de herramientas que fueron de utilidad para este trabajo. Al Dr. Israel Vaca Palomares, por brindarnos el material necesario y facilitarnos las imágenes con las que se trabajó desde el protocolo inicial, además, de apoyarnos desde el campo de la psicología resolviendo dudas o aclarando resultados. Al Dr. José Antonio Neme Castillo, por sus enseñanzas en el análisis exploratorio de datos y detección de anomalías, dos materias que sin duda sirvieron para este trabajo y que además aplico en mi trabajo actual como científico de datos. Al Dr. Fernando Arámbula Cosío, por sus enseñanzas en la materia de aprendizaje automático en imágenes médicas, de la cual reforcé los algoritmos de aprendizaje automático que se aplicaron en este trabajo.

Agradezco al Dr. Juan Fernández Ruiz y a la Facultad de Psicología y a la Facultad de Medicina de la UNAM, por las imágenes de resonancia magnética con la que pudimos realizar este proyecto.

Agradezco al CONACYT, por la beca brindada, así como también al apoyo económico de los programas UNAM-PAPIIT IA102920, IT100220 e IT101422.

Agradezco a la UNAM y a todo su personal, no tengo duda que la grandeza de esta universidad se debe por quienes la formamos en su totalidad, y esto incluye a todo el personal docente, administrativo, soporte técnico y de intendencia, muchas gracias por todo, muchas gracias UNAM.

RESUMEN

La desnutrición infantil es un problema de salud mundial que ha afectado a millones de niños y puede tener efectos negativos en el desarrollo del cuerpo humano, incluyendo el cerebro. En este trabajo se analizaron los cerebros de 14 sujetos pediátricos que padecieron desnutrición y 12 sujetos pediátricos sanos, con edades entre 6 y 8 años, por medio de imagen de resonancia magnética (MRI, por sus siglas en inglés) estructurales y de difusión. Posterior al preprocesamiento de las MRIs, se extrajeron características morfométricas, de difusividad y de conectividad de las regiones cerebrales. Para encontrar aquellas regiones cuyas características fueran significativas, se obtuvo la diferencia estadística de ambas poblaciones usando los valores p de cada una de las características, donde se tomaron aquellas inferiores a 0.05, utilizando la prueba U de Mann Withney(MW) y la prueba de Kolmogorov-Smirnov(KS). A partir de la extracción con ambas pruebas, se realizaron subconjuntos que fueron evaluados por medio de métricas de desempeño de clasificación, como lo son la exactitud, sensibilidad, precisión y la puntuación F1. Los algoritmos utilizados fueron Naive Bayes, máquinas de soporte vectorial y bosques aleatorios. A partir de aquí, se seleccionó el subconjunto formado por regiones corticales, dado que brindó un mejor desempeño. La visualización de las conexiones formadas por las regiones seleccionadas se observan a través del enfoque del conectoma humano, técnica presentada al mundo por primera vez en el 2005 por Patric Hangman y Olaf Sporns et al. donde se pueden ver las conexiones físicas por medio de la visualización espacial o por medio del conectograma, es decir, ver el cerebro y sus conexiones como un grafo o red, de las cuales se pueden comparar métricas estándar en el área de la conectividad cerebral, descubrir patrones en conexiones o regiones con mayor importancia tanto para un grupo como para ambos.

ÍNDICE

1. INTRODUCCIÓN	1
1.1. El conectoma	1
1.2. Antecedentes	4
2. MARCO TEÓRICO	10
2.1. Imagenología por Resonancia Magnética	10
2.1.1. MRI ponderada en T1 (T1W)	11
2.1.2. Imágenes de difusión (DWI y DTI)	13
2.2. Macroconectoma	16
2.2.1. Teoría de grafos y conectómica	17
2.2.2. Visualización de grafos cerebrales	18
2.3. Obtención y selección de características	19
2.3.1. Obtención de características	22
2.3.2. Selección de características con enfoque estadístico	28
2.4. Clasificación con aprendizaje automático	30
2.4.1. Naive Bayes (NB)	31
2.4.2. Máquinas de soporte vectorial (SVM)	32
2.4.3. Bosques aleatorios (RF)	34
2.4.4. Métricas de desempeño en clasificación	35
3. MARCO DE REFERENCIA	37
3.1. Desnutrición infantil y deterioro cognitivo	37
3.2. Planteamiento del problema	38
3.3. Objetivo general	38
3.4. Objetivos Específicos	38
4. MATERIALES Y MÉTODOS	40
4.1. Descripción de base de datos	40
4.2. Metodología	41
4.2.1. Preprocesamiento de las MRI	41
4.2.2. Procesamiento de las MRI	49
4.2.3. Obtención de características	51
4.2.4. Selección de características	53
4.2.5. Clasificación y validación de conjunto de características	54
4.2.6. Construcción del conectoma	54

5. RESULTADOS Y DISCUSIÓN	57
5.1. Preprocesamiento de MRI	57
5.2. Procesamiento de MRIs	60
5.3. Características obtenidas	64
5.4. Selección de características por métodos estadísticos	64
5.5. Validación por métricas de clasificación	71
5.6. Visualización del conectoma	75
6. CONCLUSIÓN	86
REFERENCIAS	88
A. Conectograma con enfoque en grafos completos por características similares	96
A.1. Construcción del conectoma	96
A.2. Visualización del conectoma	98

ÍNDICE DE FIGURAS

1.1. Representación de una red: (a) matriz de conectividad y (b) grafo.	2
1.2. Tipos de conectoma.	4
1.3. Diferencias entre escalas macro, meso y micro.	5
1.4. Neuroanatomía y conectividad de <i>C. elegans</i> adulto.	6
2.1. Modalidades de análisis de las MRIs: (a) estructurales y (b) de difusión	11
2.2. MRI estructurales: (a) T1W y (b) T2W, en vistas coronal, sagital y axial	12
2.3. Vista sagital de MRI (a) por distintos números de ROIs y (b) ROIs del atlas de Desikan.	13
2.4. Tipos de difusión: (a) isotrópica y (b) anisotrópica, con sus valores propios. (c) Mapa de FA con representación a color de la orientación.	14
2.5. Mapas generados por la estimación del tensor: (a) DM, (b) DA y (c) DR.	15
2.6. Tipos de tractografías: (a) determinista, (b) probabilística y (c) trac- tograma.	16
2.7. Tipos de grafos: (a) no dirigido, (b) dirigido y (c) no dirigido con pesos.	18
2.8. Representación del (a) grafo como (b) matriz de incidencia y (c) de adyacencia.	18
2.9. Pasos para la creación del conectoma partiendo de las MRIs en sus modalidad estructural y de difusión.	19
2.10. Visualización anatómica o espacial del grafo.	20
2.11. Visualización circular o <i>conectograma</i>	20
2.12. Proyección anatómica de conectoma (a) y su respectivo conectogra- ma (b)	21
2.13. Visualización de las DTI tomando los valores propios correspon- diente a cada medida de difusión.	25
2.14. Ejemplo de vértices y aristas a partir de la matriz de conectividad del valor de FA, donde cada esfera representa una región, localizada en el centroide de dicha región.	26
2.15. <i>Pipeline</i> de los 3 métodos de selección de características.	29
2.16. Hiperplanos: a) una línea y b) un plano	32
2.17. Función de transformación (<i>Kernels</i>) a) lineal, b) polinomial y c) gaus- siana.	33

2.18. Ejemplo del funcionamiento del algoritmo de RF.	35
4.1. Vista axial, coronal y sagital de MRIs de sujeto control: (a) T1W y (b) DWI	41
4.2. Diagrama de la metodología.	42
4.3. MRI T1W vista axial (a) con ruido riciano y (b) filtrada con ANLM	45
4.4. MRIs (a) con escala de grises original y (b) con escala de grises normalizada.	46
4.5. Representación de la malla que representa la superficie cerebral y el mapa de surcos con su inflación esférica.	47
4.6. MRIs (a) con tejido extracerebral y (b) MRI con únicamente tejido cerebral.	47
4.7. Regiones de interés de la segmentación por el método de Desikan-Kalliany-Tourville.	50
4.8. Vista axial, sagital y coronal de la tractografía.	51
4.9. MRI vista sagital con (a) todas las regiones (vértices) y (b) con las regiones rostrales de azul turquesa, superior frontal y parietal de morado y el cerebelo de naranja, todas del hemisferio izquierdo.	55
4.10. Representación espacial y conectogramas del grafo de ejemplo con las regiones seleccionadas.	56
5.1. MRIs sin preprocesamiento de tres sujetos <i>control</i>	58
5.2. Proceso entre el (a) MRI original al aplicarle alineación rígida por el (b) MRI del atlas para obtener (c) una MRI alineada.	58
5.3. Eliminación de ruido en imagen por resonancia magnética (MRI). (a) MRI original, (b) ruido "riciano" removido e (c) imagen filtrada.	59
5.4. Extracción cerebral de MRI. En la figura (a) la imagen transformada de la original al espacio del atlas infantil y (b) su respectivo tejido cerebral únicamente.	60
5.5. MRI de difusión de un paciente control: Vista axial, sagital y coronal de MRI DWI.	60
5.6. MRIs con corte sagital (a) original, (b) con un subconjunto de regiones del hemisferio izquierdo y (c) con las 101 regiones.	61
5.7. MRI con regiones activas y el tracto Cingulum Parolfactoru izquierda: (a) vista sin transparencia en las regiones, (b) vista con transparencia y (c) vista con inclinación.	62
5.8. Tracto Cingulum Parolfactoru izquierdo teniendo contacto con tres regiones corticales.	62
5.9. Ejemplo de los tractos que se consideran entre 3 regiones (parcelas)	63
5.10. MRIs (a)DTI con los tractos coloreados según su orientación y (b)T1W con los tractos registrados a la modalidad.	64
5.11. Tractografía registrada con la T1W, donde (a) es la T1W, (b) es la tractografía sobrepuesta a la T1W, subconjunto de regiones corticales del hemisferio izquierdo sobrepuestas en la tractografía y (d) las 101 regiones sobrepuestas en la tractografía.	65

5.12. Construcción generalizada del grafo formado únicamente por las regiones corticales seleccionadas usando MW	76
5.13. Conectograma de sujeto Control 1 de (a) todas las regiones y (b) un cuarto de las regiones.	77
5.14. Comparación de la reducción de conexiones del sujeto Desnutrición 1.	78
5.15. Resultados de los conectogramas de sujetos control con todas las conexiones activas.	80
5.16. Resultados de los conectogramas de sujetos control filtrado a partir del cuartil Q_3	81
5.17. Resultados de los conectogramas de sujetos desnutrición.	82
5.18. Resultados de los conectogramas de sujetos desnutrición filtrado a partir del cuartil Q_3	83
A.1. KDE de dos grupos para la la asignación del valor más probable. . .	97
A.2. Ejemplos de conectogramas (a) con una sola característica para 3 regiones y (b) con tres tipos de características de 5 regiones.	99
A.3. Conectograma de sujeto Control 1 utilizando el enfoque de grafos completos por características en común.	100
A.4. Resultados de los conectogramas de sujetos control.	101
A.5. Resultado del conectograma del sujeto 1 del grupo Desnutrición. . .	102
A.6. Resultados de los conectogramas de sujetos desnutrición.	103

ÍNDICE DE TABLAS

1.1. Citas de trabajos haciendo uso del macroconectoma para clasificación o análisis de redes.	7
4.1. Información de las MRI	40
4.2. Matriz de conectividad de las 5 regiones seleccionadas.	55
5.1. Tabla de las características obtenidas a partir de las T1W y DWI	66
5.2. Descriptores de las características seleccionadas a partir de la prueba U de Mann Whitney	67
5.3. Regiones corticales de las características seleccionadas a partir de la prueba de Mann Whitney.	68
5.4. Regiones subcorticales de las características seleccionadas a partir de la prueba de Mann Whitney.	69
5.5. Descriptores de las características seleccionadas a partir de la prueba prueba Kolmogorov Smirnov	70
5.6. Regiones corticales de las características seleccionadas a partir de la prueba de Kolmogorov Smirnov.	71
5.7. Regiones subcorticales de las características seleccionadas a partir de la prueba de Kolmogorov Smirnov.	72
5.8. Resultados de clasificación utilizando MW	73
5.9. Resultados de clasificación utilizando KS	74
5.10. Resultados de los promedios de métricas de clasificación usando el conjunto con características de regiones únicamente corticales y únicamente subcorticales.	75
5.11. Resultados por medio de KDE de las métricas de las redes completas comparando sujetos sanos (ctrl) con sujetos desnutridos (desn).	84
A.1. Tabla para formar el grafo completo de la conexión del descriptor1 con las regiones 1,2 y 3.	98
A.2. Tabla de ejemplo con el que se construyó el conectograma de la Figura A.2	98

ABREVIATURAS

- MRI** Imagen por resonancia magnética
- sMRI** Imagen por resonancia magnética estructural
- dMRI** Imagen por resonancia magnética de difusión
- fMRI** Imagen por resonancia magnética funcional
- RMN** Resonancia magnética nuclear
- DTI** MRI por tensor de difusión
- DWI** MRI ponderada por difusión
- T1W** MRI ponderada en T1
- T2W** MRI ponderada en T2
- FA** Anisotropía fraccional
- DM** Difusividad media
- DA** Difusividad axial
- DR** Difusividad radial

Abreviaturas en las regiones cerebrales

- WM** Materia blanca
- GM** Materia gris
- CSF** Líquido cefalorraquídeo
- CC** Corteza de cerebelo
- ctx-** Prefijo referente a la corteza cerebral
- lh-** Prefijo referente a hemisferio izquierdo
- rh-** Prefijo referente a hemisferio derecho

CAPÍTULO 1

INTRODUCCIÓN

El conectoma humano completo pretende brindar información sobre la conectividad total del cerebro humano. Los avances con los alcances computacionales actuales han sido útiles para encontrar patrones y biomarcadores en el mapeo de las conexiones cerebrales, y así conocer estas características de sujetos sanos y pacientes con distintas condiciones neurodegenerativas como Alzheimer, esquizofrenia, demencia, entre otras.

1.1. El conectoma

La conectómica es un área en la neuroimagen que se ocupa de aplicar las técnicas de adquisición y análisis de imágenes asistidas por computadora, así como al mapeo estructural de conjuntos de circuitos neuronales o al sistema nervioso completo de organismos seleccionados usando métodos de alta velocidad, con la organización de los resultados en bases de datos y con aplicaciones de los datos [Lichtman and Sanes, 2008].

La frase “Somos nuestro conectoma” de [Seung, 2012], hace referencia a la identidad de cada persona o ser vivo, y como ese “cableado” de nuestro cerebro se forma de manera distinta en cada individuo, ya que nadie se ha desarrollado en exactamente el mismo ambiente, por lo que es de esperarse que la formación de neuronas y los tractos de mielina que comunican una región cerebral con otra sean diferentes.

El concepto “conectoma” fue propuesto por [Sporns et al., 2005] y [Hangmann, 2005] en el año 2005; en ambos documentos respectivamente, usan este concepto para referirse a la matriz de conectividad del cerebro, véase Figura 1.1, interpre-

table también como una gráfica, grafo o red compleja. La forma más sencilla de interpretar un grafo es mediante la unión de nodos o vértices por sus respectivos bordes o aristas; conforme van aumentando los componentes, la complejidad de la red también lo hace.

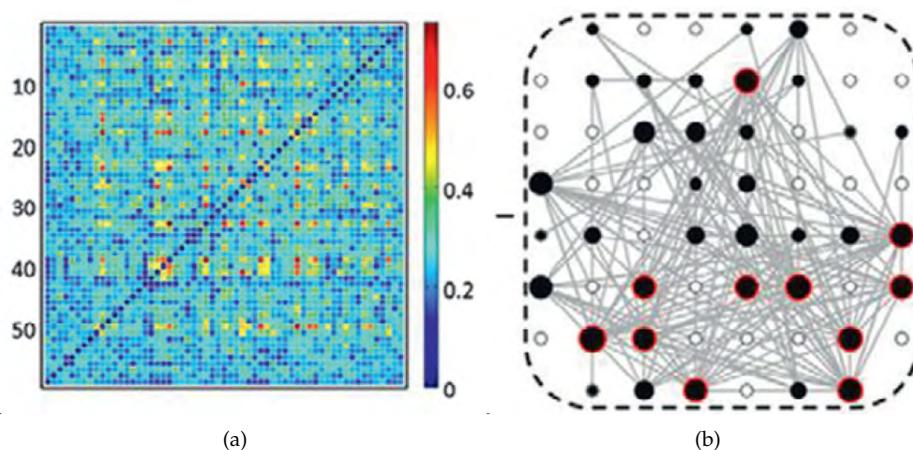


Figura 1.1: Representación de una red: (a) matriz de conectividad y (b) grafo. En la figura (a) tenemos una matriz de conectividad ponderada, donde los valores se representan con un mapa de calor siguiendo una escala de azul a rojo, siendo el primero un valor cercano a 0 y el segundo uno superior al 0.6. En la figura (b) tenemos un grafo ordenado en forma de malla que brinda información de conectividad de los vértices, así como características que aumentan el tamaño del radio del vértice, color o cambiar de color su perímetro. Fuente: [Fornito et al., 2013]

Sporns et al. propone el concepto y estrategias de análisis para conseguir el mapeo neuronal completo de un ser humano [Sporns et al., 2005]. Debido a las limitaciones computacionales de ese entonces, dicha tarea no podía completarse a corto plazo, por lo que la selección de una escala anatómica de estudio como un primer acercamiento, sería lo más adecuado. Menciona también que la obtención del conectoma podría proporcionar información estructural que pueda implementarse como parte de modelos computacionales a gran escala, así como proporcionar un conjunto de puntos de referencia estructurales que podrían facilitar la comparación e integración de los numerosos modelos especializados y de base estructural, que ya se han propuesto en neurociencia computacional.

Por otro lado, Hagmann menciona que durante la última década, han emergido una gran variedad de términos nuevos en el área de la investigación biomédica con el sufijo “-oma” (*ome* en inglés), el cual proviene del latín que significa *cuero*, denotando así el conjunto completo de algo. Algunas de estos términos son:

- **Genoma:** El cual es toda la información genética o material hereditario que posee un organismo

- **Proteoma:** La colección de proteínas funcionales sintetizadas por una célula o tejido vivo
- **Metaboloma:** El perfil metabólico de una determinada célula, tejido, fluido, órgano u organismo.
- **Conectoma:** Mapeo de las conexiones entre neuronas, incluyendo conexiones anatómicas o funcionales.

Hangmann menciona que así como el genoma es una entidad en sí mismo, el conectoma cerebral, más allá de ser un conjunto de todas las conexiones neuronales, también es una sola entidad. Enfatizando así el hecho de que la enorme capacidad de comunicación neuronal del cerebro y el poder computacional se basan críticamente en esta arquitectura de conectividad sutil e increíblemente compleja. Como se puede ver en la Figura 1.2, los autores Margulies et al. [Margulies et al., 2013] dividen los tipos de conectomas según la representación espacial de las regiones y los enlaces (vértices y aristas), los cuales pueden ser del tipo funcional, anatómico y de conexión.

El conectoma con representación funcional mapea las conexiones funcionales que reflejan la actividad sincronizada entre regiones del cerebro y se consideran como la base de funciones como percepción, movimiento y cognición. El conectoma con representación anatómica mapea las conexiones físicas considerando el espacio anatómico del cerebro y el de conexión busca la relación entre regiones y/o fibras, sin considerar su posición espacial.

Dado que el cerebro puede codificar experiencias y habilidades aprendidas en una forma que persiste durante décadas o más, la experiencia altera las conexiones entre las células nerviosas para registrar un recuerdo y recuperarlo más tarde. Tanto la experiencia sensorial que establece un recuerdo como su recuperación posterior son de hecho cadenas de potenciales de acción, pero en el medio, y persistiendo durante largos períodos, hay una entidad estructural física estable que contiene ese recuerdo. En este sentido, en la estructura se encuentra un mapa de todas las cosas que el cerebro ha puesto en la memoria: el “mapa de conectividad”. Un mapa de actividad cerebral que solo muestre cadenas de potenciales de acción, ciertamente sería un mapa incompleto, ya que la mayoría de los comportamientos y recuerdos no serán visibles en ninguna sesión de registro finita [Morgan and Lichtman, 2013]. La estructura anatómica corporal como la altura, masa, medidas del cuerpo como cintura, brazos, caderas, etc; brindan información sobre el estado

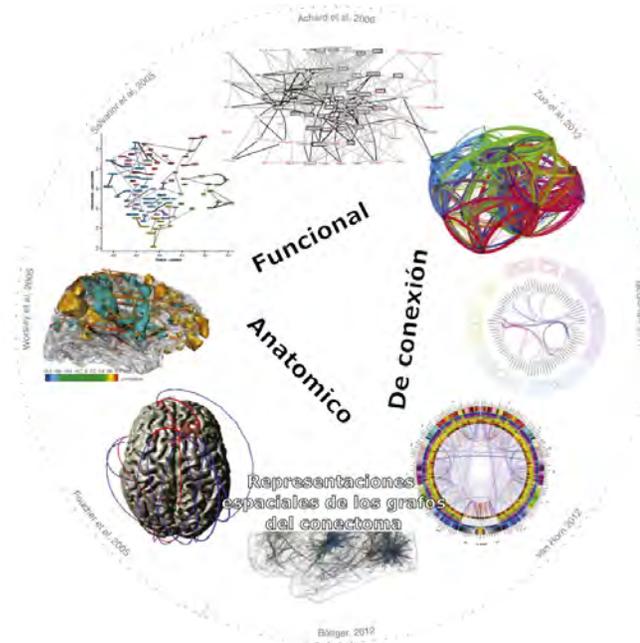


Figura 1.2: Tipos de conectoma. Se muestran algunos tipos de conectoma y la clasificación en representaciones funcionales, anatómicas y de conexión. Fuente: [Margulies et al., 2013] (Modificado)

de salud físico de un individuo, la morfometría del cerebro y sus conexiones físicas también brindan información sobre el estado de salud mental.

1.2. Antecedentes

En el año 2009 se anunció oficialmente el “Proyecto del Conectoma Humano”, el cual está formado por un grupo de investigadores de distintas áreas de investigación, universidades, ciudades y países, el cual se encuentra actualmente desarrollando esta investigación neurocientífica cuyo propósito es lograr un mapeo integral de las conexiones estructurales y dinámicas funcionales del cerebro completo [Bardin, 2012], el cual podría brindar conocimientos incomparables que respondan a la pregunta ¿Cómo funciona el cerebro?.

Los diagramas de circuitos del sistema nervioso se pueden considerar a diferentes niveles. Los avances para unir datos macro y microscópicos pueden ayudar a establecer un modelo estadístico realista que podría describir la conectividad a nivel ultra estructural, a este último, DeFelipe lo define como “sinaptoma” con el fin de diferenciar los términos, acotando el concepto “conectoma” únicamente al mapa de conexiones a nivel macroscópico [DeFelipe, 2010].

Además de la complejidad que conlleva el mapeo completo en micro escala, un

segundo desafío es que los elementos estructurales básicos del cerebro humano, en términos de conexiones y vértices de red, son difíciles de definir. Diferentes tipos de descripciones estructurales, véase Figura 1.3, podrían apuntar al menos a tres niveles bastante distintos de organización [Fornito et al., 2013], los cuales son:

- **Microescala**, en la cual el conectoma describe la disposición de las neuronas y de las sinapsis.
- **Macroescala**, cuyo objetivo es comprender la organización de las conexiones a gran escala del conjunto de las neuronas entre todas las áreas corticales y subcorticales del cerebro.
- **Mesoescala**, que comprenden los grupos o poblaciones de neuronas.

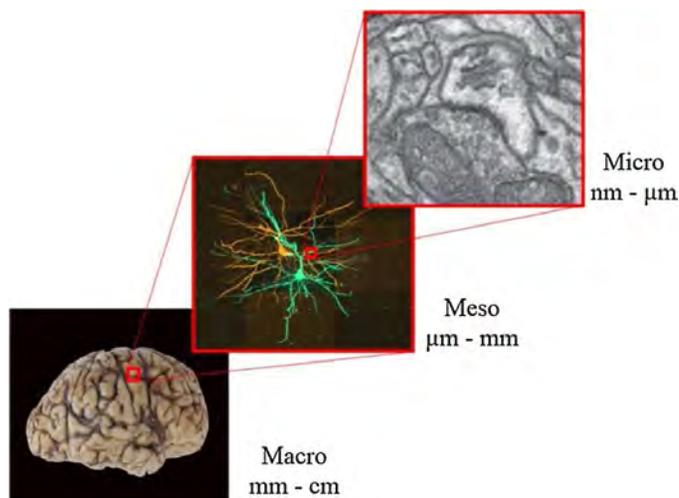


Figura 1.3: Diferencias entre escalas macro, meso y micro. Se muestran las distintas escalas en las que se trabaja en el área de neurociencias, donde la macro escala comprende medidas desde milímetros a centímetros, la meso escala de micrómetros a milímetros y la micro escala de nanómetros a micrómetros. Fuente: [Bennett et al., 2018]

Se estima que el cerebro humano contiene 100 mil millones de neuronas, estas células transmiten las señales a través de 100 billones de conexiones sinápticas [Vanderah and Nolte, 2018], por lo que la complejidad radica en el número de neuronas y conexiones, los primeros conectomas a nivel sináptico. Para la obtención del *sinaptoma* se han realizado a sistemas neuronales menos complejos como lo es la larva *C. elegans* (300 neuronas) o el cerebro central de la mosca *Drosophila melanogaster* (25,000 neuronas) [Scheffer and Xu, 2020]. Véase la neuroanatomía y conectividad una *C. elegans* adulto en Figura 1.4.

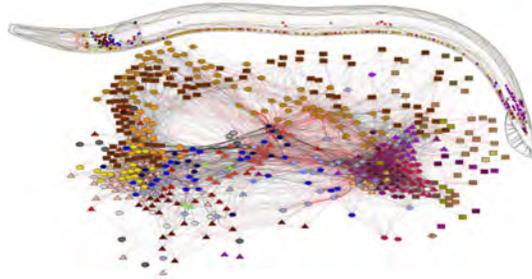


Figura 1.4: Neuroanatomía y conectividad de *C. elegans* adulto. Mapeo de todas las neuronas de la larva *C. elegans*. Fuente: [Cook et al., 2019]

Dada la complejidad del cerebro humano, los estudios a macroescala son ricos en información, por lo que la modalidad de imagen por resonancia magnética nuclear (RMN), es dominante en los estudios de neurociencia dado a su amplia disponibilidad, seguridad y resolución espacial. La imagen por resonancia magnética (MRI) estructural brinda información anatómica milimétrica y para inferir la conectividad estructural se utiliza la MRI ponderada por difusión, la cual proporciona representaciones con resolución milimétrica de tractos de materia blanca [Craddock et al., 2013]. La macroconectómica utiliza esta modalidad para el análisis de conectividad cerebral. El análisis morfométrico y de conectividad utilizando el *macroconectoma* ha sido utilizado para la clasificación entre pacientes con deterioro neurodegenerativo y pacientes cognitivamente normales (CN). Algunos trabajos reportados se encuentran en la Tabla 1.1.

Algunas condiciones y estilo de vida, están relacionados a un mayor riesgo de cambios cognitivos [Livingston et al., 2020], algunos de estos factores son la mala alimentación, la falta de sueño, alcoholismo, tabaquismo. Más de 200 millones de niños menores de 5 años en todo el mundo no logran alcanzar su pleno potencial de desarrollo [Walker et al., 2007]. Un estudio realizado a 171 pacientes infantiles sobrevivientes de desnutrición aguda severa, comprobó que estos tenían significativamente más probabilidades de estar en un grado más bajo en la escuela que los estudiantes controles y hubo diferencias significativas utilizando el test cognitivo *Cambridge Neuropsychological Test Automated Battery*, CANTAB [Lelieveld et al., 2019].

Recientemente, los avances y estudios del conectoma humano han sido utilizados progresivamente para investigar el cerebro pediátrico [Meoded et al., 2017], el cual conlleva de ciertos retos desde la adquisición de la MRI, hasta el proce-

Tabla 1.1: Citas de trabajos haciendo uso del macroconectoma para clasificación o análisis de redes.

Condición	Autor	Muestra	Resultados
Enfermedad de Alzheimer (AD)	[Shao et al., 2012]	enfermedad de Alzheimer: 17; cognitivamente normales: 21	Análisis de grafos y patrones de conectividad de los pacientes de enfermedad de Alzheimer. Precisión de clasificación.
	[Wang et al., 2019]	enfermedad de Alzheimer (AD): 110, DCL: 64, CN: 64	<ul style="list-style-type: none"> ▪ EA-CN: 97 % ▪ DCL-CN: 83 % ▪ EA-DCL: 97 %
Depresión de la vejez (DV)	[Pandya et al., 2017]	DV-DM:15; DV-MI:24; CN:30	Comparación entre propiedades de redes. Métricas de clasificación:
Trastorno del espectro autista (TEA)	[Payabvash et al., 2019]	TEA: 14; CN: 33	<ul style="list-style-type: none"> ▪ Precisión: 75.3 % ▪ Especificidad: 97 % ▪ Exactitud: 81.5 %
Trastorno de Esquizofrenia (TE)	[Gutiérrez-Gómez et al., 2019]	TE: 27; CN: 27	Clasificación entre 0 y 50 % del total de características. Exactitud \approx 65 %

DCL: deterioro cognitivo leve DM: difusividad media MI: memoria intacta. TEA: trastorno del espectro autista

samiento y análisis de las imágenes digitales. Actualmente, varios trastornos del neurodesarrollo, como el trastorno del espectro autista, el trastorno por déficit de atención con hiperactividad y la esquizofrenia, algunas malformaciones congénitas y el daño cerebral traumático, pueden considerarse *conectopatías* o síndromes de desconexión, por lo que el conectoma estructural puede ayudar a comprender la patogenia e implicación para las funciones cognitivas y conductuales de los trastornos cerebrales desde una perspectiva de red [Meoded et al., 2015] [Widjaja et al., 2015].

Dado que existen diferentes tipos de representaciones del conectoma, los componentes de una red compleja son descriptores que regularmente van más allá de únicamente vértices y aristas. Para estos descriptores, es necesario que se cuenten con biomarcadores con qué describirlos. La variedad de biomarcadores o descriptores ayudan a evaluar el comportamiento de la red [Freitas et al., 2019].

Para la extracción y selección de características, existen múltiples métodos estadísticos y de aprendizaje automático que, como su nombre lo dice, automatizan procesos para cualquiera de las dos tareas. El uso del aprendizaje automático en la neurociencia, ha sido bien recibido por la comunidad de neuroingeniería, por lo que se publican artículos cada año utilizando distintas técnicas del aprendizaje automático aplicadas en la obtención del conectoma humano, por mencionar a algunos [Shao et al., 2012] [Wang et al., 2019] [Larivière et al., 2019].

Los esfuerzos por la obtención del conectoma completo se siguen realizando hoy en día y con este trabajo se busca aplicar el enfoque del conectoma humano para encontrar diferencias en la conectividad entre pacientes pediátricos sanos y pacientes pediátricos que sufrieron desnutrición sus primeros años de vida.

En el capítulo siguiente se describen los conceptos principales para la obtención del conectoma, como lo son la imagenología por resonancia magnética en distintas modalidades, el conectoma macroscópico o macroconectoma, técnicas de selección de características así como algunos algoritmos de clasificación.

En el capítulo 3, definimos el problema de la desnutrición infantil y su relación con el deterioro cognitivo, se definieron los objetivos a alcanzar para la realización de este trabajo.

En el capítulo 4, se describen la base de datos que constó de MRIs de sujetos pediátricos sanos y con desnutrición, también se definió la metodología a seguir para lograr la construcción del conectoma que consta de preprocesar las MRIs, obtención y selección de características, clasificación y validación del subconjunto de características seleccionadas y la construcción del conectoma.

En el capítulo 5, obtenemos los resultados y la discusión de aquellos resultados que brinden información útil para el análisis de conectividad entre los sujetos sanos y desnutridos. En el capítulo 6, reportamos las conclusiones de este trabajo, así como futuras propuestas a partir de los resultados obtenidos.

CAPÍTULO 2

MARCO TEÓRICO

Para el estudio del conectoma es necesario entender cómo se obtiene y cómo interpretarlo. Existen diferentes propuestas para la construcción del mismo, según las necesidades del estudio se opta por una u otra, o se sugiere la creación de una nueva propuesta. Para la construcción del conectoma se utilizan las imágenes por resonancia magnética (MRIs, por sus siglas en inglés), para segmentar en regiones de interés (ROIs, por sus siglas en inglés) y extraer biomarcadores o descriptores que, por medio de estudios estadísticos o de aprendizaje automático, se seleccionen aquellos con mayor significancia para crear un mapa de conectividad y poder interpretarlo por la rama de teoría de grafos o de redes. Todos los conceptos para la obtención del conectoma serán definidos en este capítulo.

2.1. Imagenología por Resonancia Magnética

La imagen por resonancia magnética (MRI) se obtiene mediante un resonador magnético nuclear, siendo esta aplicación una técnica de imagen médica no invasiva utilizada para visualizar estructuras internas del cuerpo como tejidos, órganos y huesos. En el campo de la neurociencia se utilizan tres modalidades de MRI; la estructural (sMRI), de difusión (dMRI) y la funcional (fMRI). Para este trabajo nos enfocaremos únicamente en las modalidades sMRI y dMRI.

La técnica de adquisición de las sMRI es mediante la aplicación de un fuerte campo magnético y pulsos de radiofrecuencia en el cerebro que causa que los átomos de hidrógeno produzcan una señal con la que posteriormente se podrá reconstruir una imagen. Esta modalidad es utilizada para visualizar estructuras cerebrales anatómicas como tamaños y formas de los distintos tejidos cerebrales

como materia blanca (WM, por sus siglas en inglés), materia gris (GM) y líquido cefalorraquídeo (CSF). Además se pueden segmentar en regiones corticales y sub-corticales para analizarlas por separado [Maestú; et al., 2008]. Las visualizaciones de estos estudios se pueden observar en la Figura 2.1a.

Por otro lado las dMRI, son un tipo de MRI que mide la difusión de las moléculas de agua en el tejido. En esta modalidad, el movimiento aleatorio del agua es medido y usado para generar una imagen de las microestructura de los tejidos. La información de la difusión es utilizada para visualizar las fibras de la materia blanca y con el tensor de difusión se pueden información sobre orientación y direcciones de esas fibras, esta imagen es llamada imagen por tensor de difusión (DTI) [Maestú; et al., 2008]. La obtención del tensor y la visualización de las fibras de materia blanca se pueden observar en la Figura 2.1b.

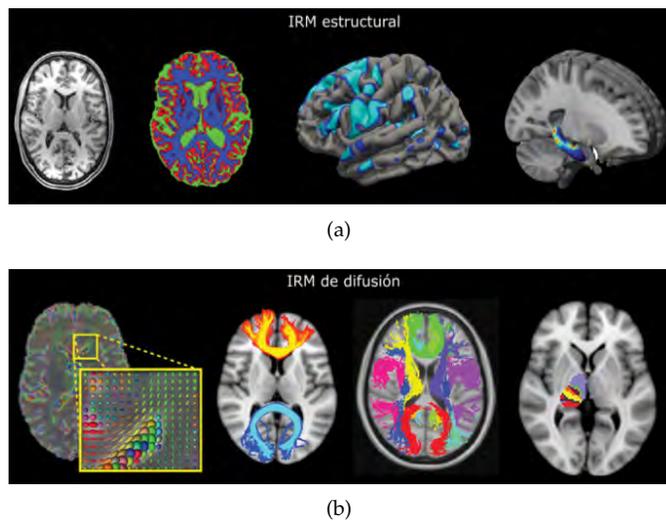


Figura 2.1: Modalidades de análisis de las MRIs: (a) estructurales y (b) de difusión. En la figura (a) se tiene de izquierda a derecha la MRI con corte axial con tejido extracerebral, el resultado de una segmentación de materia blanca, gris y CSF, el volumen correspondiente a una región y el volumen de una región visualizada en un corte sagital inclinado, respectivamente. Mientras que en la figura (b), se tiene primero el mapa de los tensores, dos fibras seleccionadas, el conjunto de distintas fibras en las últimas dos MRI. Fuente: [Jenkinson and Chappell, 2018](Modificado)

2.1.1. MRI ponderada en T1 (T1W)

En la MRI estructural, las dos ponderaciones son MRI ponderada en T1 (T1W) y MRI ponderada en T2 (T2W), véase Figura 2.2, como se pueden observar las intensidades o niveles de gris de ambas difieren. Estos niveles de grises en las imágenes dependen de la densidad de protones del tejido y el instante en que la imagen se tome. Como se observa en la Figura 2.2a, la T1W muestra en blanco la

WM, gris la GM y negro el CSF, mientras que la T2W muestra blanco el CSF, gris la GM y negro la WM.

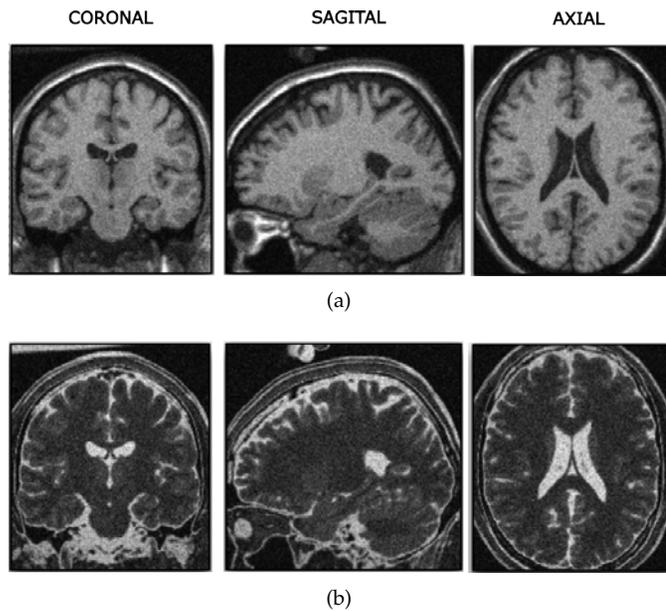


Figura 2.2: MRI estructurales: (a) T1W y (b) T2W, en vistas coronal, sagital y axial. La figura (a) muestra las características de un T1W, el cual destaca por mostrar en tonos de grises bajos la WM, mas altos la GM y más altos el CSF, y la figura (b) muestra las de MRI-T2W, que es lo contrario. Fuente: [Modat et al., 2014](Modificado)

Para esta trabajo se utilizaron MRI ponderada en T1 (T1W), la cual provee un buen contraste entre materia blanca (WM) y materia gris (GM), lo cual es de utilidad para encontrar los límites entre los distintos tejidos, lo cuál es útil para obtener las ROIs.

Regiones de Interés

Un cerebro puede dividirse en grandes volúmenes como WM y GM, y de cada volumen se pueden generar una cantidad de ROIs limitada únicamente por el número de vóxeles de la MRI. En la Figura 2.3 se puede observar una segmentación 2D de la GM, donde la parcelación o segmentación en ROIs del cerebro, implicó dividir la GM en conjuntos de distintas ubicaciones contiguas en el espacio, de modo que las ubicaciones dentro de cada parcela comparten similitudes en una o más propiedades. La segmentación de estructuras cerebral regularmente se llevan a cabo mediante el registro de un atlas con coordenadas espacialmente predefinidas y se utilizan en numerosos estudios de MRI tanto estructurales como de difusión. Un ejemplo de atlas es el realizado por [Desikan et al., 2006], cuyas ROIs se dividieron basándose en conectividad funcional y similitud anatomica.

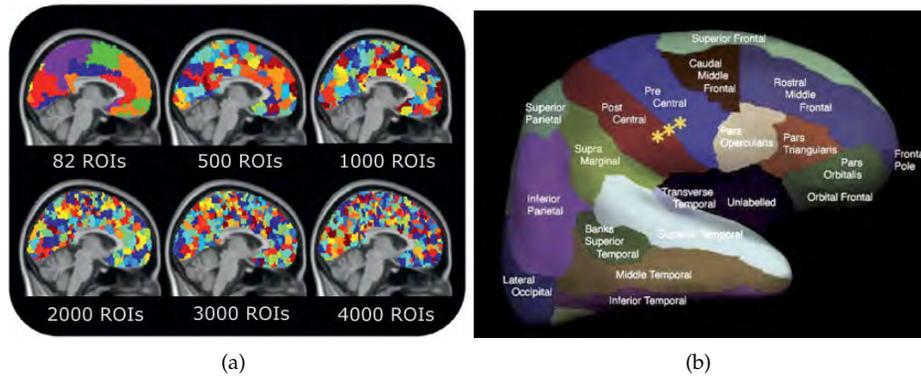


Figura 2.3: Vista sagital de MRI (a) por distintos números de ROIs y (b) ROIs del atlas de Desikan. La figura (a) muestra el distinto número de regiones de interés se pueden estudiar, sin embargo el número de ROIs por lo regular se eligen dependiendo el estudio y utilizando atlas cuyas regiones han sido estudiadas y segmentadas a mano por expertos en neurociencia, como lo es el atlas de la figura (b), el cual es utilizado en múltiples estudios cerebrales. Fuentes: [Fornito et al., 2013] y [Desikan et al., 2006]

2.1.2. Imágenes de difusión (DWI y DTI)

Con la modalidad imagen ponderada por difusión (DWI) es posible medir la difusión de las moléculas de agua en tejidos, órganos y huesos, funciona aplicando un gradiente de campo magnético para que las moléculas de agua se difundan en direcciones aleatorias, a partir del tensor de difusión, con el que se obtiene la imagen por tensor de difusión (DTI). La difusión refleja la presencia de una arquitectura estructural del sistema nervioso central (CNS) y sus cambios y alteraciones neuronales [Maestú; et al., 2008]. Las secuencias de difusión en la resonancia magnética nuclear expresan el movimiento aleatorio microscópico de las moléculas en un medio líquido, visualizando esta propiedad debido a la pérdida de la intensidad de señal de las moléculas en movimiento [Moseley et al., 1990].

Imagen de tensor de difusión

A través de la estimación del tensor de difusión, utilizando los b -valores, se obtiene la imagen por tensor de difusión. Los b -valores se utilizan para cuantificar la cantidad de ponderación de difusión aplicada a cada imagen. Para obtener medidas de difusión que permitan obtener una tractografía es necesario realizar cálculos con los valores propios, descritos por los modelos isotrópicos y anisotrópicos.

Los modelos principales de difusión son la isotrópica, en la que las moléculas se mueven en todas las direcciones por igual, donde la difusión es libre y equiparable en todas las direcciones, como se observa en la Figura 2.4a; y la anisotrópica que es el movimiento parcial o totalmente restringido, es decir que la difusión está restringida por las barreras presentes en el medio, siendo mayor una en dirección que las otras, véase Figura 2.4b.

Un valor escalar derivado de DTI es la anisotropía fraccional (FA, por sus siglas en inglés), la cual provee una medida del grado de direccionalidad de la difusión del agua en el tejido. Con este valor es posible cuantificar la organización de las fibras de WM así como identificar regiones donde las fibras están más o menos alineadas. La FA es calculada a partir de la varianza normalizada de los valores propios del tensor de difusión en cada vóxel, por lo que valores altos de FA indican que la difusión del agua tiene más restricción en cuanto a su direccionalidad, por lo tanto es un indicador de un alto grado de organización de fibras; y valores bajos indican que la difusión del agua es más isotrópica, por lo tanto su grado de organización de fibras es menor [Le Bihan et al., 2001]. Los mapas de FA proporcionan una forma cuantitativa para caracterizar la organización de las fibras de WM, lo cual es de utilidad para estudios de trastornos neurológicos y psiquiátricos, véase un ejemplo de estos mapas en la Figura 2.4c.

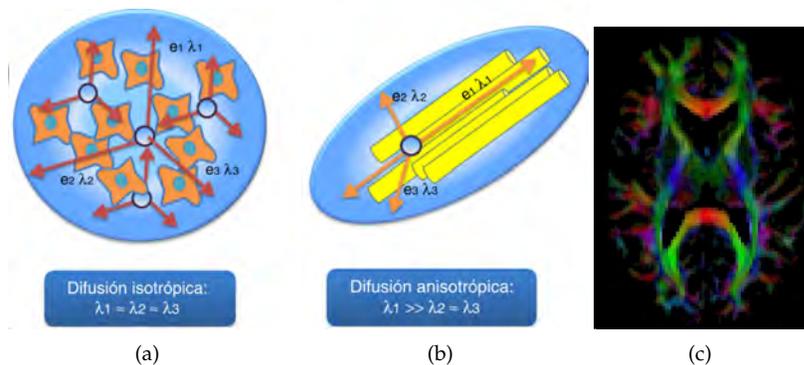


Figura 2.4: Tipos de difusión: (a) isotrópica y (b) anisotrópica, con sus valores propios. (c) Mapa de FA con representación a color de la orientación. En las figuras (a) y (b) se muestran los valores propios para cada tipo de difusión, la puramente isotrópica tendrá valores propios idénticos, mientras que la difusión anisotrópica tendrá uno mucho mayor que los otros dos. En la figura (c), se muestra el mapa de anisotropía fraccional, donde el color depende de la orientación de la difusión en ese vóxel. Fuente: [Martín Nogueroles and Martínez Barbero, 2017]

Otro concepto importante es el de los valores de difusividades, los cuales son la difusividad media (DM), difusividad axial (DA) y difusividad radial (DR). La DM se obtiene sumando los valores propios y restarlos entre el total, representa

el desplazamiento del agua en el compartimento. La DA se obtiene con el valor propio principal y su significado biológico es la integridad axonal y por último la DR, la cual se obtiene con la media de los dos menores valores propios, véase Figura 2.5.

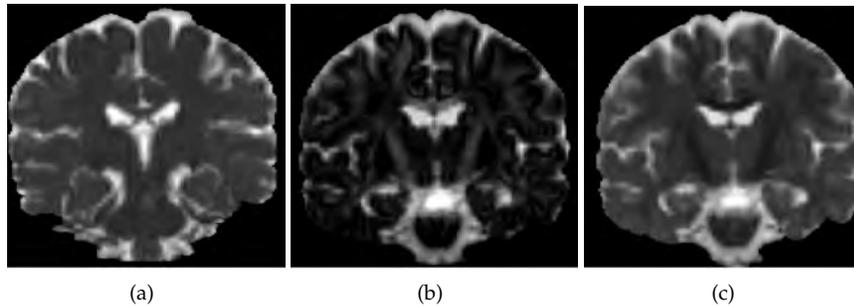


Figura 2.5: Mapas generados por la estimación del tensor: (a) DM, (b) DA y (c) DR. A partir de la estimación de los tensores se obtienen las DTIs con sus respectivas intensidades, con el cual se realizan los seguimientos de fibra. Fuente: [Moura et al., 2019](Modificado)

Tractografía

Una de las primeras aplicaciones de la DTI fue la visualización tridimensional de tractos de materia blanca utilizando algoritmos de tractografía. La variante más simple, llamada tractografía determinista, consiste en sembrar un vóxel específico y luego propagar una línea de corriente que sigue la dirección principal de difusión del agua en cada vóxel hasta que se alcanza algún criterio de terminación, vease Figura 2.6a, las líneas de corriente (líneas blancas) siguen la dirección principal de difusión del agua vóxel por vóxel. Los círculos blancos representan puntos de semilla. Los cilindros de colores representan la dirección principal de difusión del agua en cada dirección, que se utiliza como base para la propagación aerodinámica. En la tractografía probabilística, se ajusta una distribución de probabilidad de la dirección preferida de difusión del agua en cada vóxel y determina la probabilidad de que una línea de corriente se propague en cualquier dirección particular. Por lo tanto, donde la tractografía determinista usa una estimación puntual de la dirección preferida de difusión en cada vóxel, la tractografía probabilística modela la incertidumbre de esta estimación y, por lo tanto, proporciona límites de confianza en la trayectoria de la vía reconstruida, como se muestra en la Figura 2.6b, se generan y propagan muchas líneas de corriente de acuerdo con un modelo de densidad de probabilidad de las orientaciones de las fibras en cada vóxel, lo que permite una estimación de la incertidumbre en la trayectoria de un tramo determinado. En el

recuadro, las densidades de probabilidad están representadas por los conos que se muestran dentro de cada vóxel. Las estimaciones precisas de la distribución de probabilidad requieren el muestreo de más direcciones de difusión y, por lo tanto, secuencias de resonancia magnética especializadas para la adquisición de datos ponderados por difusión. Los tractos se pueden sembrar desde todos los vóxeles dentro del volumen de materia blanca, o entre todos los pares de regiones de materia gris, para generar tractogramas de todo el cerebro, véase Figura 2.6c, [Fornito et al., 2013].

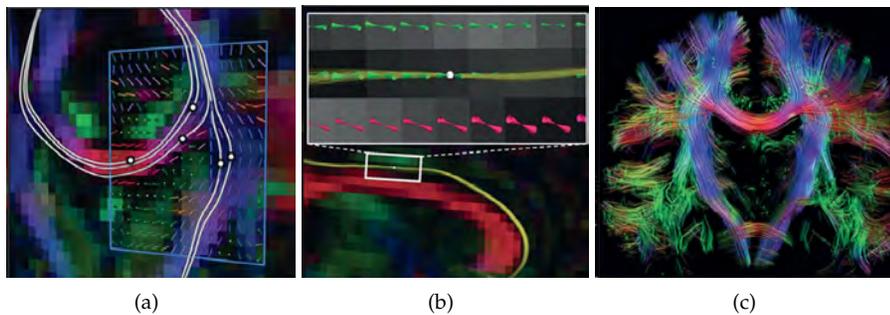


Figura 2.6: Tipos de tractografías: (a) determinista, (b) probabilística y (c) tractograma. Métodos para obtener las tractografías con (a) enfoque determinista y (b) con enfoque probabilístico. Usando cualquiera de los dos se obtiene un tractograma como el que se observa en la figura (c) que tiene código de colores dependiendo su sentido de elongación. Fuente: [Tournerier et al., 2011]

2.2. Macroconectoma

La idea central del conectoma es tomar cada neurona como el nodo o vértice del grafo y cada sinapsis como un borde o arista, pero en los métodos de imagenología no invasivos no es posible llegar a esa resolución, por lo que la escala macro es la utilizada cuando se trabaja con MRI, que tiene un vóxel de 1 mm^3 , el cuál contiene un aproximado de 20,000 a 30,000 neuronas realizando miles de millones de sinapsis [Logothetis, 2008]. Aunque el conectoma con mayor número de vértices es aquél cuyo número de region de interés (ROIs) sea el equivalente a cada vóxel, en la macroconectómica se trabaja con regiones cerebrales parceladas por un atlas que define funciones del cerebro, por lo tanto es necesario definir los vértices y las aristas que se usarán en el estudio del grafo.

2.2.1. Teoría de grafos y conectómica

Los vértices y aristas son la pieza fundamental de la teoría de grafos y definir los pertenecientes al sistema nervioso no hay una única respuesta. Las ROIs definen el número de vértices del grafo y las conectividades entre esas regiones por medio de la tractografía definen el peso de las aristas o si existe o no una conexión por medio de algún criterio.

El análisis de redes, y en particular la teoría de grafos, ofrece nuevas formas de caracterizar cuantitativamente los patrones anatómicos. De acuerdo con la teoría de grafos, las redes cerebrales estructurales pueden describirse como gráficos compuestos por nodos (vértices) que denotan elementos neuronales (neuronas o regiones cerebrales) que están vinculados por bordes (aristas) que representan conexiones físicas (sinapsis o proyecciones axonales) [Bullmore and Sporns, 2009].

Gráfica o grafo

Un grafo no dirigido G consiste en un conjunto V de vértices y un conjunto E de aristas tal que cada arista $e \in E$ se asocia con un par no ordenado de vértices. Si existe una arista única e asociada con los vértices u y w , se escribe $e = (v, w)$ o $e = (w, v)$. En este contexto, (v, w) denota una arista entre v y w en una gráfica no dirigida y no es un par ordenado, ejemplo en la Figura 2.7a. Un grafo dirigido (o digrafo) G consiste en un conjunto V de vértices y un conjunto E de flechas tales que cada arista $e \in E$ está asociada con un par ordenado de vértices. Si hay una arista única e asociada con el par ordenado (v, w) de vértices, se escribe $e = (v, w)$, que denota una arista de v a w , ejemplo en la Figura 2.7b. También existen los grafos ponderados, es decir, que a su arista le corresponde un peso, véase Figura 2.7c. Estas gráficas son útiles cuando se desean encontrar rutas de menor peso o costo [Johnsonbaugh, 2005].

Matrices de conectividad

Una forma de estudiar las gráficas es a través de matrices. La matriz de adyacencia nos indica el número de conexiones que hay entre un vértice, véase Figura 2.8b, en esta representación también podemos encontrar su versión ponderada en la cual se sustituirá el número de conexiones y se colocará el peso que existe en la arista que conecta dichos vértices. La matriz de incidencia, véase Figura 2.8c

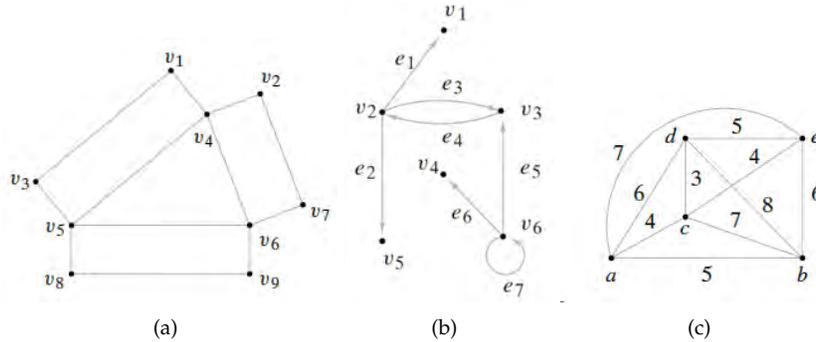


Figura 2.7: Tipos de grafos: (a) no dirigido, (b) dirigido y (c) no dirigido con pesos.
Fuente: [Johnsonbaugh, 2005]

se obtiene acomodando los vértices en las filas y las aristas en las columnas, cada columna determinará con un 1 o con el peso de dicha conexión.

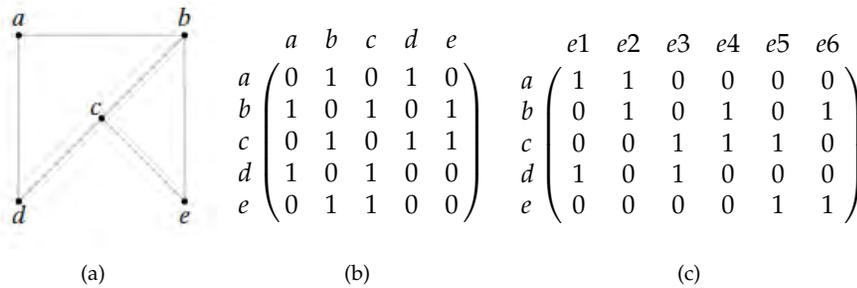


Figura 2.8: Representación del (a) grafo como (b) matriz de incidencia y (c) de adyacencia.
Fuente (a): [Johnsonbaugh, 2005].

2.2.2. Visualización de grafos cerebrales

Existen diversas formas de representar un grafo y las características del mismo; como se mencionó anteriormente, una matriz de adyacencia o incidencia es una de esas formas, sin embargo interpretar una matriz y conocer ciertos atributos del grafo por medio de la misma se vuelve complejo conforme aumentan el número de vértices o el número de aristas. Como se observa en la Figura 2.9, los pasos para construir la matriz de conectividad de un grafo por medio de las regiones de la parcelación o segmentación anatómica (vértices) y tractos de la tractografía (aristas).

Algunas de estas visualizaciones son la anatómica y el conectograma. El grafo anatómico utiliza la información espacial de las regiones o puntos de interés para analizar del cerebro enfatizando el espacio donde ocurre cierto efecto, mientras que el conectograma es una técnica de visualización circular que ayuda a visualizar la

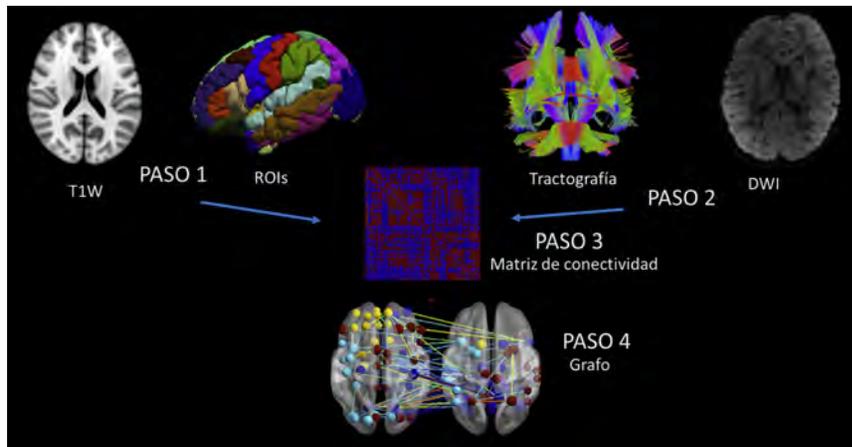


Figura 2.9: Pasos para la creación del conectoma partiendo de las MRIs en sus modalidades estructural y de difusión. Se observa que el primer paso es la segmentación anatómica que representará los vértices, el segundo obtener la tractografía que representará las aristas, el tercero la unión de vértices y aristas por medio del registro entre las imágenes T1W y DWI para crear una matriz de conectividad y como paso 4, se agregó la visualización del grafo. Fuente: [Meoded et al., 2020] (modificado).

conectividad entre ciertas regiones.

Como se muestra en la Figura 2.10, en las proyecciones anatómicas, los vértices están en las coordenadas del centroide de la región cerebral y las aristas se representan con la unión simple entre un vértice y otro.

Como se observa en la Figura 2.11, el *conectograma* es un diagrama circular utilizado para representar patrones de conectividad de una red. El *conectograma* en neuroimagen se utiliza para representar conexiones entre regiones del cerebro configurando parámetros como el color de la arista o el grosor para sobresaltar ciertas características o pesos de la conexión [Fornito et al., 2013].

En la Figura 2.12, se observa la relación entre una visualización espacial con el conectograma circular, cada una con su propósito y ambas parten de la misma matriz de conectividad.

2.3. Obtención y selección de características

Las características son propiedades medibles o que categorizan un objeto. Estas, pueden ser de tres tipos de datos: binarios, categóricos o continuos. El objetivo de la selección de características es la disminución de la dimensionalidad encontrando aquellas que definan un objeto con información relevante que sea útil para la construcción de un modelo de aprendizaje automático. Para estas tareas se pueden usar el enfoque paramétrico y el no paramétrico, pero dado que se ha demos-

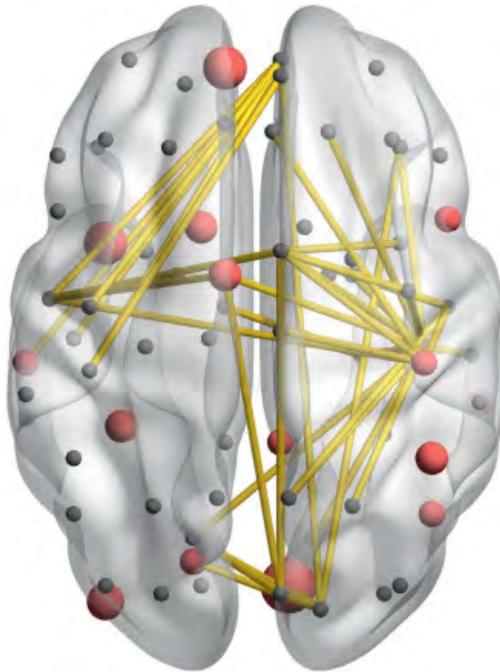


Figura 2.10: Visualización anatómica o espacial del grafo. En la visualización anatómica es posible observar el punto exacto donde se encuentra el centroide de la región de interés. Usando esta técnica se pueden observar conexiones o importancias de las regiones, destacándolas por color o por tamaño. Fuente: [Fornito et al., 2013]

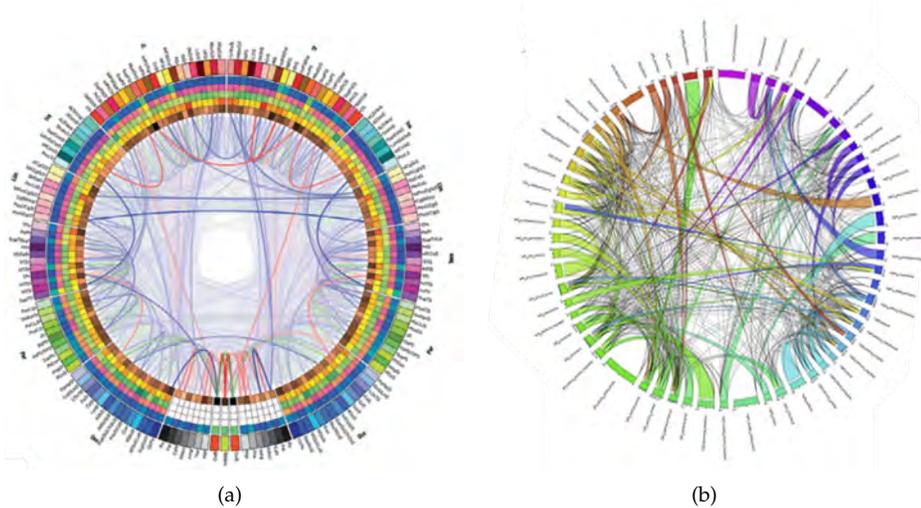


Figura 2.11: Visualización circular o *conectograma*. En la figura (a) se muestra un conectograma con diversas características dadas sus respectivas escalas de colores, la figura (b) es un conectograma sencillo donde cada región tiene su respectivo espacio en el perímetro del círculo. En ambas se muestran el grado de conectividad tanto de cada región como el del grafo. Fuente: [Fornito et al., 2013] [Meoded et al., 2020] (Modificados)

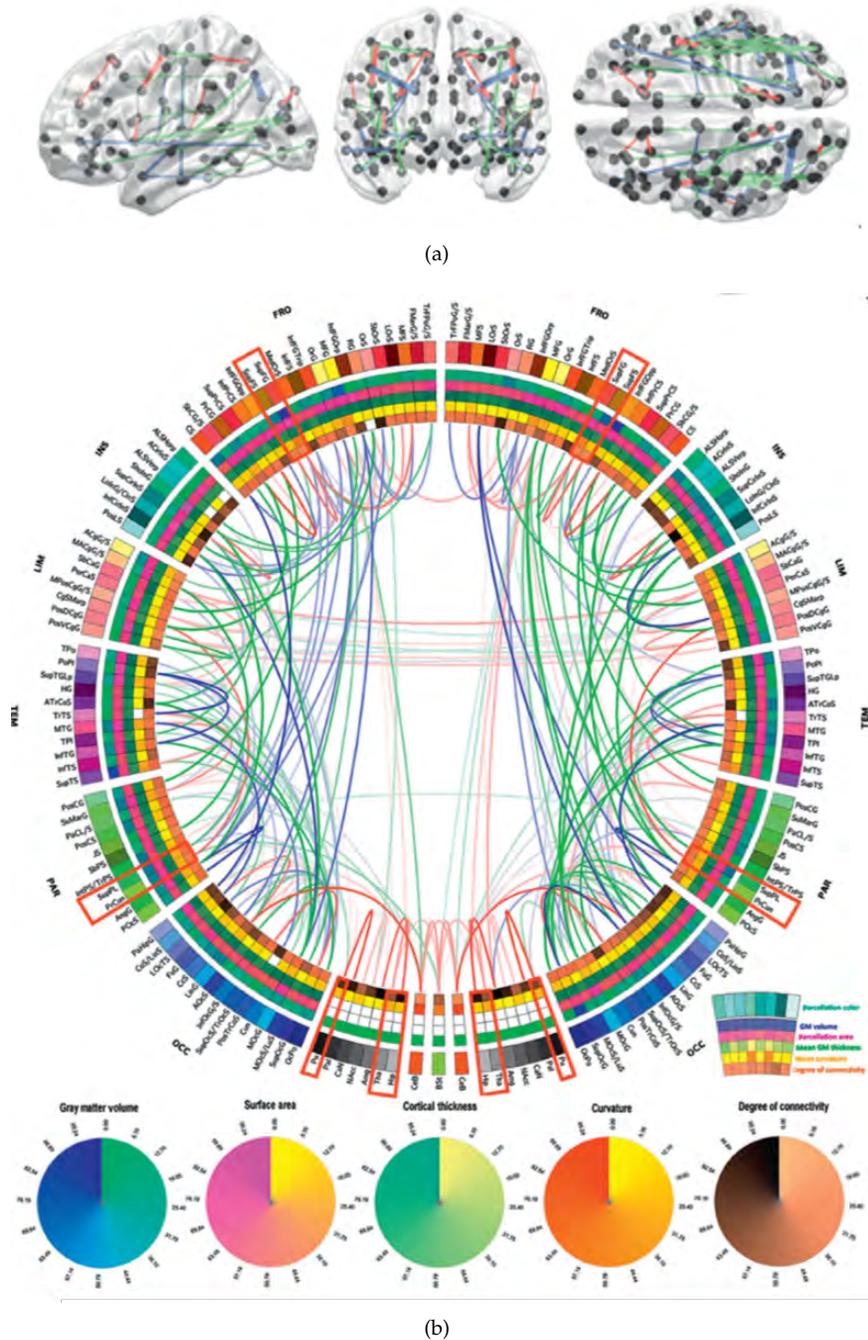


Figura 2.12: Proyección anatómica de conectoma (a) y su respectivo conectograma (b). Aquí vemos una proyección anatómica (a) y un conectograma (b) de conexiones cuya eliminación conduce a un cambio estadísticamente significativo en la topología de la red global. La proyección anatómica se ha umbralizado a un nivel más estricto que el conectograma para facilitar la visualización. Los colores de borde en (a) y (b) representan enlaces con anisotropía fraccional alta (rojo), promedio (verde) o baja (azul). Fuente: [Fornito et al., 2013]

trado que los métodos estadísticos no paramétricos pueden ser mas robustos que los métodos paramétricos, cuando el tamaño de la muestra es pequeña [Wilcox, 2016], se explicarán únicamente los métodos no paramétricos que se utilizarán. Estos, son útiles para hacer inferencias cuando se tienen dudas respecto a las suposiciones que forman la base de una distribución normal, por lo que una de sus múltiples ventajas es la capacidad de detectar diferencias entre poblaciones cuando no se satisface normalidad.

2.3.1. Obtención de características

A partir de distintos métodos y técnicas aplicadas a las MRIs, es posible obtener características que describan una población o grupo de estudio. Para este trabajo, definimos como:

- **Descriptores:** Aquellos que se obtendrán de cada región, es decir el atributo en su forma general, ejemplo: Volumen es un descriptor morfométrico, Anisotropía Fraccional es un descriptor de difusión y la centralidad *pagerank* es un descriptor de redes.
- **Característica:** La unión entre un solo descriptor con una sola región del cerebro, siendo un ejemplo: El *volumen* del *tronco encefálico* es una característica morfométrica, la media de la Anisotropía Fraccional de la materia blanca izquierda es una característica de difusión y la centralidad *pagerank* es una característica de red.

Descriptores morfométricos

El análisis por descriptores morfométricos intenta encontrar aquellos parámetros descriptivos, generalmente numéricos, que representan la información de importancia en la imagen. Para la extracción de estas características se utilizo la medida volumétrica del voxel, que en MRI estructural equivale a 1 mm^3 y se obtuvieron para cada región segmentada.

- **Volumen:** El volumen de la región puede resumirse al conteo o suma de los vóxeles que conforman el objeto tridimensional, por lo que una región de n vóxeles tendrá un volumen de $n \text{ mm}^3$.
- **Volumen normalizado:** Es el porcentaje que representa dicha región de vóxeles dividida entre el total de vóxeles del cerebro completo.

- **Área de la superficie:** Dado que la cara de cada vóxel es de 1 mm^2 , corresponde a la suma de las áreas de las caras visibles de la región.
- **Redondez:** Es una medida de cuán cerca se parece la forma de una región cerebral a la de una esfera perfecta. La redondez se rige por la siguiente fórmula:

$$\psi = \frac{\pi^{\frac{1}{3}}(6V_p)^{\frac{2}{3}}}{A_p}. \quad (2.1)$$

Donde:

ψ coeficiente de redondez

V_p volumen

A_p área de la superficie

La esfericidad de una esfera es 1 y cualquier objeto tridimensional que no sea esfera será menor a 1.

- **Aplanamiento:** Se interpreta el volumen como un elipse y se calcula con los dos ejes del elipse con la siguiente fórmula:

$$f = \frac{a - b}{a}. \quad (2.2)$$

Donde:

f coeficiente de aplanamiento (*flatness*)

a eje mayor

b eje menor

El aplanamiento es 1 cuando el objeto es totalmente plano (bidimensional) debido a que su eje b sería 0 y la división $\frac{a}{a} = 1$.

- **Diámetro de Feret:** Es una medida del tamaño de un objeto a lo largo de una dirección específica. En general, se puede definir como la distancia entre los dos planos paralelos que restringen el objeto perpendicular a esa dirección.
- **Elongación:** Se calcula a partir de la raíz cuadrada de la relación entre el segundo momento principal más grande y el más pequeño.

$$e = \sqrt{\frac{\gamma_1}{\gamma_0}}. \quad (2.3)$$

Donde:

e coeficiente de elongación

γ_1 momento más grande

γ_2 momento más pequeño

- **Compacidad discreta:** La compacidad relaciona el área de superficie envolvente con el volumen y es minimizada por una esfera. La medida de compacidad discreta es maximizada por un cubo. Considerando el volumen directamente proporcional al área de superficie de contacto, la compacidad es la relación del área de la superficie envolvente con respecto al área de superficie de contacto [Carrillo-Bermejo et al., 2018].

El área de la superficie (A) envolvente se obtiene considerando un objeto tridimensional compuesto por un número finito de n vóxeles, donde A corresponde a la suma de las áreas de los polígonos del plano externo de los vóxeles que forman las caras visibles del objeto, mientras que la superficie de contacto es la suma de las áreas de las superficies que son comunes a dos vóxeles [Bribiesca, 2008].

- **Tortuosidad:** Se refiere a la suma de todos los valores absolutos de los elementos de la cadena definida por [Bribiesca, 2000], expresada en la siguiente ecuación:

$$\tau = \sum_{i=1}^n |a_i|. \quad (2.4)$$

Donde:

τ tortuosidad discreta

a valores absolutos

La tortuosidad se mide a partir de los contornos discretos de las regiones, donde los segmentos de línea recta tienen una longitud de acuerdo con el tamaño y el número de vóxeles. La τ discreta para objetos 3D se define como la suma de todos los valores absolutos de las curvaturas de todos los contornos concatenados presentes en la imagen [Carrillo-Bermejo et al., 2018].

Descriptores de difusión

Existen varias formas de cuantificar la forma de los tensores en cada vóxel. En DTI hay 4 medidas que son las más utilizadas; anisotropía fraccional, difusividad media, axial y radial. Estas medidas se relacionan directamente con el valor de los tres principales valores propios del tensor, indicados en la Figura 2.13 con λ_1 , λ_2 y λ_3 .

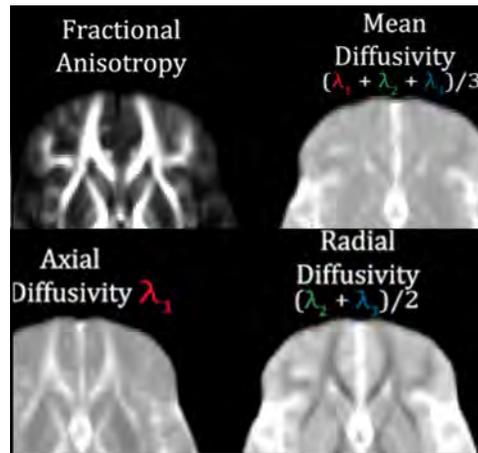


Figura 2.13: Visualización de las DTI tomando los valores propios correspondiente a cada medida de difusión. Como se observa, la difusividad media se obtiene por medio del promedio de valores propios, la axial únicamente corresponde al primer valor propio y la radial al promedio de los valores propios 2 y 3. Fuente: [Fornito et al., 2013]

- Anisotropía fraccional:** describe el grado de anisotropía de difusión en cada vóxel y refleja el grado de direccionalidad de las estructuras celulares dentro de los tractos de fibras. Para ello, los tensores de WM se reconstruyen para rastrear la orientación de la fibra tridimensional de vóxel a vóxel, lo que demuestra la arquitectura de la fibra asumiendo así, que la dirección de la difusión menos restringida presenta la orientación de las fibras nerviosas. En la tractografía de fibra, la propagación de línea es el método más común, en él, cuando establecemos una o varias ROIs en el curso de las proyecciones de fibra en relación con un punto de referencia anatómico conocido, el seguimiento comienza con un vóxel desde el cual se propaga según el eje principal [Le Bihan et al., 2001]. La FA proporciona la diferencia relativa entre el valor propio más grande en comparación con los demás; cuantifica la fracción de difusión que es anisotrópica. Su ecuación es la siguiente:

$$FA = \sqrt{\frac{1}{2} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}}{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \quad (2.5)$$

- **Difusividad media:** el valor promedio de los tres valores propios. Ésta representa el desplazamiento del agua en el compartimiento.
- **Difusividad axial:** se obtiene con el valor propio principal y su significado biológico es la integridad axonal. Cuantifica qué tan rápido se difunde el agua a lo largo de las fibras axonales.
- **Difusividad radial:** se obtiene con la media de los dos menores valores propios. Cuantifica la rapidez con la que el agua se difunde a través de los haces axonales. Se considera como una aproximación de las difusividades generales.

Descriptores de redes para vértices

Algunas de las métricas de redes a las que más importancia se les ha dado en estos últimos años están definidas a mayor detalle en [Fornito et al., 2013], [Sporns, 2018] y [Meoded et al., 2020]. Aquí podemos ver las definiciones de aquellas que se extraerán. Para ello, es necesario primero formar el grafo que representa la conectividad del cerebro, como se observa en la Figura 2.14.



Figura 2.14: Ejemplo de vértices y aristas a partir de la matriz de conectividad del valor de FA, donde cada esfera representa una región, localizada en el centroide de dicha región.

- **Centralidad de grado:** Es equivalente al grado del vértice, el cual es la suma de todas las aristas con las que incide. Esta definición de centralidad asume que el vértice con muchas conexiones ejercen más influencia sobre la función de la red y, por lo tanto, tienen una centralidad topológica más alta en comparación con los vértices con menos conexiones.

$$C_D(i) = k_i = \sum_{i \neq j} A_{ij}. \quad (2.6)$$

Donde:

k centralidad de Katz

A matriz de adyacencia

- Centralidad de eigenvector:** Medida sobre la cantidad y calidad de las conexiones de un vértice. Considera tanto el grado de ese vértice como el grado de sus vecinos. Para estimar esta medida, debemos considerar los vectores propios y los valores propios de la matriz de adyacencia A . Un vector propio de A es un vector x distinto de cero que, cuando se multiplica por A , satisface la condición $Ax = \lambda x$. El escalar λ es un valor propio de A y representa un factor que, cuando se multiplica por x , es igual a Ax . La centralidad del vector propio, $C_E(i)$, del vértice i se define como la i -ésima entrada en el vector propio que pertenece al valor propio más grande de A , que se denota como λ_1 . Alternativamente, podemos pensar en la centralidad del vector propio del vértice i como equivalente a la centralidad sumada de sus vecinos. Dado que $Ax = \lambda x$

$$C_E(i) = x_i = \frac{1}{\lambda_1} \sum_{h=1}^N A_{ij} x_j. \quad (2.7)$$

Donde:

x vector distinto de cero

λ_1 valor propio de la matriz A

- Centralidad PageRank** Un vértice es importante si está enlazado desde otros vértices importantes o si está altamente enlazado.

$$C_{PR} = x = (I - \alpha \mathbf{A} \mathbf{D}^{-1})^{-1} \mathbf{1} = \mathbf{D}(\mathbf{D} - \alpha \mathbf{A})^{-1} \mathbf{1}. \quad (2.8)$$

Donde:

α centralidad alpha

\mathbf{D} matriz diagonal del exgrado

\mathbf{A} matriz de adyacencia

Descriptores de la red completa

Así como los otros descriptores muestran valores correspondientes al vértice en cuestión, los siguientes descriptores analizan la totalidad de la red de todos los sujetos a evaluar.

- Coefficiente de asortatividad:** Medida de resiliencia y el coeficiente de correlación para el grado de vértices vecinos; las redes con un coeficiente de asortatividad positivo son resilientes, mientras que las redes con un coeficiente de asortatividad negativo son probablemente vulnerables [Fornito et al., 2013].

- **Small worldness:** En el cerebro existen regiones altamente conectadas entre sí, dentro de sus vecindarios locales y al mismo tiempo estas regiones están conectadas con otras regiones que también comparten una vecindad con otras regiones, por lo que existe una alta eficiencia de transmisión de la información, ya que según la analogía en redes sociales, aunque la distancia entre dos personas sea la máxima en cuanto a geolocalización, ahí se podría conocer a alguien que sea amigo de algún amigo en común de la persona que está lejos, por lo tanto se diría que está a distancia 2 de él, y no a los miles de kilómetros. Esto permite que las redes *small worldness* o mundo pequeño, tengan una organización local eficiente con una integración global efectiva [Bassett and Bullmore, 2006].
- **Coefficiente de agrupamiento:** Es una medida para cuantificar la tendencia de los vértices a buscar formar grupos. Este, indica el grado de agrupamiento local en una red y con ello analizar la estructura de las redes y estudiar su organización. [Fornito et al., 2013]
- **Diámetro:** Es la longitud del camino más largo entre vértices en un grafo. O sea, la máxima distancia entre dos vértices en el grafo. [Johnsonbaugh, 2005]pro

2.3.2. Selección de características con enfoque estadístico

La selección de características se ha vuelto una herramienta importante debido a la masividad de información con la que se cuenta hoy en día. Esta, se encarga de seleccionar y conservar solo las características más importantes del modelo. Las ventajas de implementarla es que simplifica el modelo, lo que conlleva a reducción de datos, menos almacenamiento y mejor visualización, también reduce el tiempo de entrenamiento, evita el sobreajuste y mejora la precisión del modelo, pero sobre todo evita la “maldición de la dimensionalidad”. La selección de características se divide en tres categorías: Filtro(*filter*), *wrapper* y *embedded*, véase en Figura 2.15.

Seleccionar las características principales en pequeñas muestras es una tarea que debe realizarse con precaución [Kuncheva et al., 2020] y esta precaución radica en el par (selector de características, clasificador).

Cuando la muestra es muy pequeña se complica asumir normalidad de las distribuciones o similitudes de varianza, por lo que los procedimientos estadísticos no paramétricos son útiles para hacer inferencias en estos casos. Estos métodos aunque son casi igual de capaces de detectar diferencias entre poblaciones como

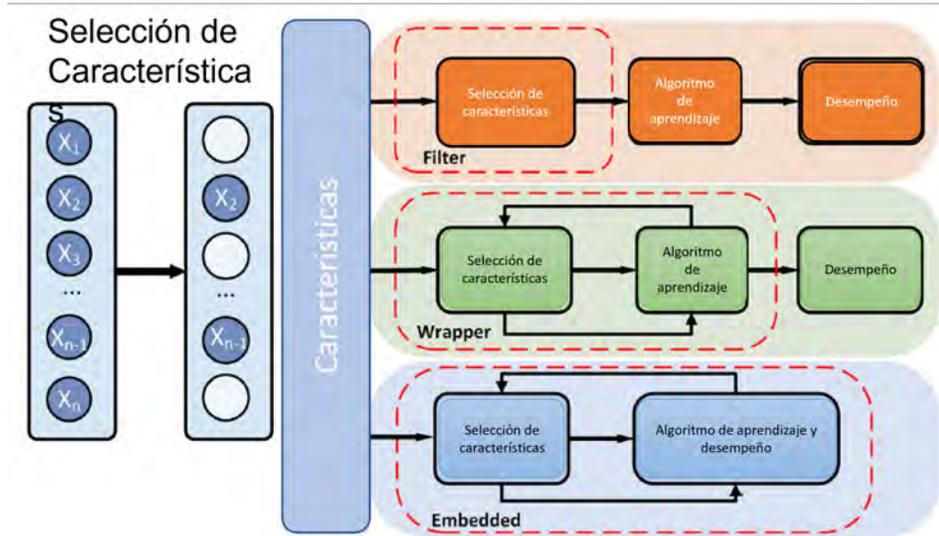


Figura 2.15: Pipeline de los 3 métodos de selección de características. En la figura se muestra el diagrama de funcionamiento para cada tipo de selección de características. En orden de arriba a abajo, tenemos primero el método *filter*, luego el *wrapper* y por último el *embedded*. Fuente de la figura: [Xie et al., 2020] (modificado)

los métodos paramétricos tienen más probabilidades de detectarlas cuando no se satisfacen los supuestos mencionados.

Prueba de Kolmogorov Smirnov

Para comparar las distribuciones de dos muestras con esta prueba se realizan los siguientes pasos:

1. Calcular las funciones de distribución acumuladas (CDFs, por sus siglas en inglés) de cada muestra. Esta función representa la probabilidad acumulada de presentar valores menores o iguales a un valor establecido. Esta se calcula calculando la fracción de valores de la muestra que son menores o iguales a este valor.
2. Se comparan ambos CDF y se calculan las diferencias en cada dato y se registra la diferencia máxima.
3. Se calcula el valor crítico para la estadística de prueba y se le asigna un nivel de significancia α .
4. Se comparan con el valor crítico y si este es mayor se rechaza la hipótesis nula de que ambas pruebas provienen de la misma población. Cuando no lo supera se considera que ambas provienen de la misma población.

Prueba U de Mann Whitney

Para comparar las distribuciones de dos muestras con esta prueba se realizan los siguientes pasos:

1. Se combinan ambas muestras en una sola y se ordenan de menor a mayor posicionándolas del 1 a n , siendo n el tamaño de la muestra.
2. Se calcula la suma de las posiciones de cada muestra. Si llamáramos a las muestras A y B , el conjunto de sus posiciones sería R_A y R_B .
3. Se calcula la prueba estadística U usando las siguientes fórmulas:

$$U_A = n_A * n_B + \frac{(n_A(n_A + 1))}{2} - R_A \quad (2.9)$$

$$U_B = n_A * n_B + \frac{(n_B(n_B + 1))}{2} - R_B \quad (2.10)$$

donde n_A y n_B son los tamaños de las muestras.

4. La hipótesis nula será rechazada si el valor p es inferior al que se defina. El valor p se obtiene basándose en la prueba U con respecto al tamaño de la muestra.

2.4. Clasificación con aprendizaje automático

Es una tarea del aprendizaje automático (ML) supervisado, que busca predecir una etiqueta categórica de un subconjunto de datos que nunca había aprendido de él. Para cumplir esta tarea regularmente se realiza un entrenamiento o ajuste del modelo usando atributos o características que contienen una etiqueta previa y se va optimizando haciendo pruebas con la submuestra nunca antes estudiada. Existen diversos algoritmos que cumplen con el mismo objetivo, por lo que no existe uno superior al otro en general sino en casos específicos tomando en cuenta parámetros como el tamaño de la muestra, cantidad de características o tipo de datos. A continuación se definirán aquellos que se utilizaron para este trabajo, tomando en cuenta que se seleccionaron algunos por su simpleza y otros por la posibilidad de ajustar hiperparámetros que permiten mejorar el modelo evitando el sobre ajuste.

El algoritmo para clasificación realiza los siguientes pasos:

1. Dividir el conjunto total en conjunto de entrenamiento y pruebas.
2. Entrenar algoritmo usando las características y las etiquetas del conjunto de entrenamiento.
3. Realizar una validación con el conjunto de validación y medir desempeño de clasificación.

2.4.1. Naive Bayes (NB)

El algoritmo Naive Bayes es un método basado en el teorema de Bayes, utilizado para clasificar datos. Su enfoque se basa en la suposición de independencia condicional entre las características (variables). Aunque esta suposición puede ser poco realista en algunos casos, Naive Bayes ha demostrado ser efectivo en muchas aplicaciones prácticas, lo que le da origen al nombre “Naive” (ingenuo).

El algoritmo funciona de la siguiente manera: de un conjunto tamaño N de datos etiquetados de entrenamiento, cada dato está representado por un conjunto de características $\mathbf{X} = (X_1, X_2, \dots, X_n)$ y una clase C . La probabilidad a priori $P(C)$, es en la que de una instancia seleccionada de manera aleatoria pertenezca a la clase C . Para cada clase C y cada característica X_i , calculamos la probabilidad condicional $P(X_i|C)$, que es la probabilidad de que la característica X_i tome un determinado valor siendo que la instancia pertenece a la clase C . Por lo tanto, es posible estimar esta probabilidad utilizando el conjunto de entrenamiento.

Dado un dato del conjunto de prueba con características $\mathbf{X} = (X_1, X_2, \dots, X_n)$, se requerirá encontrar la probabilidad a posteriori $P(C|\mathbf{X})$, que es aquella en la que el dato pertenece a la clase C dadas sus características. Aplicando el teorema de Bayes, la probabilidad a posteriori se define como:

$$P(C|\mathbf{X}) = \frac{P(\mathbf{X}|C)P(C)}{P(\mathbf{X})} \quad (2.11)$$

Donde:

$P(\mathbf{X}|C)$ probabilidad conjunta de que todas las características tomen sus valores dados

$P(C)$ probabilidad a priori de la clase C

$P(\mathbf{X})$ probabilidad marginal de las características.

Dado que $P(\mathbf{X})$ es constante para todas las clases, se ignora al comparar las probabilidades a posteriori para simplemente calcular: $P(C|\mathbf{X}) = P(\mathbf{X}|C)P(C)$. Dada

la asunción de independencia condicional entre las características, de puede expresar $P(\mathbf{X}|C)$ como el producto de las probabilidades condicionales de cada característica:

$$P(\mathbf{X}|C) = \prod_{i=1}^n P(\mathbf{X}_i|C) \quad (2.12)$$

Para la clasificación, seleccionamos la clase C que maximiza la probabilidad a posteriori $P(C|\mathbf{X})$, es decir clasificamos el ejemplo como perteneciente a la clase C que tiene el valor máximo de $P(\mathbf{X}|C)P(C)$.

2.4.2. Máquinas de soporte vectorial (SVM)

Es un clasificador discriminativo que busca encontrar el hiperplano separador óptimo que maximiza la separabilidad de los datos de entrenamiento, categorizando así las nuevas muestras. En las ocasiones donde no es posible trazar un hiperplano separador se aplica una transformación añadiendo una dimensión más (*kernels*), facilitando así la clasificación [Timm et al., 2009]. Un hiperplano es un borde de decisión que separa los datos en diferentes clases; en un espacio de dos dimensiones (R^2), el hiperplano es una línea, en R^3 un plano de dos dimensiones, siendo así que en un espacio n -dimensional generalmente el hiperplano será de dimensión $n - 1$.

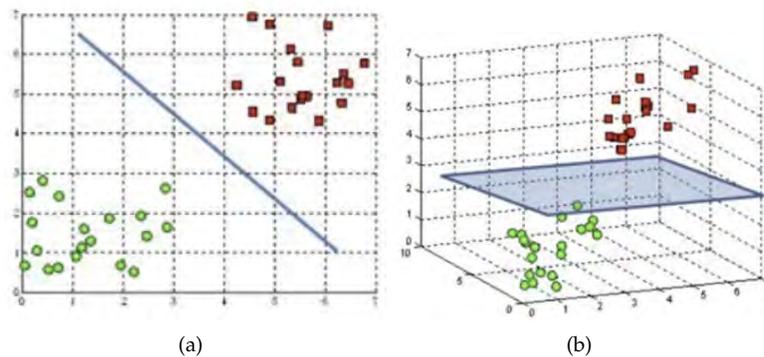


Figura 2.16: Hiperplanos: a) una línea y b) un plano. Se muestra en la figura la separación por medio de los hiperplanos cuando se trabajan con (a) dos dimensiones y (b) tres dimensiones. La dimensión del hiperplano en estos casos es $n - 1$ cuando la dimensión de características es n . Fuente: [Ippolito, 2019]

Lo que diferencia a las SVM es por la forma en que determina el límite de decisión que maximiza la distancia desde los puntos de datos más cercanos de todas las clases. Ese límite de decisión se denomina hiperplano de margen máximo. El margen es la distancia entre el plano y los puntos de datos más cercanos de cada clase. Estos últimos puntos son los llamados “vectores soporte”. En caso de que

nuestros datos no sean linealmente separables, se realiza esta transformación.

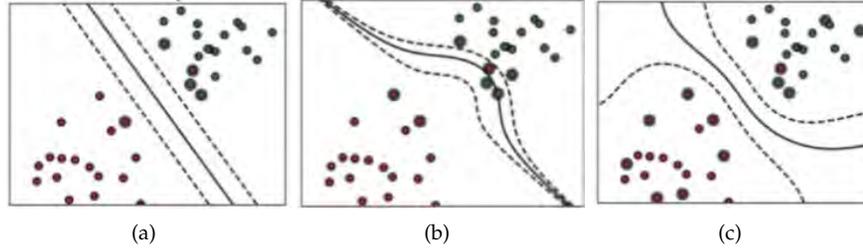


Figura 2.17: Función de transformación (*Kernels*) a) lineal, b) polinomial y c) gaussiana. En las figuras se muestra como este tipo de algoritmos dependen de la función de *kernel*, dependiendo los datos a clasificar. Fuente: [Divakar, 2023]

Si tenemos un conjunto de datos de entrenamiento etiquetados, donde cada muestra está representada por un vector de características \mathbf{X} y una etiqueta de clase y , SVM busca encontrar un hiperplano que maximice el margen entre las clases.

Considerando un conjunto de datos de entrenamiento $\mathbf{D} = \{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \dots, (\mathbf{X}_n, y_n)\}$ donde \mathbf{X}_i es el vector de características de la muestra i y y_i es su etiqueta de clase correspondiente. Para la clasificación binaria, asumimos que las etiquetas son $y_i \in \{-1, +1\}$, [Cortes and Vapnik, 1995].

Dado que SVM se basa en el principio del margen máximo, que busca encontrar el hiperplano que maximice la distancia entre las muestras más cercanas de cada clase. El hiperplano se denota como:

$$\mathbf{w} \cdot \mathbf{x} - b = 0 \quad (2.13)$$

Donde

\mathbf{w} vector de pesos

b sesgo

Las muestras que están más cerca del hiperplano y que influyen en su ubicación se llaman vectores de soporte, y para encontrar el hiperplano óptimo, se utiliza la optimización convexa:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \quad (2.14)$$

sujeto a $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n$

Donde:

\mathbf{w} vector

C parámetro de regularización

ξ variable de holgura

El primer término de la función objetivo es una medida de la complejidad del modelo y busca minimizar la norma del vector de pesos w , el segundo término penaliza las violaciones de las restricciones de separabilidad lineal, donde C es un parámetro de regularización que controla el equilibrio entre el margen y el error de clasificación.

Las restricciones aseguran que las muestras estén calificadas de manera correcta y se encuentren en el lado correcto del hiperplano. Las variables de holgura ξ , sirven para permitir el grado de solapamiento y errores de clasificación.

Para resolver este problema de optimización convexa, se utiliza el método de multiplicadores de Lagrange.

2.4.3. Bosques aleatorios (RF)

Los bosques aleatorios es un algoritmo que consiste en ensamblar un largo número de árboles de decisión donde cada árbol brinda una predicción de clase y la clase con más votos se convierte en la predicción del modelo. Un árbol de decisión es aquél que se representa por vértices de decisión y cada rama representa el resultado de la decisión. Al final las hojas del árbol representan la predicción final. Las decisiones que se toman en cada vértices se basan en una condición en las características, lo que dirige el flujo del algoritmo a la rama apropiada hasta que llegue a un nodo hoja con el resultado [Tony Yiu, 2019].

Los bosques aleatorios crea múltiples árboles de decisión tomando características aleatorias con reemplazo y el ensamble lo realiza al promediar las predicciones de estos árboles para obtener una predicción final o haciendo un conteo de votos, lo cual reduce considerablemente el sobre ajuste.

Como se puede observar en la Figura 2.18 se tienen 9 árboles de decisión, de los cuales seis predijeron la clase 1 y tres la clase 0. Por lo tanto, la predicción final del bosque sería 1, por sistema de votación.

El proceso de construcción comienza con un muestreo aleatorio seleccionando subconjuntos del conjunto de datos de entrenamiento. Posteriormente se construyen los árboles para cada árbol del bosque, se construye un árbol de decisión utilizando la muestra seleccionada anteriormente. En cada nodo, se realiza una división utilizando una selección aleatoria de características. El árbol se construye de manera recursiva hasta alcanzar una condición de parada, como puede ser

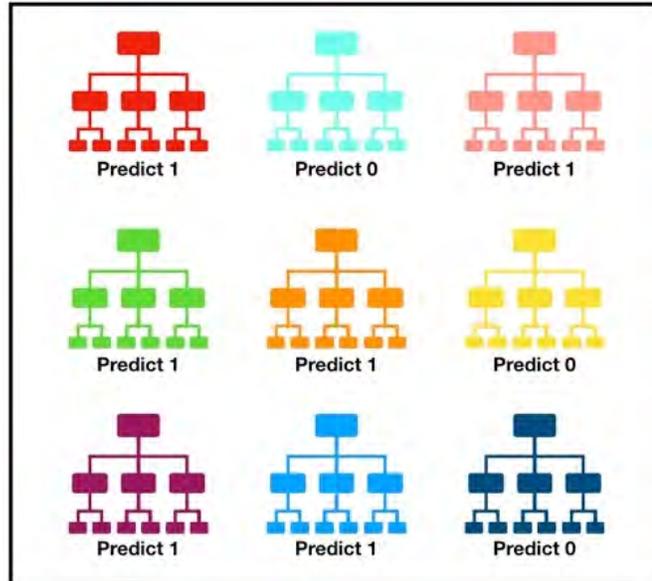


Figura 2.18: Ejemplo del funcionamiento del algoritmo de RF. En la figura se muestra como este método funciona por medio del ensamble de 9 árboles de decisión, donde en conjunto predijeron seis votos a la clase 0 y tres a la clase 1. Fuente: [Tony Yiu, 2019]

alcanzar una profundidad máxima.

Para clasificación, la predicción final se realiza por mayoría de votos, donde cada árbol del bosque vota por una clase, es decir, la moda. Los bosques aleatorios combinan las predicciones de los árboles individuales mediante votación, por lo que si tuviéramos N árboles en el bosque, la predicción para clasificación se puede calcular así:

$$\hat{y} = \text{moda}(f_1(x), f_2(x), \dots, f_N(x)) \quad (2.15)$$

Donde

\hat{y} predicción final

$f_1(x)$ predicción del árbol

2.4.4. Métricas de desempeño en clasificación

La elección de una métrica de clasificación depende de diversos parámetros como el tipo de datos o el problema en específico. Cuando se tienen pocas muestras es importante considerar que pueden haber un tamaño de muestra mayor que el otro, por lo que las métricas de exactitud (accuracy), precisión (*presicion*), sensibilidad (*recall*) y puntuación F1 (*F1-score*) pueden brindar más información.

Para obtener estas métricas es necesario realizar un conteo de: falsos positivos (FP), verdaderos positivos (TP), falsos negativos (FN) y verdaderos negativos (TN). A partir de aquí, podemos obtener la exactitud con la siguiente fórmula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.16)$$

La precisión con la siguiente fórmula:

$$Precision = \frac{TP}{TP + FP} \quad (2.17)$$

Y la sensibilidad por la siguiente:

$$Recall = \frac{TP}{TP + FN} \quad (2.18)$$

A partir de ahí se obtiene la puntuación F1 con la siguiente fórmula:

$$F1 - Score = 2 \frac{Precision * Recall}{Precision + Recall} \quad (2.19)$$

Estas métricas sirven tanto para saber si nuestro clasificador está realizando el desempeño buscado, o para comparar entre 2 o más clasificadores el que realice un mejor desempeño.

CAPÍTULO 3

MARCO DE REFERENCIA

3.1. Desnutrición infantil y deterioro cognitivo

La desnutrición ocurre cuando la ingesta de nutrientes no satisface la necesidad de funciones normales del cuerpo. [Larson-Nath and Goday, 2019]. Se sabe que la mala nutrición del infante está relacionada con factores sociales como la pobreza, depresión materna, bajo coeficiente intelectual materno y el maltrato infantil, tanto la mala nutrición como los factores sociales afectan fuertemente el desarrollo cognitivo, el lenguaje y socioemocional [Anoop et al., 2004].

La desnutrición infantil es un problema de salud pública en algunas regiones de México, y una de sus principales consecuencias es el deterioro en la anatomía y función del Sistema Nervioso Central (SNC). Este deterioro resulta más perjudicial en la fase de crecimiento acelerado desde la trigésima semana de gestación hasta el fin del segundo año de vida. Es en este momento del desarrollo donde podrían generarse importantes modificaciones estructurales del sistema nervioso provocando daño neurológico permanente. Este deterioro anatómico provoca efectos duraderos que impactan el comportamiento y la cognición de los infantes, por ejemplo, ocasiona problemas en el desarrollo normal del control motor y funciones del lenguaje llevando a un bajo rendimiento escolar [Georgieff, 2007]. El impacto negativo de la desnutrición en el SNC puede ser particularmente evidente en el periodo de 6 a 8 años de edad, ya que la línea de crecimiento físico va en aumento e implica adquisición y consolidación de nuevas habilidades cognitivas. Así, se ha encontrado que los procesos cognitivos más gravemente afectados son los mediados por la corteza prefrontal dorsolateral (como atención, fluidez de pensamiento y memoria de trabajo); la corteza parietal derecha (con la función visuoespacial) y

la corteza temporal bilateral (con aprendizaje y memoria verbales, así como comprensión del lenguaje) [Kar et al., 2008].

3.2. Planteamiento del problema

Se conoce que existen diferencias entre pacientes con enfermedades neurodegenerativas y sujetos sanos por medio del estudio de su anatomía cerebral, así como de las conexiones entre regiones de la corteza y subcorteza cerebral. Esta clase de estudios se ha enfocado en enfermedades como el Alzheimer o esquizofrenia, siendo de mayor proporción el estudio en adultos o jóvenes que en niños o bebés. Las conexiones cerebrales se estudian a partir del conectoma, el cual brinda información a partir de imágenes de resonancia magnética estructural, funcional o de difusión, dependiendo el estudio. Dada la poca información que existe sobre el conectoma de cerebro infantil es necesaria la realización de una propuesta de conectoma que pueda ser utilizada como biomarcador de descubrimiento, lo que podría conducir a herramientas clínicamente útiles para la diferenciación entre sujeto pediátricos sanos y aquellos que sufrieron desnutrición. Para la elaboración del conectoma se propone extraer las características morfométricas, de difusión y de conectividad a partir de MRI estructural, identificar las características con mayor aporte estadístico y elaborar un método de visualización que permita identificar el conectoma de un infante sano y diferenciarlo de uno perteneciente a un sujeto desnutrido.

3.3. Objetivo general

Desarrollar un macro conectoma cerebral por medio de técnicas estadísticas y de aprendizaje automático a partir de biomarcadores morfométricos, de difusión y de conectividad extraídos de las imágenes de resonancia magnética T1W y DWI en pacientes pediátricos.

3.4. Objetivos Específicos

- Preprocesar las imágenes y extraer características tanto para las estructurales ponderadas en T1 (T1W) como las ponderadas es difusión (DWI).

- Segmentar las T1W en regiones de interés (ROIs) y obtener características morfométricas.
- Obtener las tractografías por medio de las MRI DTI y sus respectivas características de difusión.
- Realizar registro entre la tractografía y las ROIs.
- Obtener las matrices de conectividad y características de redes.
- Seleccionar distintos conjuntos de características estadísticamente relevantes
- Evaluar el desempeño para clasificar por medio de aprendizaje automático supervisado para cada conjunto seleccionado.
- Generar conectogramas para cada sujeto.
- Análisis comparativo entre conectomas de los grupos control y con desnutrición infantil.

CAPÍTULO 4

MATERIALES Y MÉTODOS

4.1. Descripción de base de datos

Se realizaron estudios en la Facultad de Psicología y la Facultad de Medicina de la UNAM a pacientes pediátricos mexicanos, del estado de Querétaro, entre los 6 y 8 años de edad. Estos pacientes se clasificaron en los grupos “control” y “desnutrición”, siendo 12 y 14 niños, respectivamente, el número de pacientes que se consideraron para este trabajo. Los sujetos del grupo control son pacientes sanos mientras que los del grupo desnutrición padecieron desnutrición sus primeros años de vida.

Tabla 4.1: Información de las MRI

Tipo de MRI	Información
T1W (Secuencia <i>Fast Field Echo</i>)	<ul style="list-style-type: none">▪ TR/TE=8/3.7 ms▪ FOV = $256 \times 256 \text{ mm}^2$▪ Matriz de reconstrucción de 256×256
DTI (Secuencias de disparo único <i>Eco Planar</i>)	<ul style="list-style-type: none">▪ 33 volúmenes de 70 rebanadas axiales (grosor de 2mm sin separación) para cada una de las 32 direcciones independientes de difusión con: $b = 800 \text{ s/mm}^2$ y $b = 0 \text{ s/mm}^2$▪ TR/TE=8467/60 ms▪ FOV $256 \times 256 \text{ mm}^2$▪ Matriz de reconstrucción de 128×128

La base de datos consta de MRIs adquiridas en el Instituto de Neurobiología de la UNAM usando un equipo 3.0 Tesla Achieva MRI (Philips), con el cual se adquirieron imágenes de alta resolución T1-3D y secuencias de DWIs. Información sobre la adquisición en tabla 4.1 y visualización de las MRIs se observan en la Figura 4.1.

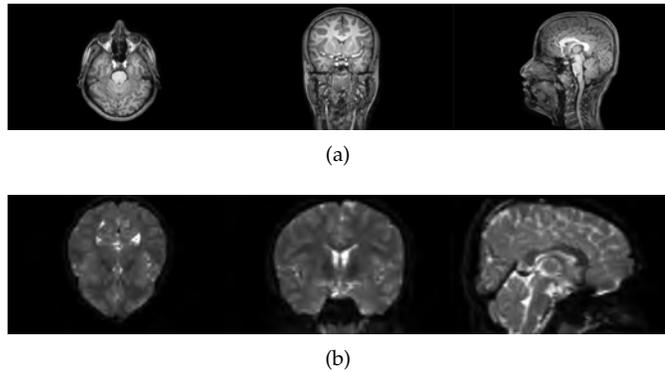


Figura 4.1: Vista axial, coronal y sagital de MRIs de sujeto control: (a) T1W y (b) DWI

4.2. Metodología

A continuación, se muestra un diagrama general de los pasos que se siguieron hasta obtener la visualización de la propuesta de conectoma. En términos generales está dividido en 5 procesos: el primero es preprocesar las imágenes tanto estructurales como de difusión, seguido del procesamiento de las mismas que consiste en la segmentación en regiones corticales y subcorticales y la obtención de la tractografía, posteriormente la extracción de características de forma, de difusión y de redes, luego la selección de características con mayor importancia estadística y por último la visualización del conectoma. Cada método lleva sus procedimientos, como se observa en el diagrama de la Figura 4.2.

4.2.1. Preprocesamiento de las MRI

Preprocesar las MRIs sirve para preparar los datos de las imágenes para el análisis buscando minimizar errores causados por artefactos, una errónea alineación o ruido. Para este trabajo se utilizaron dos modalidades de MRI, la estructural y la de difusión, y aunque existen similitudes en algunos de los pasos del preprocesamiento, en general será diferente dado que estos se realizan especialmente a una modalidad de imagen y se buscan extraer datos distintos en ambos. Para la

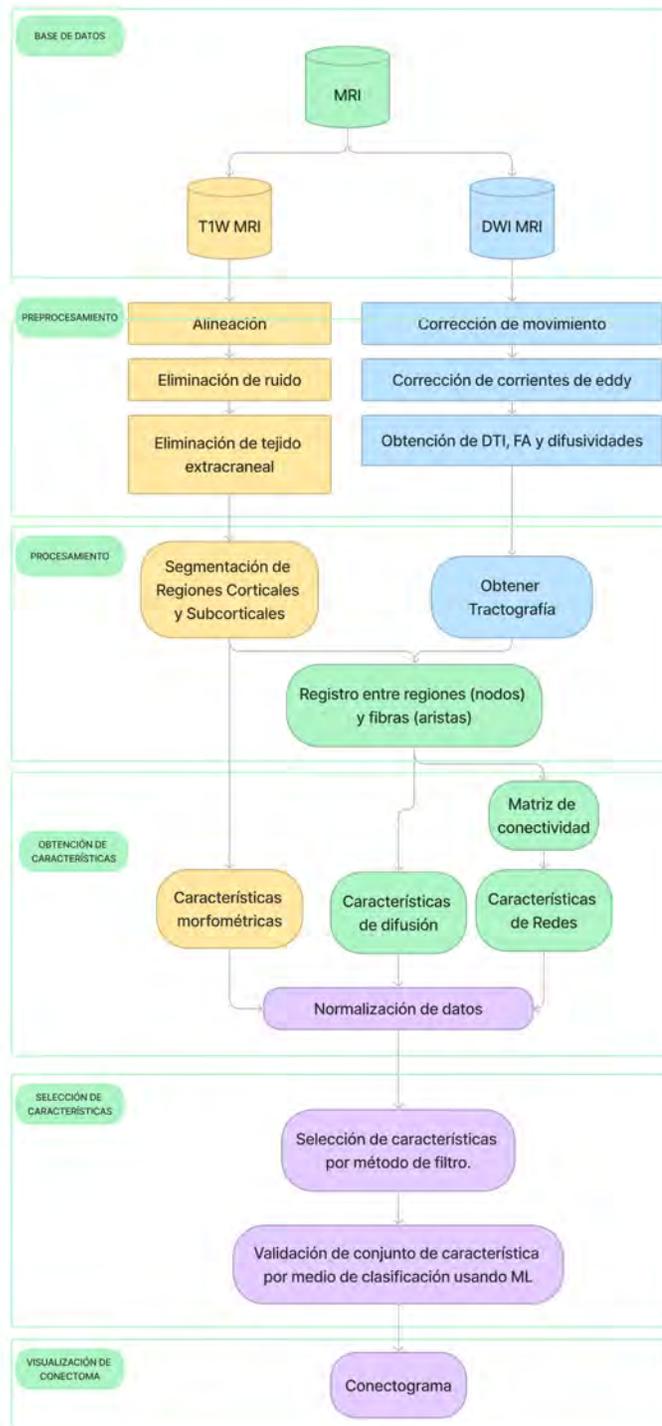


Figura 4.2: Diagrama de la metodología. Comenzando por los materiales o base de datos, seguido de un preprocesamiento en paralelo entre las dos modalidades, el procesamiento consta también de un registro entre las regiones y la tractografía para obtener todos los tipos de métricas, se realiza una selección de las mejores características por medio de aprendizaje automático y con ellas se seleccionan las regiones con mayor significancia para el análisis del conectoma por medio del conectograma.

identificación de los procedimientos únicos para cada modalidad se colocará entre paréntesis cuando sea exclusivamente para ella, para el preprocesamiento y segmentación de las imágenes T1W se utilizó *FreeSurfer* [Fischl et al., 2002] y para las imágenes DWI se utilizó *DSI studio* [Maier-Hein et al., 2017].

Alineación rígida

Cuando se desea analizar MRIs en grupos con tamaños y edades variadas, es útil hacer una alineación que no cambie la forma interna de los tejidos pero que sí existan similitudes en las dimensiones generales de sus imágenes. La alineación rígida o registro de cuerpo rígido consiste en realizar únicamente transformaciones como traslación, rotación y escalamiento para preservar las estructuras y exclusivamente cambiar el tamaño y orientaciones de la imagen.

Cada tipo de alineación consiste en alinear una o más imágenes con una imagen de referencia (atlas, regularmente), encontrando aquellos movimientos que alinee las imágenes en la orientación y posición de la de referencia. A continuación, se presentan las representaciones para cada tipo de alineación:

- **Traslación:** Matemáticamente es representado mediante la matriz de transformación tridimensional \mathbf{T} , que contiene los valores de traslación en las dimensiones x , y y z , escrita de la siguiente forma:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.1)$$

- **Rotación:** Es representado por una matriz de rotación tridimensional \mathbf{R} , que contiene los valores de movimiento en los ejes x , y y z , escrita de la siguiente forma:

$$\mathbf{R} = \mathbf{R}_z(\theta_z)\mathbf{R}_y(\theta_y)\mathbf{R}_x(\theta_x) \quad (4.2)$$

Donde:

\mathbf{R} rotación en los ejes

θ ángulos de rotación al rededor de los ejes correspondientes.

Las matrices de rotación son escritas de la siguiente forma:

$$\mathbf{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{bmatrix} \quad (4.3)$$

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \quad (4.4)$$

$$\mathbf{R}_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.5)$$

- **Escalamiento:** Es cambiar el tamaño de la imagen en una o más dimensiones, el cual es representado en una matriz, escrita de la siguiente forma:

$$\mathbf{S} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \quad (4.6)$$

Eliminación de ruido (T1W)

Las MRIs suelen adquirir ruido como el gaussiano o riciano, sobretodo al momento de la reconstrucción. La eliminación de ruido debe realizarse antes de comenzar otros procesos donde este mismo pueda alterar resultados en los métodos siguientes. La eliminación de ruido por medio del filtro adaptativo de medios no locales (ANLM, por sus siglas en inglés) restaurando cada píxel de la imagen calculando un promedio ponderado de los píxeles circundantes mediante una medida de similitud que tiene en cuenta los píxeles vecinos que rodean al píxel que se compara [Manjón et al., 2010].

El algoritmo ANLM funciona de la siguiente forma [Manjón et al., 2010]:

1. Se inicializa una función de ponderación estimando el ruido local obtenido por la desviación estándar de las intensidades es un reducido vecindario de píxeles alrededor de cada pixel.

2. Se calcula el promedio ponderado de pixeles similares en la imagen, el cuál se calcula en función de la similitud de sus intensidades y función de ponderación.
3. Se actualiza la función de ponderación basandose en la imagen sin ruido, tomando en cuenta el nivel de ruido local de cada pixel.
4. Repetir paso 2 y 3, hasta converger.

En la Figura 4.3a se puede observar una MRI con el ruido llamado riciano, y una vez aplicado el algoritmo ANLM se obtiene como resultado una imagen como la Figura 4.3b.

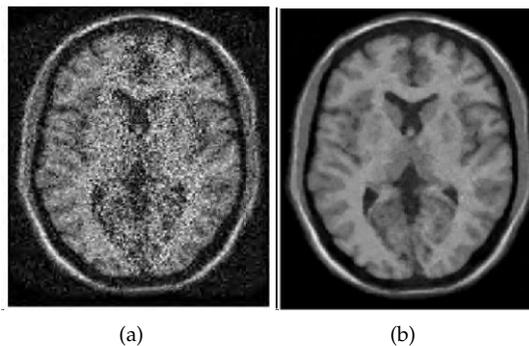


Figura 4.3: MRI T1W vista axial (a) con ruido riciano y (b) filtrada con ANLM
Fuente: [Manjón et al., 2010] (Modificado)

Extracción de tejido extracerebral (T1W)

A través del software FreeSurfer [Fischl et al., 2004] se realizaron la extracción de tejido extracerebral. Esta función remueve los tejidos que no pertenecen a la WM, GM o CSF, utilizando un conjunto de técnicas basadas en umbrales, operaciones morfológicas y deformaciones de superficie. Los pasos que se realizaron son los siguientes:

1. Mediante un modelo de forma activa tridimensional (ASM, por sus siglas en inglés) se realiza una primer extracción de tejido inicial. El ASM está pre entrenado con un conjunto de datos de MRI para encontrar la ubicación aproximada del tejido cerebral y genera una máscara cerebral inicial que se va refinando con un algoritmo de segmentación que se basa en las intensidades.
2. Se aplica el algoritmo de segmentación primero normalizando las intensidades para que los valores en las MRIs sean consistentes en toda la imagen,

posteriormente se genera un umbral para que los valores de intensidades que no pertenezcan a los valores normales sean eliminados.

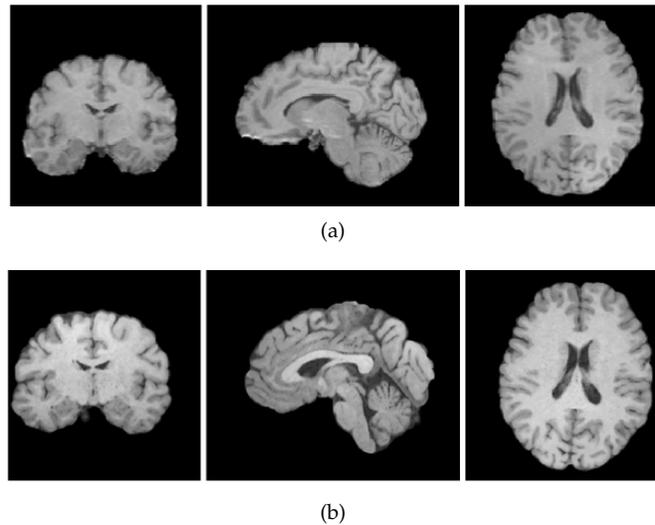


Figura 4.4: MRIs (a) con escala de grises original y (b) con escala de grises normalizada.
Fuente: [Sun et al., 2015]

3. Se realiza un análisis de componentes conexas, para evaluar cuál grupo de vóxeles que conectados están en un espacio tridimensional y están considerados para ser parte del mismo objeto. A partir de estos resultados se eliminan vóxeles solitarios y artefactos que pudieron haber sido introducidos este proceso.
4. Se realiza una reconstrucción de la superficie cerebral usando la técnica de inflación esférica, la cual genera una malla tridimensional que cubre el cerebro completo, con vértices y aristas que representan la forma de la superficie cerebral, véase Figura 4.5a. La técnica de inflación esférica comienza en el centro del cerebro como punto de partida, seguido de la expansión de la superficie cerebral de una forma esférica hasta llegar a tejido no cerebral. Una vez expandida la superficie cerebral se realiza un método de interpolación para estimar las intensidades de los vóxeles dentro y se comparan con los de la imagen original, eliminando aquellos vóxeles que no coincidan.
5. Se comparan las intensidades de cada vértice en la superficie del cerebro y aquellos con diferencia significativa a sus vecinos es considerado tejido no deseado y es eliminado, véase el resultado en Figura 4.6b.

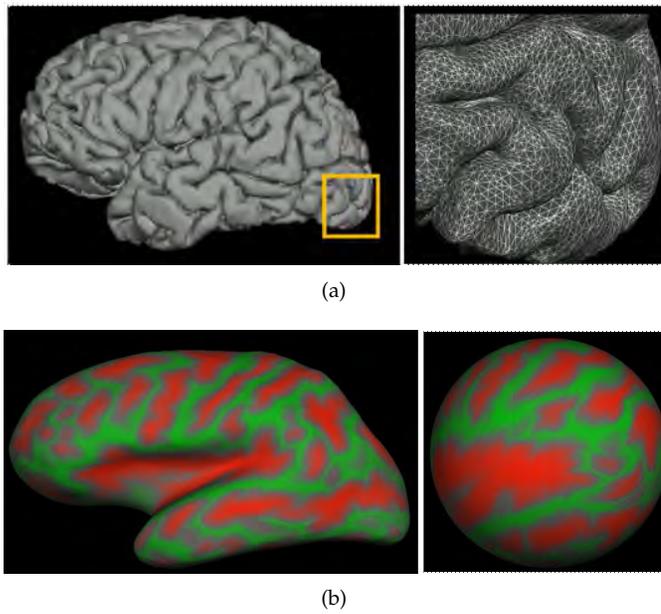


Figura 4.5: Representación de la malla que representa la superficie cerebral y el mapa de surcos con su inflación esférica. La Figura (a) presenta la malla que mediante vértices y aristas representa la superficie. La Figura (b) presenta el mapa de surcos y su inflación esférica. Fuente: [Andy Brain Blog, 2018]

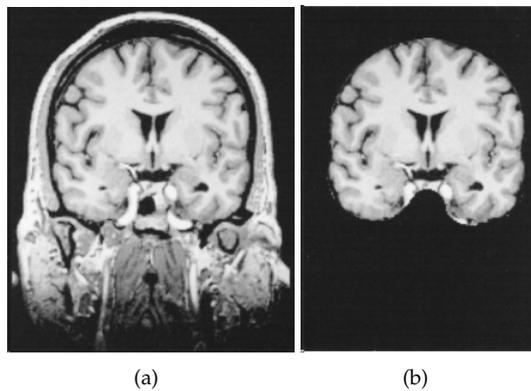


Figura 4.6: MRIs (a) con tejido extracerebral y (b) MRI con únicamente tejido cerebral. Fuente: [Dale et al., 1999]

Corrección de movimiento (DTI)

La importancia de este procedimiento se debe a que busca reducir los efectos causados por los movimientos de la cabeza al momento de la adquisición, dado que estos pueden resultar en una imagen con artefactos. Este paso busca alinear las imágenes DTI parte de una imagen anatómica y luego se realinean las imágenes. Para hacer una corrección de movimiento se deben realizar los siguientes pasos:

- **Coregistro de DTI con T1W:** Para estimar el movimiento de la cabeza durante la adquisición de la DTI.
- **Estimación de movimiento de cabeza:** Estima el movimiento calculando la matriz transformación que describe el movimiento.
- **Corrección de movimiento de cabeza:** La matriz de transformación es aplicada a los datos DTI.

Corrección de corrientes eddy (DTI)

Esta modalidad de imágenes médicas por lo regular tienen distorsiones causadas por la sensibilidad a inhomogeneidades del campo magnético [Sotiropoulos et al., 2013]. En las DTI es necesario corregir la distorsión ocasionada por las corrientes eddy y artefactos por movimiento de la cabeza [Tax et al., 2022]. Las corrientes eddy, también conocidas como corrientes de Foucault, son el resultado de campos magnéticos de gradiente que cambian rápidamente y que inducen corrientes parásitas en los materiales conductores circundantes, formándose de acuerdo con la ley de inducción de Faraday. Las corrientes eddy no son deseadas ya que generan sus propios campos magnéticos, que se oponen al campo magnético original a través de la Ley de Lenz y, por lo tanto, distorsionan el rendimiento espacial y temporal del campo magnético general deseado [Maniega et al., 2007]. Se sabe que la resonancia magnética de difusión tiene una gran distorsión de corrientes eddy y distorsión de fase. Para solucionar esto se debe corregir la distorsión parcial mediante una transformación lineal.

Estimación de tensor de difusión (DTI)

El cálculo del tensor de difusión es obtenido a partir de la siguiente ecuación lineal que describe la difusión del agua en cada dirección.

$$\frac{\partial P(r, t)}{\partial t} = D \nabla^2 P(r, t). \quad (4.7)$$

Donde:

$P(r, t)$ *concentración de las moléculas de agua*

r *posición*

t *tiempo*

D *coeficiente de difusión*

∇^2 *operador Laplaciano*

La ecuación describe el movimiento de las moléculas de agua debido a su movimiento térmico aleatorio. La definición del tensor de difusión en sí es una matriz definida simétrica que describe la difusión en cada dirección y la utilidad de este es realizar el seguimiento de fibra o tractografía para cada vóxel. Para esto se debe cuantificar la cantidad de ponderación de difusión aplicada a la MRI, en la que se utilizan los b -valores (coeficiente de ponderación de difusión), y este se mide en s/mm^2 .

4.2.2. Procesamiento de las MRI

Dado que encontrar los descriptores morfométricos, de difusión y de red de cada región, primero es necesario segmentar el tejido cerebral en ROIs y posteriormente encontrar sus descriptores ya sea analizando solo las regiones como también las regiones registradas con la tractografía y así extraer las características correspondientes.

Segmentación de T1W

Segmentar y etiquetar la anatomía macroscópica del cerebro humano es fundamental al realizarse un “mapeo cerebral”, con el objetivo de localizar regiones para su identificación, comparación y análisis. Una vez subdivido y etiquetado es posible cuantificar datos dentro de las regiones [Klein and Tourville, 2012].

La segmentación es basada en un atlas probabilista independiente y valores medidos específicos del sujeto. El atlas se construye a partir de un conjunto de entrenamiento, es decir, un conjunto de sujetos cuyos cerebros (superficies o volúmenes) que se han etiquetado a mano. Estas etiquetas se mapean en un espacio común para lograr una correspondencia punto a punto para todos los sujetos. Un “punto” es un vóxel en el volumen o un vértice en la superficie. En cada punto del espacio, existe la etiqueta que se asignó a cada sujeto y el valor medido para cada



Figura 4.7: Regiones de interés de la segmentación por el método de Desikan-Kalliany-Tourville. Fuente: [Klein and Tourville, 2012]

sujeto. Luego se calculan las probabilidades de que el punto pertenezca a cada una de las clases de etiquetas, de que un punto dado pertenezca a una etiqueta dada la clasificación de sus puntos vecinos y la función de distribución de probabilidad del valor medido se estima por separado para cada etiqueta en cada punto [Fischl et al., 2002].

En la segmentación de la corteza, mediante un enfoque basado en la curvatura de resolución múltiple, se realiza un registro de un modelo de superficie cerebral del sujeto en un modelo de superficie de atlas delineado; las ROIs del atlas y las curvas de los surcos se asignan a la superficie del cerebro del sujeto. A continuación, se utiliza un flujo de curvatura geodésica en la superficie cortical para refinar las ubicaciones de las curvas de los surcos y etiquetar los límites aún más, de modo que sigan el verdadero fondo de los sulcales. El flujo se formula utilizando un método basado en conjuntos de nivel en la superficie cortical, que representa las curvas como conjuntos de nivel cero. La ponderación es basado en curvatura y lleva las curvas a la parte inferior de los valles de los surcos en los pliegues corticales [Joshi et al., 2012].

Tractografía de MRI-DTI

El seguimiento de fibra de MRI DTI utiliza direcciones de fibra locales medidas en cada vóxel para rastrear la trayectoria de una vía de materia blanca. Las direcciones axonales permiten seguir las trayectorias de los haces de fibras. Los métodos de seguimiento se pueden clasificar en seguimiento de fibra determinista

y seguimiento de fibra probabilístico [Basser et al., 2000]. Para encontrar los tractos se realizará el enfoque determinista para el seguimiento de las fibras propuesto por el Dr. Basser, el cual delinea la trayectoria de la fibra empezando en un punto semilla en la materia blanca y va siguiendo las direcciones de las fibras locales en un proceso recursivo hasta que se cumplen los criterios de terminación. Los criterios pueden incluir un umbral angular, un umbral de anisotropía o un antecedente anatómico. El cálculo no involucra ningún parámetro aleatorio y, por lo tanto, cada vez que se realice un seguimiento de fibra desde el mismo punto semilla, la trayectoria resultará ser la misma. Una visualización de lo que se obtiene se muestra en la Figura 4.8, en las vistas axial, sagital y coronal, respectivamente.

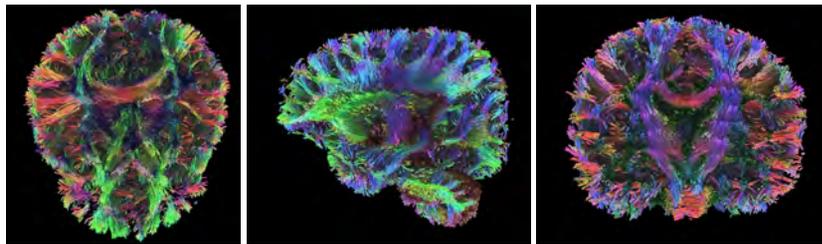


Figura 4.8: Vista axial, sagital y coronal de la tractografía.

Registro de DTI con cerebro de T1W

El análisis de la tractografía puede realizarse únicamente con los tractos, sin embargo existen otros enfoques como lo es el basado en regiones. Para ello es necesario realizarse un registro entre los segmentos y las fibras de la tractografía, por lo que será necesario realizar un registro entre las imágenes de referencia MRI T1W y la MRI DTI del cerebro de cada sujeto. Una vez registrada la MRI DTI en la MRI T1W, se calcula la tractografía, con la cuál se puede realizar al análisis de difusión por región, e interpretando las regiones como vértices de un grafo y los tractos (filtrados por las regiones en las que empiezan y terminan) como aristas, se procede a un análisis de conectividad usando la matriz de adyacencia ponderada (según el número de tractos) y no ponderada (umbralizada).

4.2.3. Obtención de características

La obtención de características conlleva usar la información extraída de la T1W y las DWI, como lo son las regiones cerebrales y la tractografía. Con las regiones se obtendrán las características morfométricas, mientras que, a partir del registro en-

tre la tractografía y las regiones (a partir del registro de T1W y DWI), se obtuvieron las características de difusión y de redes.

Morfométricas

Para obtener los descriptores morfométricos de cada región se utilizaron los valores de los vóxeles como el número de vóxeles, se compararon las distancias vóxel a vóxel para obtener los valores propios, los cuales nos brindaban información de elongación, redondez y aplanamiento. También a partir del conteo de las caras y aristas para brindarnos información como la compacidad o tortuosidad. Más detalles, véase marco teórico, capítulo 2.3.1.

De difusión

A partir de la DTI, se utilizan los valores propios del tensor para obtener tanto el valor de FA como el de las difusividades. Dado que se realizó un registro entre los tractos y las regiones, se caracteriza cada región con sus valores estadísticos como el promedio, el mínimo y el máximo. Lo que nos da una idea de los valores que se manejan en esa región. Más detalles, véase marco teórico, capítulo 2.3.2.

De redes

Para obtener estas características fue necesaria una matriz de conectividad cuyos pesos en las intersecciones se obtenía con información del número de tractos normalizados (aristas) que empezaban en una región (vértice) y terminaban en otra región. Se descartaron todos los tractos menores a 20 mm, ya que pequeños errores en el registro entre las regiones y tractos podrían indicar que una región tiene una conectividad mayor con sus regiones vecinas de lo que en realidad podría tener. Más detalles, véase marco teórico, capítulo 2.3.3.

Una vez obtenidas las características, es necesario normalizar el conjunto de datos completo para cada característica. Lo que se hace en este proceso es ordenar en un formato tabular todas las características de cada sujeto, por lo que se tendrían 16 filas (número de sujetos) y cada característica individual conformará una sola columna. Cada columna del conjunto de datos se normalizará con el método mínimo y máximo, siendo el valor mínimo de dicha columna el equivalente al 0 y el máximo el equivalente a 1. Por lo que todas las columnas tendrán valores entre 0 y 1, únicamente.

4.2.4. Selección de características

La selección de características consiste en seleccionar un subconjunto de ellas que sean relevantes para el estudio y no redundantes. Es una práctica que se realiza en tareas de aprendizaje automático cuyas ventajas son la mejora del desempeño de un modelo, la reducción de la complejidad y la reducción del costo computacional.

De los tres métodos de selección de características se utilizó el método de filtro (*filter method*), que consta de pruebas estadísticas que evalúan la importancia de cada característica y descarta aquellas estadísticamente menos relevantes.

Dado que el número de sujetos para el estudio es reducido y existe una alta variabilidad en los datos, se utilizaron pruebas no paramétricas para la selección de características. Se realizó el método de filtro usando la prueba de Kolmogorov Smirnov (KS) y la prueba U de Mann Whitney (MW). Ambas son usadas para comparar la distribución de dos muestras e identificar si las muestras pertenecen o no a la misma población.

Para armar el conjunto de todas las características se consideró como característica única a la unión del biomarcador con la región. Es decir, si se extrajeron m descriptores o biomarcadores de n regiones, entonces de cada sujeto se tienen mn características. Una vez teniéndolas se separó un conjunto para selección de características y otro de validación. La validación de los subconjuntos de características se hará por medio del desempeño en la tarea de clasificación utilizando las características seleccionadas por KS, las seleccionadas usando MW y del subconjunto que comprenden únicamente regiones subcorticales. Dado que los 26 sujetos totales, siendo 12 los de Control y 14 los de Desnutrición, se tomaran 20 sujetos al azar para el conjunto de entrenamiento y los 6 restantes formarán el conjunto de prueba, tomando aquellas características cuyo valor p fuera menor a 0.05. seleccionadas todas las veces.

Para la creación de estos subconjuntos se consideraron los siguientes criterios:

- Todos los seleccionados con la prueba de MW
- Todos los seleccionados con la prueba de KS
- Aquellas características resultantes de la unión entre los dos subconjuntos anteriores
- Todos aquellas características de regiones corticales de aquellas resultantes de la unión

- Todos aquellas características de regiones subcorticales de aquellas resultantes de la unión

4.2.5. Clasificación y validación de conjunto de características

Una vez teniendo el subconjunto de características a considerar se implementaron algoritmos clasificadores cuyo desempeño sirve de indicador acerca de sí el subconjunto de características seleccionado es útil para el estudio.

La selección de algoritmos cuando se tienen pequeñas muestras y más características que ellas, depende de diversos factores como la naturaleza de los datos o el balanceo entre clases. Se optó por usar un algoritmo simple como lo es Naive Bayes (NB), así como también otros dos con hiperparámetros que configurar para evitar el sobre ajuste, un problema muy común cuando se trabaja con muestras pequeñas, los cuales fueron máquinas de soporte vectorial (SVM) y bosques aleatorios (RF).

El preprocesado de los datos para nuestros clasificadores será el siguiente:

- Normalizar el conjunto de entrenamiento de 20 sujetos usando minmax, siendo 0 el mínimo y 1 el máximo.
- Usando los valores mínimos y máximos de cada característica, normalizar el conjunto de validación de 6 sujetos.

Una vez preprocesados los datos, se implementaron los algoritmos NB, SVM y RF en los subconjuntos de características. Esto para poder medir su desempeño por medio de métricas de clasificación como lo son la exactitud, precisión, sensibilidad y puntuación F1 aquel conjunto de características que logra clasificar mejor y tener una mayor certeza sobre la elección de las mismas. Dado que dependiendo de los sujetos de entrenamiento, serán las características que se seleccionarán, se considerarán como características seleccionadas aquellas que hayan sido significantes a menos de 0.05 de valor p en todas las iteraciones.

4.2.6. Construcción del conectoma

De las características seleccionadas, se tiene la particularidad de ser la unión de un descriptor con una región del cerebro. Para la construcción del conectoma son de interés las regiones, las cuales serán nuestros vértices, así como su conexión por medio de la tractografía para la obtención de la matriz de conectividad y generar

así el conectograma. En la Figura 4.9 se puede observar la reducción de vértices para el estudio, dejando únicamente 5 regiones.

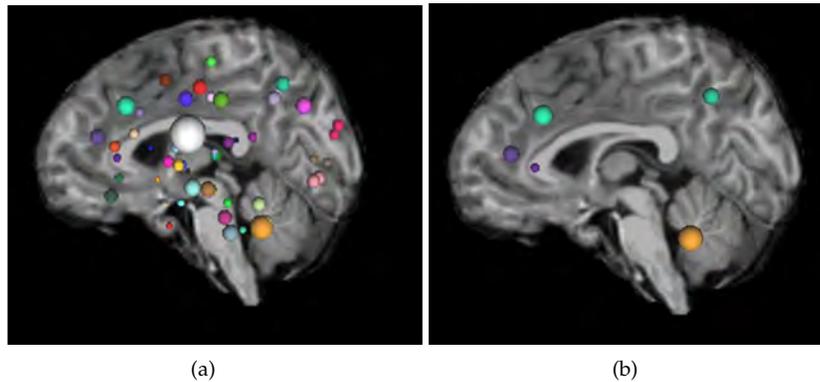


Figura 4.9: MRI vista sagital con (a) todas las regiones (vértices) y (b) con las regiones rostrales de azul turquesa, superior frontal y parietal de morado y el cerebelo de naranja, todas del hemisferio izquierdo.

Para obtener el grafo, necesitamos saber si existe o no conexión entre regiones, utilizando la FA es que podemos obtener la matriz de conexión entre dichas regiones con sus respectivos coeficientes de correlación, véase la matriz en tabla 4.2. De esta misma matriz es que se obtiene el grafo que se puede ver de manera espacial, como en la Figura 4.10a, así como también visto por sus valores de conexión, para lo cual se usa la técnica del conectograma, véase Figura 4.10b.

Tabla 4.2: Matriz de conectividad de las 5 regiones seleccionadas.

REGIONES	L-CC	ctx-lh-rac	ctx-lh-rmf	ctx-lh-sf	ctx-lh-sp
L-CC	0	0	0.392925	0.423129	0.391147
ctx-lh-rac	0	0	0	0.250292	0.392652
ctx-lh-rmf	0.392925	0	0	0.287393	0.402162
ctx-lh-sf	0.423129	0.250292	0.287393	0	0.436146
ctx-lh-sp	0.391147	0.392652	0.402162	0.436146	0

L-CC: Corteza cerebral rac: rostral cíngulo anterior rmf: Rostral medio frontal sf: Superior frontal sp: Superior parietal.

Como se observa en la matriz presentada en la tabla 4.2, los valores rondan aproximadamente entre el 0.2 y el 0.4, por lo que mediante la técnica de umbralización es posible filtrar algunas conexiones para únicamente tener aquellas con mayor coeficiente. Esta técnica también puede servir para obtener una matriz binaria y por ende un grafo binario. Algunas de estas técnicas de umbralización es eliminar todas aquellas por debajo cierto cuartil o percentil.

En la Figura 4.10a se observa la vista sagital con los vértices de tipo esfera tri-dimensional representando las 5 regiones seleccionadas para visualizarse, y en la

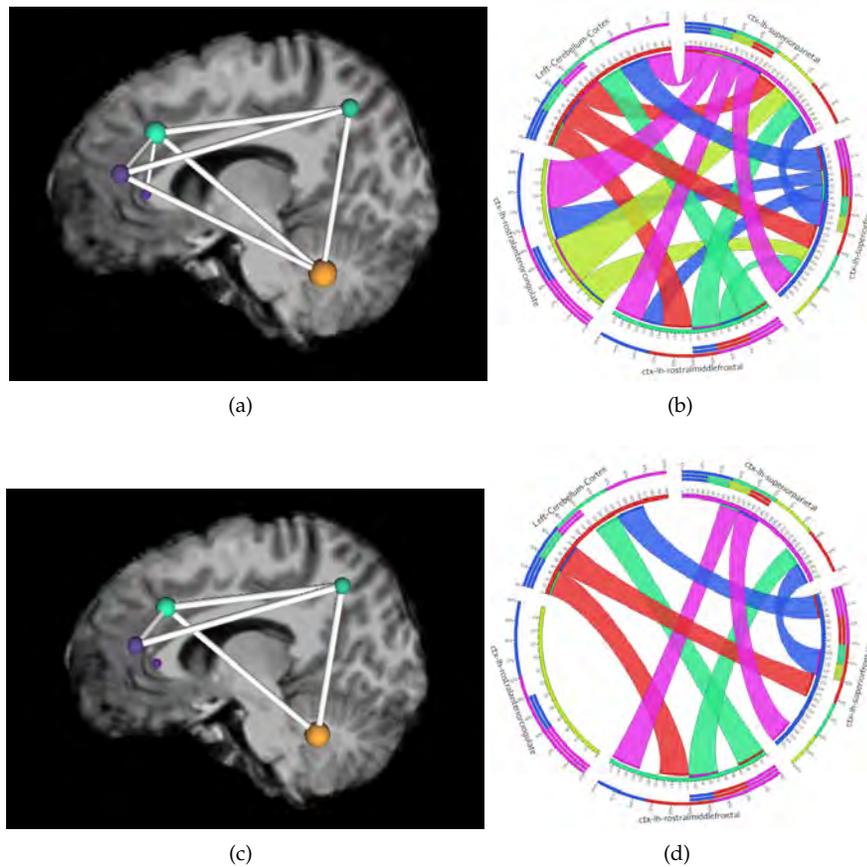


Figura 4.10: Representación espacial y conectogramas del grafo de ejemplo con las regiones seleccionadas. El grafo espacial mostrado en (a) y conectograma mostrado en (b), corresponden a la matriz vista en la tabla 4.2, y el grafo (c) y el conectograma (d), son del resultado de filtrarla hasta el cuartil Q2.

Figura 4.10b se tiene el conectograma de las 5 regiones seleccionadas, destacando la conectividad la corteza izquierda superior parietal, la región con más y mayor peso en sus conexiones y el rostral cíngulo anterior izquierdo el que tuvo menos y menor peso en sus conexiones.

El objetivo de únicamente ver aquellas conexiones con alto valor de conectividad es, además de filtrar, poder observar con mayor facilidad la conectividad de cada vértice, evitando saturarse de colores y tender a la confusión cuando se trata de observar todas las conexiones. Como se observa en la figura 4.10 b y d, existen diversos colores que conectan una región con otra, esto significa que cada región porta su color característico, por lo que será muy común encontrar dos colores de aristas uniendo los mismos dos vértices, esto debido a que provienen de una matriz de adyacencia simétrica.

CAPÍTULO 5

RESULTADOS Y DISCUSIÓN

En este capítulo se irán presentando los resultados junto con la discusiones del mismo. Los cuatro métodos generales que se llevaron a cabo y que se mostrarán a continuación, son:

- Preprocesamiento de las imágenes
- Procesamiento de las imágenes
- Extracción de características
- Selección de características
- Visualización del conectoma

5.1. Preprocesamiento de MRI

El método de preprocesamiento se dividió en 2 secciones, las que se realizan a las T1W y las que se aplican a DTI, dado que estos se realizaron en paralelo y en este mismo orden se presentarán.

Preprocesamiento de T1W

Para este tipo de imágenes se buscó que el preprocesamiento no alterara la forma de las ROIs, por lo que se optó por hacer alineación rígida que únicamente hace traslaciones, rotaciones y escalamientos, esto permite cambiar el tamaño y orientaciones sin cambiar la forma o posiciones internas de las estructuras. Para la eliminación de ruido se optó por un filtro ANLM y para la eliminación de tejido extracraneal se realizó por medio de la técnica de inflación esférica.

Alineación rígida

Como se observa en la Figura 5.1, existen diferencias de tamaño y alineación entre las distintas MRIs de cada paciente, por lo que se aplicó una alineación rígida con el atlas pediátrico construido a partir de MRIs desde 4.5 a 8.5 años [Fonov et al., 2011], como se muestra en la Figura 5.2. En esta última Figura, el MRI correspondiente a un paciente del grupo de estudio se alinea con el atlas al que todas las demás MRIs se alinearán. Y en la Figura 5.2c, se aprecia el resultado de esa alineación.

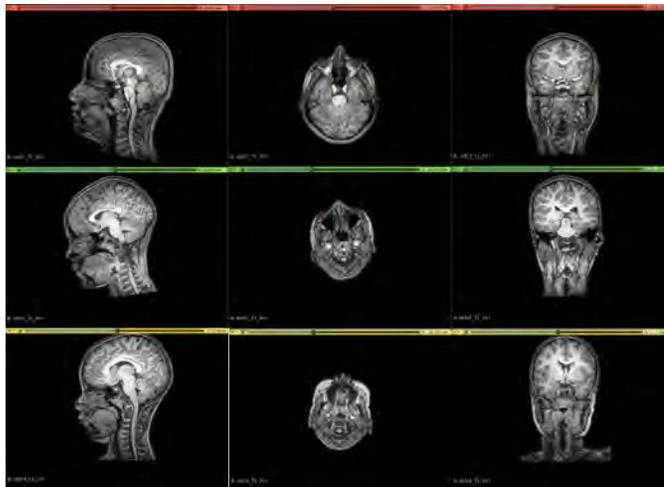
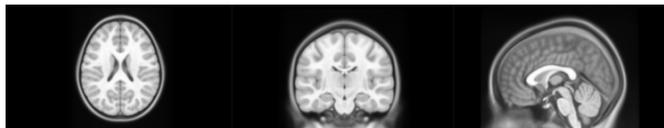


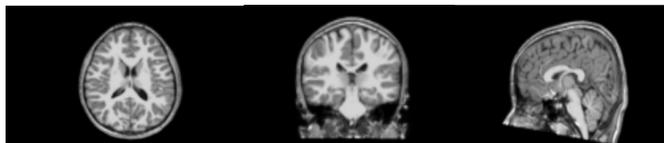
Figura 5.1: MRIs sin preprocesamiento de tres sujetos *control*.



(a)



(b)



(c)

Figura 5.2: Proceso entre el (a) MRI original al aplicarle alineación rígida por el (b) MRI del atlas para obtener (c) una MRI alineada.

Eliminación de ruido

Las MRI suelen sufrir ruidos como el ruido gaussiano o “riciano”, sobretodo al momento de la reconstrucción. Se utilizó un filtro de medias adaptativas no locales (ANLM), el cual restaura cada píxel de la imagen calculando un promedio ponderado de los píxeles circundantes mediante una medida de similitud que tiene en cuenta los píxeles vecinos que rodean al píxel que se compara [Manjón et al., 2010].

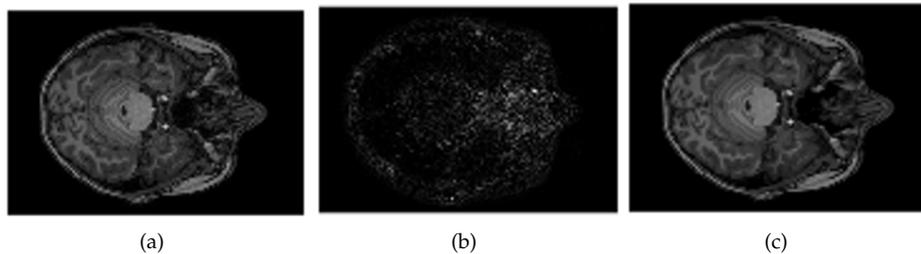


Figura 5.3: Eliminación de ruido en MRI. (a) MRI original, (b) ruido “riciano” removido e (c) imagen filtrada.

Eliminación de tejido extra-cerebral

La extracción cerebral, permite que usemos únicamente los pixeles correspondientes al tejido cerebral para la extracción de características y así, al momento de segmentar o realizar un registro, este se realice con el mejor desempeño posible, ya que se disminuyen la cantidad de artefactos que podrían perjudicar el análisis, véase el resultado del MRI de un paciente en la Figura 5.4b, como se observa, se eliminaron los pixeles que no corresponden al tejido cerebral.

Preprocesamiento de DWI

El preprocesamiento que se realizó a las DWIs fue corregir el movimiento y corrientes de Eddy, para posteriormente estimar el tensor de difusión y obtener las DTIs. La importancia de estos pasos es obtener tractografías de todas las DTIs partiendo desde imágenes con condiciones aproximadamente similares y así poder analizar los grupos Control y Desnutrición evitando a medida de lo posible el sesgo de una imagen con más movimiento o distorsión.

Corrección de movimiento, corrientes de Eddy y DTI

Así como la alineación en el procesamiento de T1W es de importancia, la corrección del movimiento entre imágenes por cada secuencia es importante. El objetivo

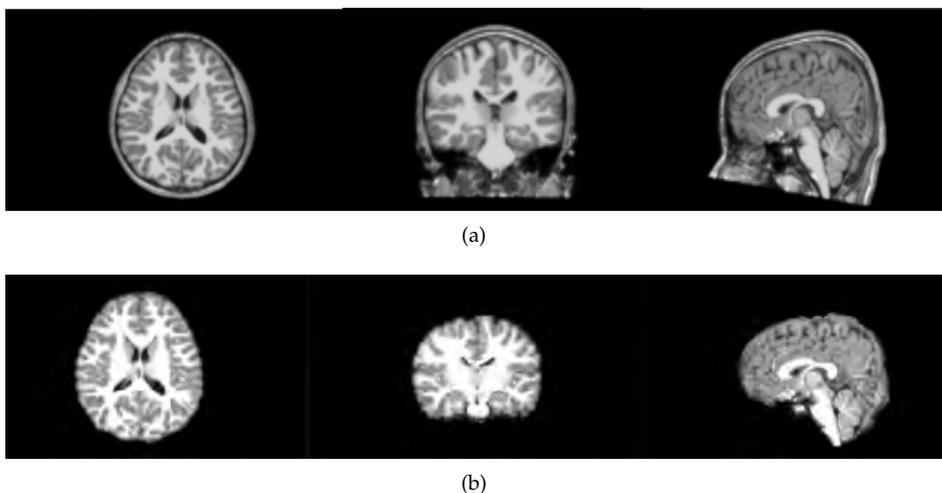


Figura 5.4: Extracción cerebral de MRI. En la figura (a) la imagen transformada de la original al espacio del atlas infantil y (b) su respectivo tejido cerebral únicamente.

de este paso es que cada una de las 32 secuencias esté correctamente orientada. Así también, los valores del tensor de difusión se calcularon usando b -valores inferiores a los $1750s/mm^2$, con los que fue posible construir una imagen DTI a partir de la DWI, como se observa en la Figura 5.5.

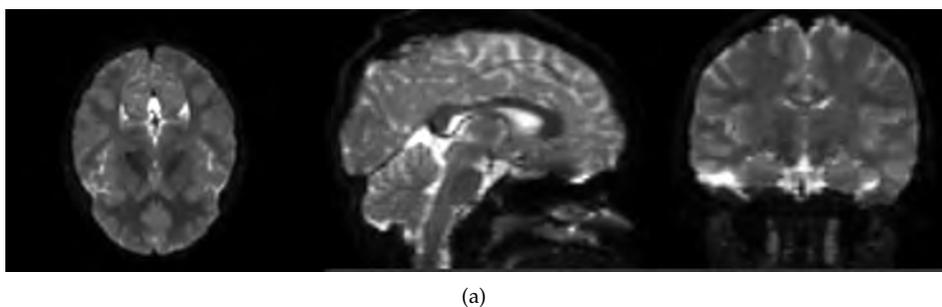


Figura 5.5: MRI de difusión de un paciente control: Vista axial, sagital y coronal de MRI DWI.

5.2. Procesamiento de MRIs

Una vez teniendo los vóxeles correspondientes al tejido cerebral, y se hayan generado las DTIs, será momento de segmentar la MRI estructural y de realizar el seguimiento de fibras a partir de las DTIs. Una vez realizado ese proceso, usando sus imágenes preprocesadas, se deberá realizar un registro a partir de la minimización de la distancia entre distintos marcadores o puntos de referencia (landmarks).

Segmentación en ROIs

Las MRIs preprocesadas fueron segmentadas utilizando el atlas de Desikan se segmentaron las MRI en 101 regiones, de acuerdo con [Klein and Tourville, 2012]. Como se observa en la Figura 5.6a, tenemos únicamente la vista sagital en 2 dimensiones; en la Figura 5.6b, se observa la misma vista del mismo sujeto con algunos segmentos como la corteza izquierda superior frontal y superior parietal de un color turquesa o la ínsula de corteza izquierda de amarillo; por último la Figura 5.6c, muestra las 101 regiones que conforman las ROIs en el MRI del paciente.

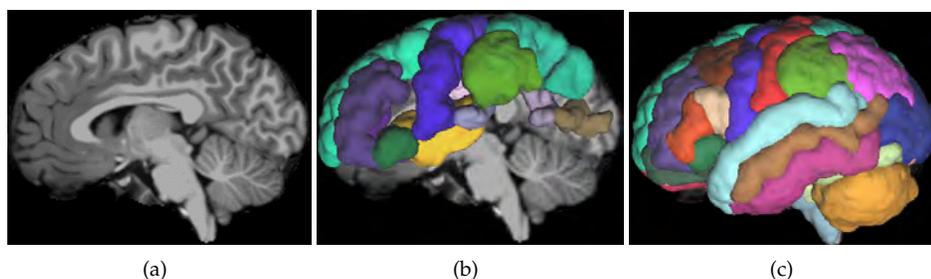


Figura 5.6: MRIs con corte sagital (a) original, (b) con un subconjunto de regiones del hemisferio izquierdo y (c) con las 101 regiones. En estas figuras se pueden ver como se enciman las regiones en el corte sagital dado que son volúmenes.

Seguimiento de fibras

En la figura 5.7 se pueden observar algunas regiones sobrepuestas en la MRI con vista sagital del sujeto control, a simple vista en a y b, parece que el tracto llamado *Cingulum Parolfactoru izquierdo* atraviesa ciertas regiones, sin embargo como se observa en c, esto no es así. En la Figura 5.8, se observa el mismo tracto atravesando las regiones corticales izquierdas: *isthmus-cingulate*, *posterior-cingulate* y *rostral-anterior-cingulate*. Los tractos que se consideran en el análisis de conectividad región con región, o desde el enfoque de redes vértice con vértice, son aquellos que empiezan y terminan entre dos regiones. Como se observa en la Figura 5.9, los k tractos se consideran en la conectividad de esas dos regiones o parcelas en la imagen superior, mientras que en la imagen inferior se muestra como se le restan m tractos al total a considerar.

Registro entre regiones y tractos de WM

A partir de las MRIs preprocesadas se realiza un registro usando una matriz afín para guardar el posicionamiento de cada vóxel. En este caso registraremos

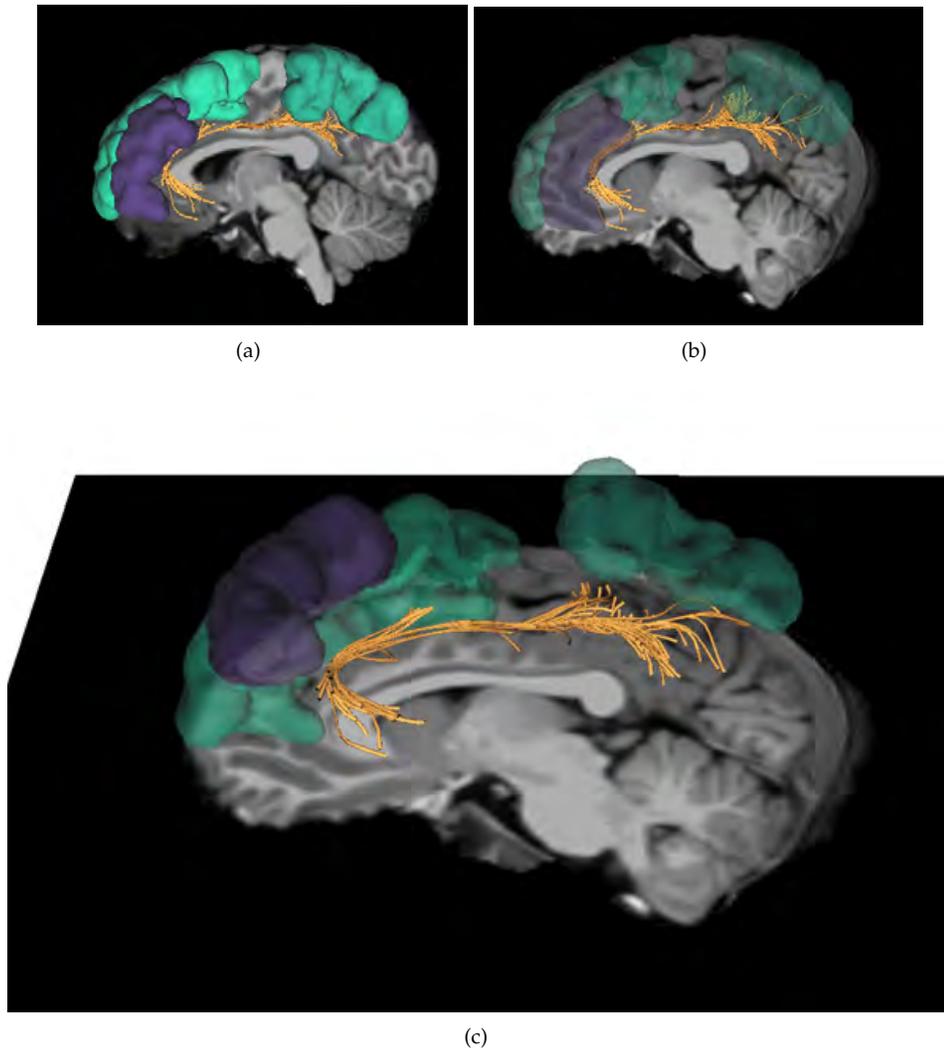


Figura 5.7: MRI con regiones activas y el tracto Cingulum Parolfactoru izquierda: (a) vista sin transparencia en las regiones, (b) vista con transparencia y (c) vista con inclinación.

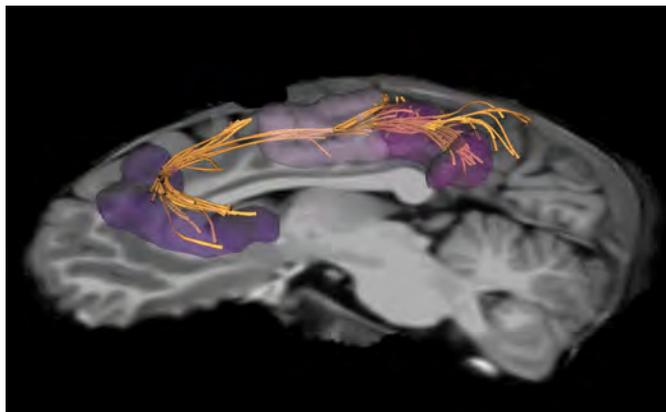


Figura 5.8: Tracto Cingulum Parolfactoru izquierdo teniendo contacto con tres regiones corticales.

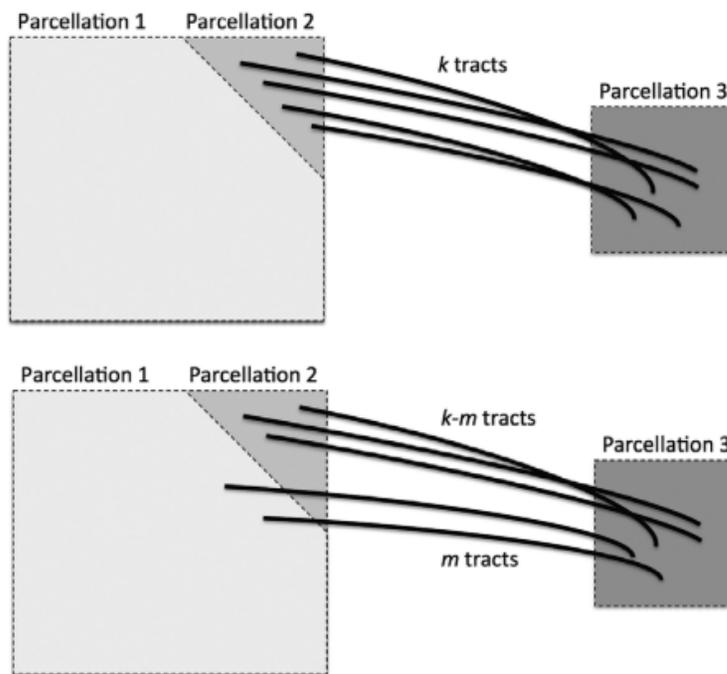


Figura 5.9: Ejemplo de los tractos que se consideran entre 3 regiones, o también llamadas parcelas (“*parcellation*” en inglés), siendo únicamente aquellas que empiezan o terminan en las regiones seleccionadas. Como se muestra en la figura superior, los k tractos empiezan y terminan en la región 2 y 3, mientras que en la inferior se muestra como los m tractos no se consideran dentro de las regiones 2 y 3, dado que no empiezan y terminan en dichas regiones. Fuente: [Chung, 2019]

la DTI en la T1W con únicamente el tejido correspondiente al cerebro, una vez teniendo la matriz de transformación se aplica a los voxeles correspondientes a los tractos obtenidos. Esto posicionará cada tracto en las respectivas regiones donde atraviesa. Véase Figura 5.11

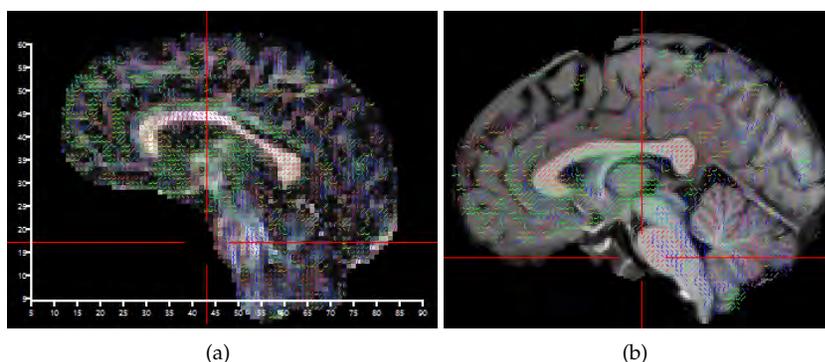


Figura 5.10: MRIs (a)DTI con los tractos coloreados según su orientación y (b)T1W con los tractos registrados a la modalidad.

5.3. Características obtenidas

Para la obtención de características fue necesario obtener primero las regiones cerebrales y la tractografía, ya que respectivamente representan los vértices y aristas del grafo donde se obtendrán características de red. Por sí solos los segmentos brindan información morfométrica y mediante la unión de vértices y aristas, por medio del registro entre modalidades, fue posible obtener biomarcadores de difusión para cada segmento, así como las características de redes antes mencionadas.

Las características mostradas en la Tabla 5.1 se obtuvieron por distintas técnicas como conteo de vóxeles y sus caras, distancias entre vóxeles, valores de difusividades, correlación y conectividad con la vecindad de vóxeles u otras regiones.

5.4. Selección de características por métodos estadísticos

La selección de características se hizo por medio de una técnica de filtro, en la cual cada característica es evaluada para verificar su relevancia estadística, es decir si poseen los valores estadísticos que puedan garantizar con mayor peso una diferencia notable entre un grupo y otro. Para esto se utilizaron métodos no paramétricos por la cantidad de sujetos se muestra. Un paso muy importante antes de

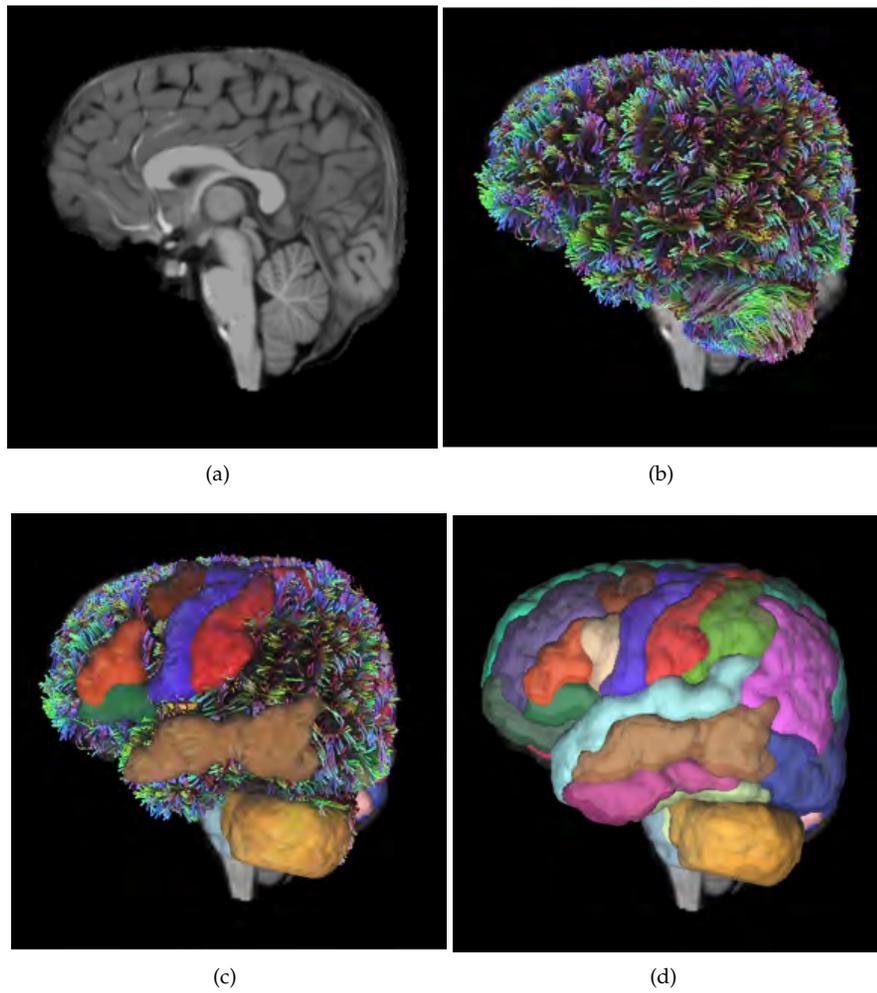


Figura 5.11: Tractografía registrada con la T1W, donde (a) es la T1W, (b) es la tractografía sobrepuesta a la T1W, subconjunto de regiones corticales del hemisferio izquierdo sobrepuestas en la tractografía y (d) las 101 regiones sobrepuestas en la tractografía.

Tabla 5.1: Tabla de las características obtenidas a partir de las T1W y DWI. Las características de difusión son la media, mínimo y máximo de las mencionadas.

Morfométricas	Difusión	Redes
<ul style="list-style-type: none"> ▪ Volumen ▪ Volumen normalizado ▪ Área de la superficie ▪ Redondez ▪ Aplanamiento ▪ Diámetro de Feret ▪ Elongación ▪ Compacidad discreta ▪ Tortuosidad 	<ul style="list-style-type: none"> ▪ Anisotropía Fraccional ▪ Difusividad Media ▪ Difusividad Radial ▪ Difusividad Axial 	<ul style="list-style-type: none"> ▪ Centralidad de grado o grado ▪ Centralidad de eigenvector ▪ Centralidad Pagerank

empezar la selección y evaluación, es formar un conjunto de datos para el entrenamiento y otro conjunto de datos para la validación final.

Se obtuvieron las características estadísticamente relevantes únicamente de los conjuntos de entrenamiento. Una vez obtenidos se realizaron clasificaciones cuyas métricas nos brindaron un indicador acerca del desempeño de esas características para discriminar entre las clases.

Para evitar el riesgo que puede tener haber elegido al azar los sujetos para la validación, se realizó el proceso de clasificación con sujetos elegidos de forma aleatoria y repitiendo el experimento 10 ocasiones. Aquellas características seleccionadas más de 7 ocasiones se agregaron a la lista.

Se obtuvieron 2 conjuntos a evaluar, el seleccionado por la prueba de Kolmogorov Smirnov y el seleccionado por la prueba de Mann Withney. Estos conjuntos contienen las características seleccionadas por el método de filtro (*filter method*, definido en la subsección 2.3.2 del marco teórico), cuyos pasos a seguir fueron los siguientes:

1. Se realizó la prueba individual de cada característica, teniendo como hipótesis que ambas poblaciones (grupo control y desnutrición) pertenecen a la misma población.

2. Se calculó su valor p y se seleccionaron únicamente aquellas con un valor p menor a 0.05.
3. Se formó la lista de características de KS y MW.

El resultado de eso, fueron las siguientes características.

Mann Withney

Como se pueden observar en la Tabla 5.2, cuando filtramos aquellas características cuyo valor p es inferior a 0.05 usando la prueba U de Mann Whitney, obtenemos 101 características donde predominan las morfométricas, seguido de las de redes y por último las de difusión.

Tabla 5.2: Descriptores de las características seleccionadas a partir de la prueba U de Mann Whitney. La cantidad que se muestra son de las combinaciones entre la unión del descriptor y una región. Por lo que existen 53 características morfométricas cuyos descriptores son los mencionados, y así con los otros tipos de características.

Tipo	Descriptor	Cantidad
Morfométricas	▪ Volumen	53
	▪ Diámetro de Feret	
	▪ Área de la superficie	
	▪ Redondez	
	▪ Elongación	
Difusión	▪ Difusividades media, axial y radial	14
	▪ Anisotropía fraccional	
Redes	▪ Grado del vértice (Grafo binario)	34
	▪ Coeficiente de agrupamiento (grafo binario)	
	▪ Centralidad eigenvector, pagerank y betweenness	

Las regiones de las características seleccionadas las presentamos dividiendo las regiones pertenecientes a la corteza y las debajo de la corteza. El número de características pertenecientes a las regiones corticales es aproximado al 60%, y el

biomarcador tipo de característica predominante en estas regiones son las características de redes, véase Tabla 5.3. En cuanto a las regiones subcorticales, el hemisferio izquierdo fue el tuvo más regiones seleccionadas, siendo más del 90 % las características morfométricas, véase Tabla 5.4.

Tabla 5.3: Regiones corticales de las características seleccionadas a partir de la prueba de Mann Whitney.

Estructura	Hemisferio Izquierdo	Hemisferio derecho
Corticales	■ <i>caudal-anterior-cingulate</i>	■ <i>cuneus</i>
	■ <i>caudal-middle-frontal</i>	■ <i>caudal-anterior-cingulate</i>
	■ <i>entorhinal</i>	■ <i>caudal-middle-frontal</i>
	■ <i>lateral-occipital</i>	■ <i>entorhinal</i>
	■ <i>lingual</i>	■ <i>inferior-parietal</i>
	■ <i>medial-orbitofrontal</i>	■ <i>inferior-temporal</i>
	■ <i>paracentral</i>	■ <i>isthmus-cingulate</i>
	■ <i>parahippocampal</i>	■ <i>lateral-orbitofrontal</i>
	■ <i>postcentral</i>	■ <i>paracentral</i>
	■ <i>posterior-cingulate</i>	■ <i>parsopercularis</i>
	■ <i>rostral-anterior-cingulate</i>	■ <i>posterior-cingulate</i>
	■ <i>superior-frontal</i>	■ <i>rostral-anterior-cingulate</i>
	■ <i>transverse-temporal</i>	■ <i>rostral-middle-frontal</i>
		■ <i>superior-frontal</i>
	■ <i>supramarginal</i>	
	■ <i>transverse-temporal</i>	

Kolmogorov Smirnov

Mientras que las características seleccionadas por medio de la prueba de MW, la proporción de características morfométricas y de redes fue similar en el subconjunto de características seleccionadas por la prueba de KS. Véase Tabla 5.5.

Dentro de las regiones pertenecientes a este subconjunto predominaron las regiones corticales del hemisferio izquierdo. Siendo minoría las características subcorticales, véase la Tabla 5.3 para regiones corticales y la Tabla 5.7 para subcortica-

Tabla 5.4: Regiones subcorticales de las características seleccionadas a partir de la prueba de Mann Whitney.

Estructura	Región
Subcorticales	Hemisferio izquierdo
	<ul style="list-style-type: none"> ▪ <i>Caudate</i> ▪ <i>Cerebral-WM</i> ▪ <i>Choroid-plexus</i> ▪ <i>Pallidum</i> ▪ <i>Putamen</i> ▪ <i>VentralDC</i>
	Hemisferio derecho
	<ul style="list-style-type: none"> ▪ <i>Cerebral-WM</i> ▪ <i>Hippocampus</i> ▪ <i>VentralDC</i>
	Hemisferio cerebelosos
	<ul style="list-style-type: none"> ▪ <i>Cerebellum-Cortex</i> ▪ <i>Cerebellum-WM</i>
	Diencéfalo
	<ul style="list-style-type: none"> ▪ <i>Thalamus-Propor</i>
	Metencéfalo
	<ul style="list-style-type: none"> ▪ <i>Brainstem</i>
	Telencéfalo
	<ul style="list-style-type: none"> ▪ <i>Corpus Callosum middle-anterior</i> ▪ <i>Corpus Callosum middle-posterior</i>

Tabla 5.5: Descriptores de las características seleccionadas a partir de la prueba prueba Kolmogorov Smirnov. La cantidad que se muestra son de las combinaciones entre la unión del descriptor y una región. Por lo que existen 36 características morfométricas cuyos descriptores son los mencionados, y así con los otros tipos de características.

Tipo	Descriptor	Cantidad
Morfométricas	▪ Volumen	36
	▪ Diámetro de Feret	
	▪ Área de la superficie	
	▪ Redondez	
	▪ Elongación	
Difusión	▪ Difusividades media, axial y radial	24
	▪ Anisotropía fraccional	
Redes	▪ Grado del vértice (Grafo binario)	39
	▪ Coeficiente de agrupamiento (grafo binario)	
	▪ Centralidad eigenvector y <i>pagerank</i> .	

Tabla 5.6: Regiones corticales de las características seleccionadas a partir de la prueba de Kolmogorov Smirnov.

Estructura	Hemisferio Izquierdo	Hemisferio derecho
Corticales	▪ <i>caudal-anterior-cingulate</i>	▪ <i>caudal-anterior-cingulate</i>
	▪ <i>caudal-middle-frontal</i>	▪ <i>caudal-middle-frontal</i>
	▪ <i>entorhinal</i>	▪ <i>entorhinal</i>
	▪ <i>fusiform</i>	▪ <i>inferior-parietal</i>
	▪ <i>inferior-parietal</i>	▪ <i>inferior-temporal</i>
	▪ <i>lateral-occipital</i>	▪ <i>isthmus-cingulate</i>
	▪ <i>lingual</i>	▪ <i>lateral-orbito-frontal</i>
	▪ <i>medial-orbitofrontal</i>	▪ <i>paracentral</i>
	▪ <i>paracentral</i>	▪ <i>parsopercularis</i>
	▪ <i>parahippocampal</i>	▪ <i>posterior-cingulate</i>
	▪ <i>postcentral</i>	▪ <i>rostral-anterior-cingulate</i>
	▪ <i>posterior-cingulate</i>	▪ <i>rostral-middle-frontal</i>
	▪ <i>rostral-anterior-cingulate</i>	▪ <i>superior-frontal</i>
	▪ <i>superior-frontal</i>	▪ <i>supramarginal</i>
	▪ <i>transverse-temporal</i>	▪ <i>transverse-temporal</i>

les.

5.5. Validación por métricas de clasificación

Dado que el objetivo es encontrar ese subconjunto de características que permitan la mejor diferenciación entre sujetos sanos y desnutridos, se cuantificó el desempeño de ambos subconjuntos de características clasificando ambos sujetos por medio de algoritmos de aprendizaje automático.

Para fines prácticos nombraremos MW al subconjunto de características seleccionadas por la prueba de Mann Withney y KS al subconjunto seleccionado por la prueba de Kolmogorov Smirnov. Una vez decretado el conjunto con la mejor selección, se procede a hacer la división entre regiones corticales y subcorticales.

Los resultados fueron los siguientes:

Tabla 5.7: Regiones subcorticales de las características seleccionadas a partir de la prueba de Kolmogorov Smirnov.

Estructura	Región
Subcorticales	Hemisferio izquierdo
	▪ <i>Cerebral-WM</i>
	▪ <i>VentralDC</i>
	Hemisferio derecho
	▪ <i>VentralDC</i>
	Hemisferio cerebelosos
	▪ <i>Cerebellum-Cortex (l - r)</i>
	Metencéfalo
	▪ <i>Brainstem</i>
	Telencéfalo
▪ <i>Corpus Callosum middle-anterior</i>	

Desempeño de clasificación utilizando MW

Como podemos observar en la Tabla 5.8, las métricas de desempeño rondaron entre 66 y 100, donde el promedio de las medianas dio de resultado una exactitud de 88, una sensibilidad de 90, una precisión de 91 y una puntuación F1 de 87, dando un total de 356.

Desempeño de clasificación utilizando KS

Como podemos observar en la Tabla 5.9, las métricas de desempeño rondaron igual, entre 66 y 100, pero el promedio de las medianas dio de resultado una exactitud de 82, una sensibilidad de 93, una precisión de 91 y una puntuación F1 de 87, dando un total de 355.

Subconjunto del conjunto ganador

Hasta aquí, las características del conjunto MW ganaron por suma de puntuación, lo que queda ahora es evaluar si considerar únicamente aquellas características con regiones corticales o subcorticales es beneficioso para el modelo. Se muestran los resultados en la Tabla 5.10, en la cual se observa que utilizar las regiones

Tabla 5.8: Resultados de clasificación utilizando MW

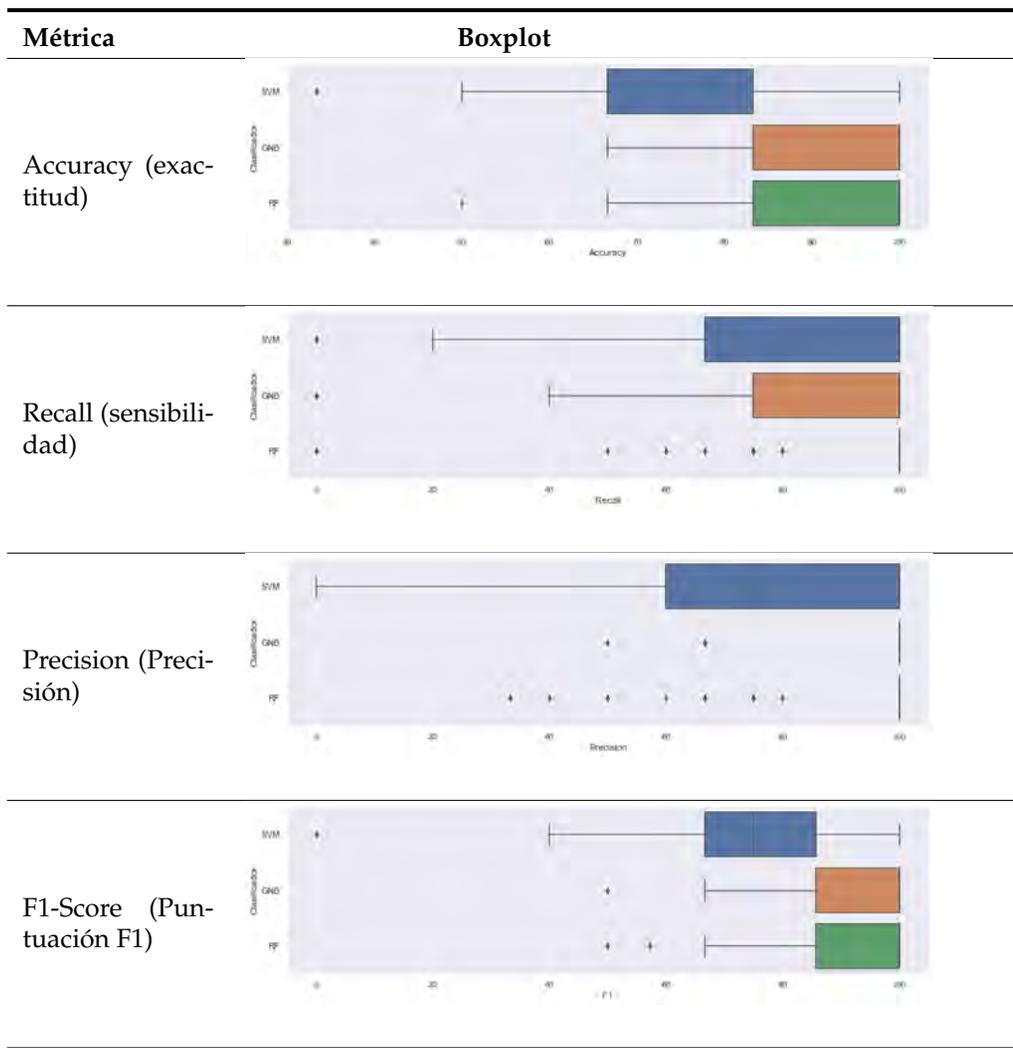
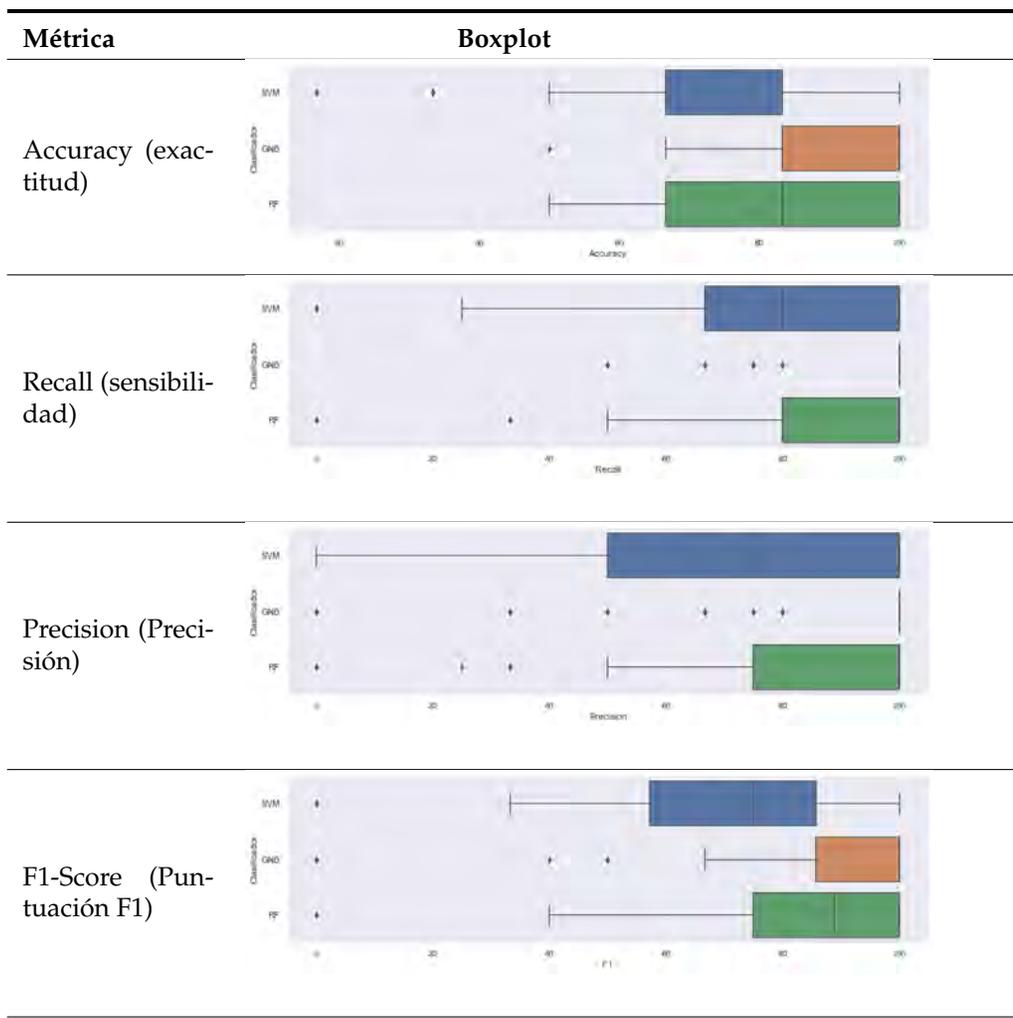


Tabla 5.9: Resultados de clasificación utilizando KS



corticales o subcorticales conlleva a unos resultados favorables, pero el conjunto con mejor desempeño fue aquel que considera únicamente características de las regiones corticales.

Tabla 5.10: Resultados de los promedios de métricas de clasificación usando el conjunto con características de regiones únicamente corticales y únicamente subcorticales.

Subconjunto	Métrica	Resultado
Regiones corticales	▪ Accuracy:	89
	▪ Recall:	91
	▪ Precision:	93
	▪ F1 Score:	90
Regiones subcorticales	▪ Accuracy:	84
	▪ Recall:	88
	▪ Precision:	87
	▪ F1 Score:	86

5.6. Visualización del conectoma

Dado que las regiones corticales tuvieron un mejor desempeño que las subcorticales en cuanto a tasa de clasificación, a partir de esas regiones y sus conexiones físicas, se obtuvieron los conectogramas. Por medio de la técnica del conectograma podremos encontrar diferencias y similitudes en cuanto a conectividad cerebral. De cada sujeto se obtuvo la matriz de conectividad a través de las conexiones como se muestra en la Figura 5.12.

Así como se observa el grafo de la Figura 5.12c, la visualización brinda información espacial, más su información de conectividad depende de la manipulación tridimensional, por lo que visualizarlo desde el conectograma resulta ser más rico en cuanto a la información de la red. Existen varios puntos a observar como lo son la distancia que abarca cierta región en el perímetro del círculo, la cual determina el peso total de las conexiones totales de dicho vértice, con lo que podemos saber qué regiones son más importantes bajo el criterio de conectividad y centralidad. Véase en la Figura 5.13a el conectograma del sujeto Control 1, del cual podríamos desta-

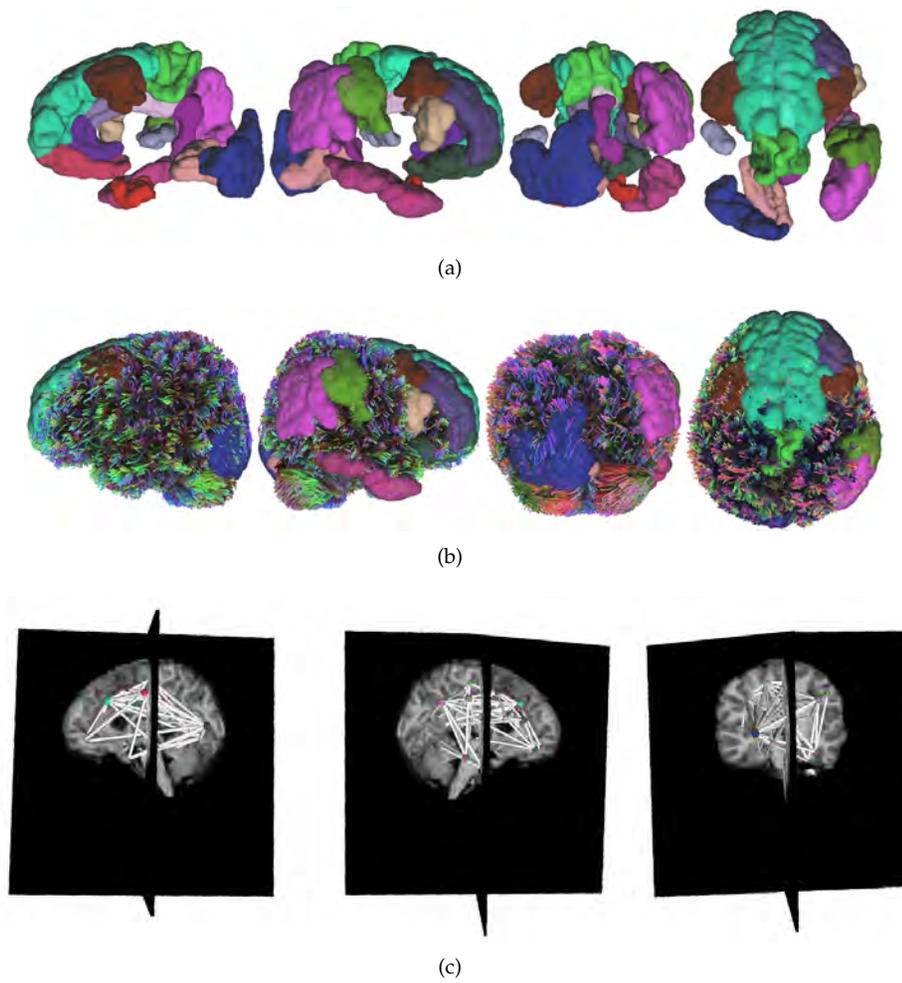


Figura 5.12: Construcción generalizada del grafo formado únicamente por las regiones corticales seleccionadas mostradas en la Tabla 5.4. Donde, (a) es la representación espacial de dichas regiones, (b) es la imagen registrada entre las regiones y la tractografía y (c) el grafo espacial formado por la matriz de conectividad.

car la importancia de la región cerebral superior frontal derecha de color morado o la lateral orbito frontal derecha que se muestra en la Figura 5.13.

Sujetos control

Como se observa en la Figura 5.15a, la alta conectividad dificultan visualizar ciertos patrones de los grafos, por lo que información como el grado del vértice son de los únicos que se aprecian. La forma en que se interpreta el grado del vértice es mediante la distancia que abarca en el perímetro del círculo, como se muestra en la Figura 5.15b.

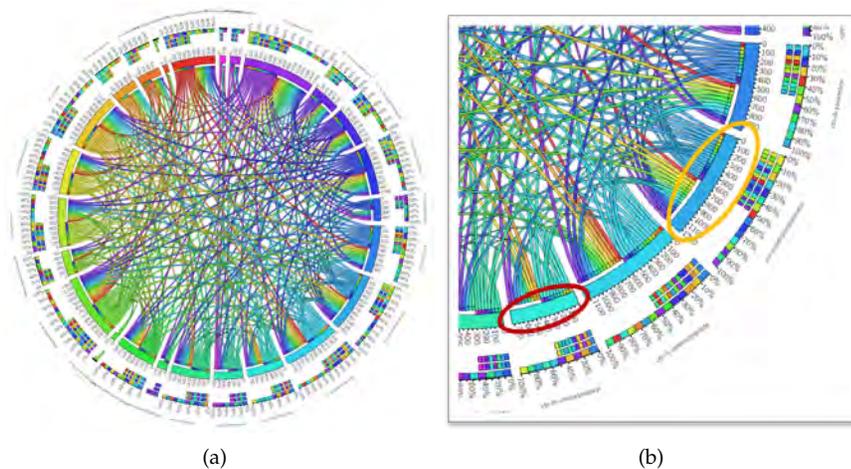


Figura 5.13: Conectograma de sujeto Control 1 de (a) todas las regiones y (b) un cuarto de las regiones. Cada color de las aristas representa el origen, es decir, de la región o vértice que viene. Véase en la figura (b), dos tonalidades de azul distintas entre la región rodeada de amarillo y la región rodeada de rojo. En la figura (b), también se presenta la importancia de la distancia en el perímetro del círculo, la cual representa mayor peso en las conexiones. Se muestra rodeada por el óvalo amarillo la región lateral orbito-frontal, la cual abarca una distancia mayor que la región inferior-temporal, ambas regiones de la corteza cerebral del hemisferio derecho.

El objetivo de estos gráficos es auxiliar a una interpretación rápida y efectiva acerca de la conectividad de una red o grafo. Esta tarea se facilita conforme se reducen las conexiones haciendo un grafo más simple pero esa eliminación de aristas debe de realizarse por medio de algún parámetro. Para el fin de este proyecto, es de interés aquellas regiones con conectividad fuerte, por lo que filtrar con umbral por encima del cuartil Q3 nos brindó la simpleza que se aprecia en los conectogramas de la Figura 5.16. Aquí es posible darse cuenta que la región transversal temporal derecha o la región supramarginal derecha tienen muy pocas uniones con el resto de las regiones, caso contrario a la región superior frontal izquierda o medial orbito-frontal derecha que tienden a ser las regiones con mayor grado

en el grafo de conexiones filtradas.

Sujetos desnutrición

Dentro de las regiones con similitudes a los sujetos control en cuanto a conectividad de alto valor tenemos a las regiones superior frontales izquierda y derecha, y en las que se tiene una conectividad mayor en los de desnutrición es en la región rostral medio frontal derecha. Véase los conectogramas completos de los sujetos desnutrición en la Figura 5.17. Esto si comparamos los conectogramas completos de ambos grupos, dada la dificultad que esto conlleva por la alta conectividad de los grafos, la comparación la haremos con los grafos filtrados por los valores superiores al cuartil Q3, véanse en la Figura 5.18, y véase con más detalle en la figura 5.14.

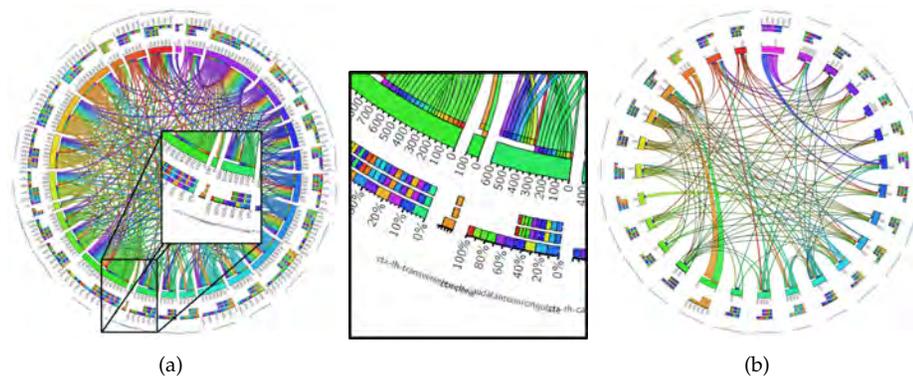


Figura 5.14: Comparación de la reducción de conexiones del sujeto Desnutrición 1. En la figura (a) vemos el conectograma sin reducción de conexiones, dando un énfasis en la región ctx-lh-transversal temporal, en la cual únicamente existe una conexión con la región ctx-lh-lateral occipital, representada de color naranja, esto se observa en la figura (b). Un punto a observar en la figura (c) es la presencia del grosor en algunas de las conexiones. Este engrosamiento de las conexiones significa la relevancia de la arista con respecto a la región en la que se conecta con el mayor grosor. Quiere decir, que cuando una conexión representa un porcentaje alto de las conexiones totales de esa región, existe un engrosamiento en la arista incidente a la región. Por lo que la arista verde y naranja tienden a engrosarse en la región ctx-lh-transversal temporal, dado que representan su 100 % de conectividad.

A diferencia del grupo control, en el grupo desnutrición tenemos una presencia constante de la región transversal temporal derecha, manteniendo una unión con las regiones del hemisferio derecho: inferior parietal derecha, superior frontal y *parsopercularis*.

Comparación entre conectogramas de sujetos Control y Desnutrición

La visualización del conectoma por medio del conectograma necesitó un tratamiento de eliminación de aristas, dado que la alta densidad de conectividad sin esta dificulta la tarea de distinguir un grupo de otro o de encontrar patrones entre sí. Eliminar aquellas inferiores al cuartil Q_3 permite una visualización de únicamente aquellas conexiones fuertes o con mayor peso. Con esto, es posible comparar tanto las regiones entre sí de un mismo grupo como la comparación entre dos o más grupos.

En la Figura 5.15, se muestran los 12 sujetos Control, en dicha figura se observó que las regiones con más conectividad en estos sujetos son las regiones lateral-occipital, lingual y superior-frontal del hemisferio izquierdo, y del hemisferio derecho: lateral orbito frontal, inferior temporal y superior frontal. Mientras que en la Figura 5.16, se pueden observar regiones que no cuentan con aristas que incidan en ellas, estas regiones son: la ctx-rh-supramarginal, ctx-lh-caudal medio frontal y la ctx-lh-transversal temporal.

En la Figura 5.17, se puede visualizar la similitud, en cuanto a importancia de vértice se refiere, de las regiones superior frontal de ambos hemisferios y la lingual del hemisferio izquierdo. A diferencia de los sujetos control, en este grupo se encontró que la región ctx-lh-postcentral, cuenta con más importancia en este grupo. Una vez se filtran las aristas con Q_3 , véase Figura 5.18 se puede observar la importancia de ciertas regiones a sus únicas aristas, pero siendo la región ctx-lh-transversal temporal la que no presentó aristas en la mayoría de los sujetos.

Otro enfoque a considerar acerca de la conectividad son los descriptores de redes, los cuales pueden brindar información de un vértice, así como de un conjunto de vértices. Los descriptores de cada vértice fueron de utilidad para encontrar aquellas regiones cuyos descriptores difieren entre ambos grupos con relevancia estadística y las métricas de red, nos brindan coeficientes que describen la red en su totalidad, estas métricas fueron el coeficiente de asortatividad, *small worldness*, coeficiente de agrupamiento y diámetro del grafo.

A continuación se presentan las distribuciones por medio de estimadores de densidad de kernel (KDE, por sus siglas en inglés), donde se muestran las diferencias entre utilizar todas las regiones a utilizar únicamente las regiones corticales seleccionadas en la Tabla 5.3. Como se observa en la Tabla 5.11, los valores obtenidos usando todas las regiones de ambos sujetos, tienden a tener un rango de

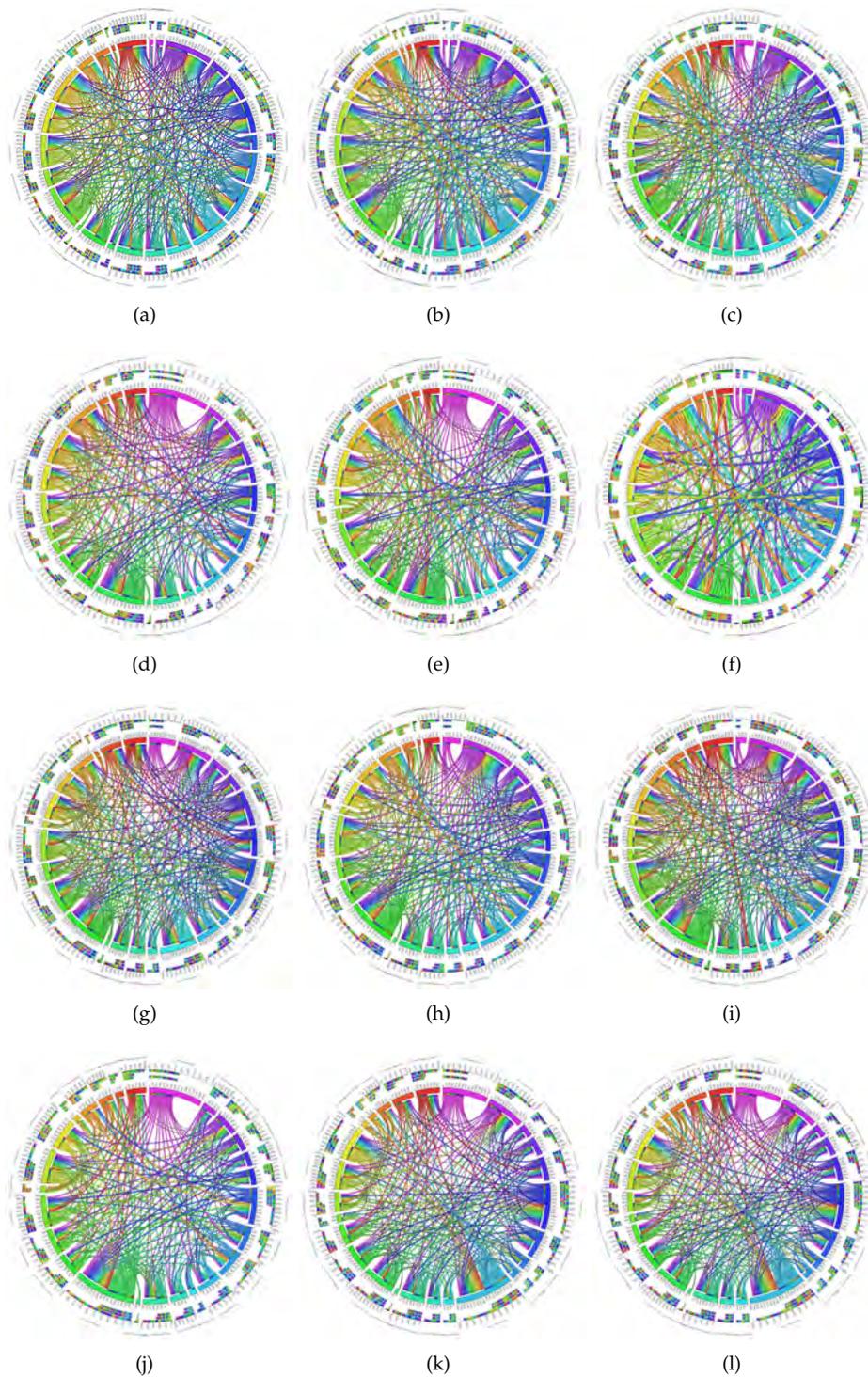


Figura 5.15: Resultados de los conectogramas de sujetos control con todas las conexiones activas. Conectogramas de las regiones de importancia, donde se observa qué regiones tienden a tener más conectividad que otras y en qué regiones tienen esa conectividad. Las regiones con mayor conectividad resultaron ser del hemisferio izquierdo: lateral occipital, lingual y superior frontal; mientras que del hemisferio derecho: lateral orbito frontal, inferior temporal y superior frontal.

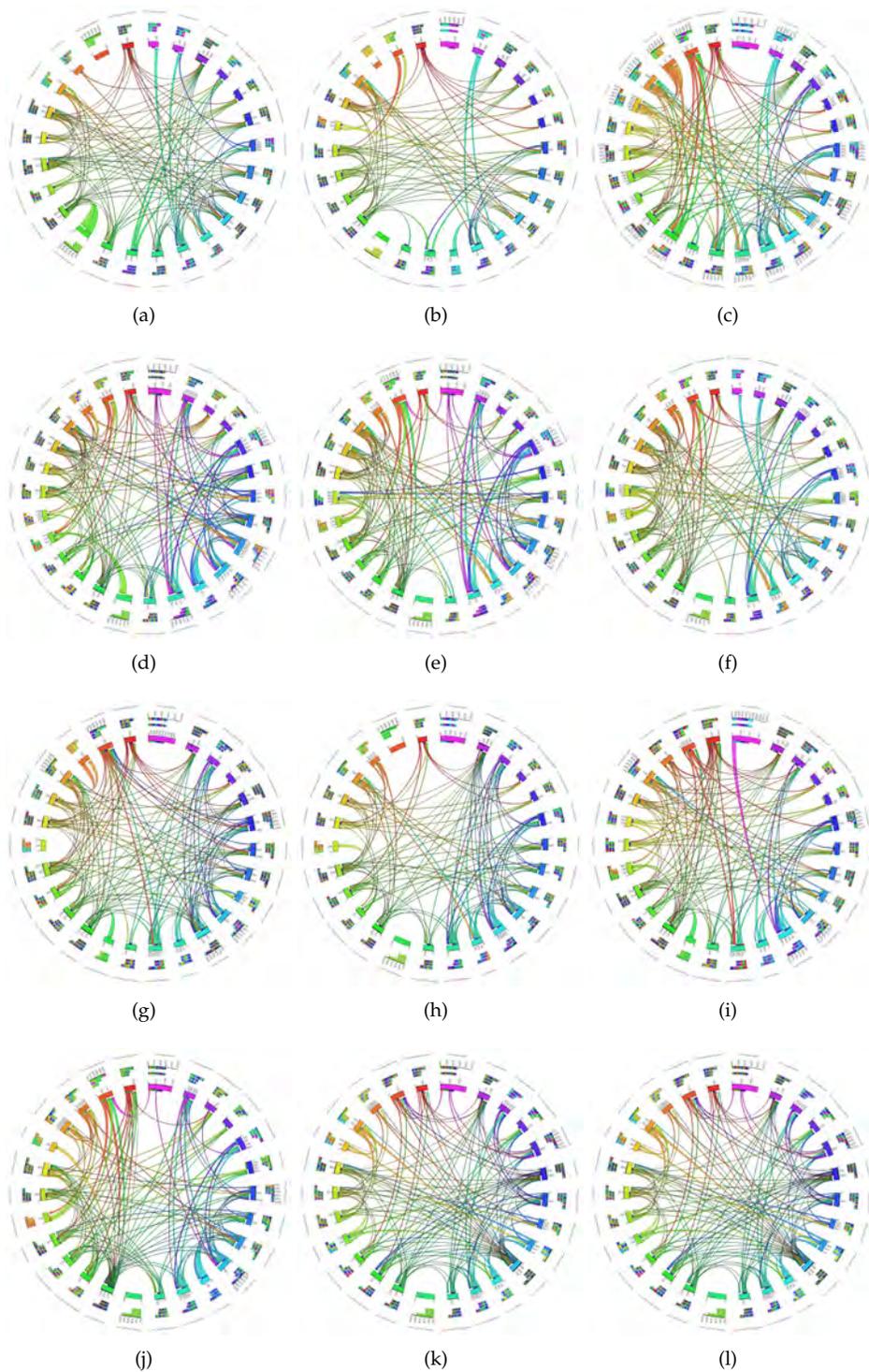


Figura 5.16: Resultados de los conectogramas de sujetos control filtrado a partir del cuartil Q_3 . Con este método se observan las conexiones con mayor fuerza, por lo que es posible darse cuenta qué regiones, a pesar de tener una conectividad alta con respecto a otras regiones, la fuerza de su unión no es la suficiente para mantener aristas en esta visualización. Esto se observa en (b)(c)(g)(h) en la región ctx-rh-supramarginal, (a) en la región ctx-lh-caudal medio frontal en (a)(b) y en la región ctx-lh-transversal temporal en (b)(f)(h)(j)(l).

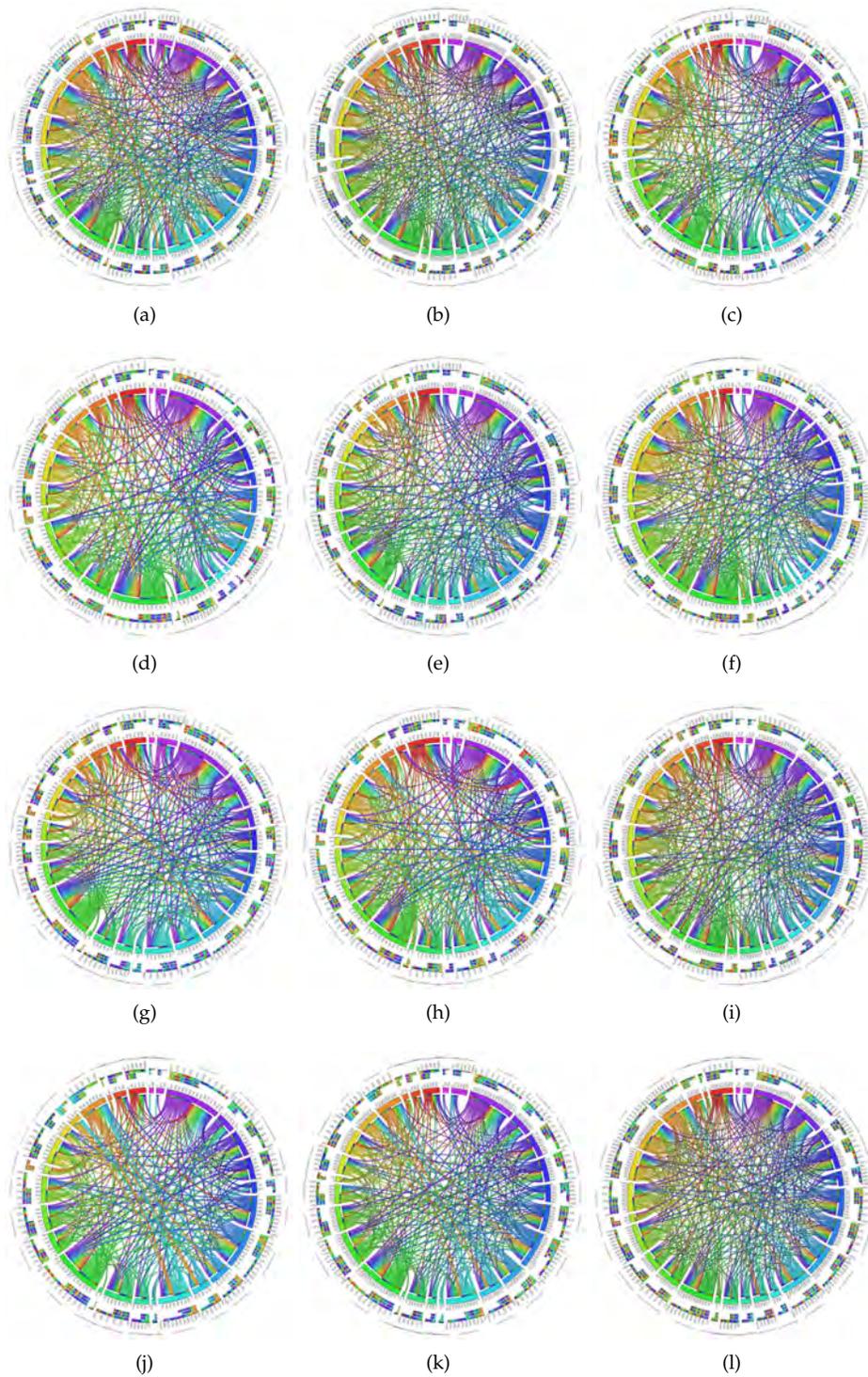


Figura 5.17: Resultados de los conectogramas de sujetos desnutrición. Al igual que en los conectogramas de los sujetos control, se encuentra que las regiones con más importancia en cuanto a conectividad es la región superior frontal en ambos hemisferios, así como el lingual del hemisferio izquierdo. Sin embargo, una región que destaca por su densidad de conexiones a diferencia de los sujetos control es la ctx-lh-postcentral.

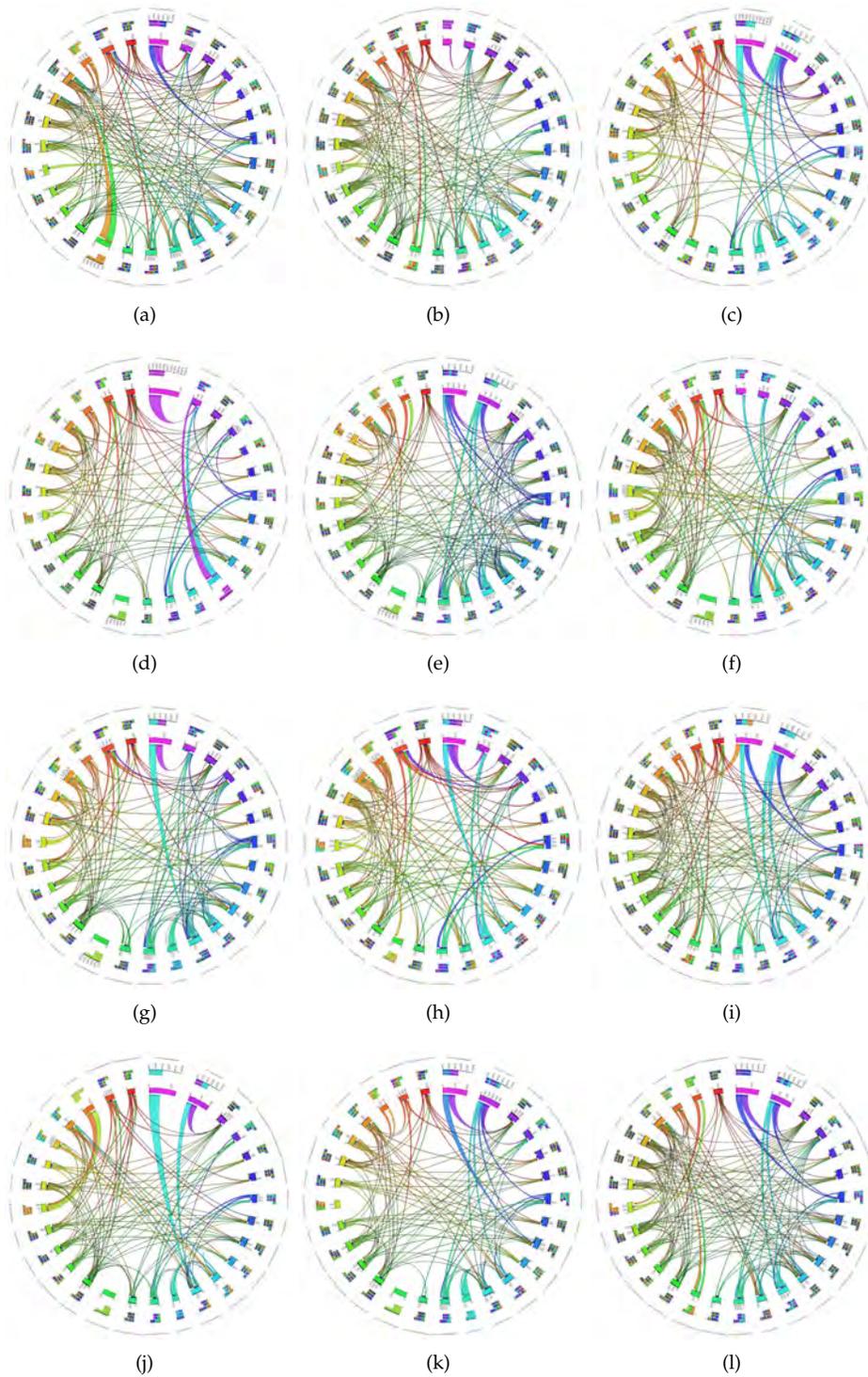
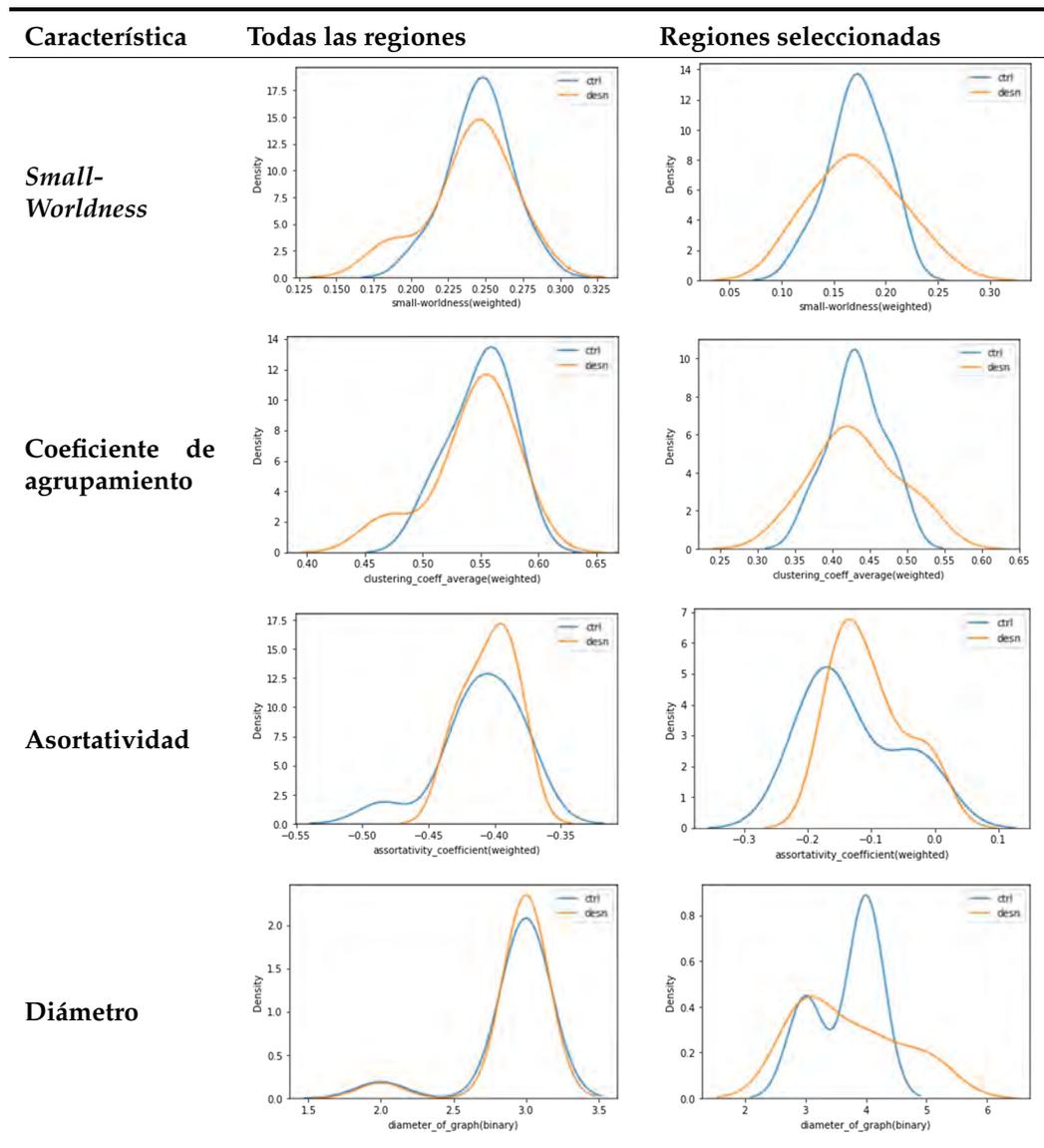


Figura 5.18: Resultados de los conectogramas de sujetos desnutrición filtrado a partir del cuartil Q_3 . A diferencia de los sujetos control, solo una región se vio afectada de no tener aristas incidentes, la región ctX-lh-transversal temporal, véase (c)(d)(e)(f)(g)(j)(k). Otro factor diferenciador es la presencia de fuerza de la conexión demostrada por las aristas que inciden con la región ctX-rh-transversal temporal y ctX-lh-entorhinal.

Tabla 5.11: Resultados por medio de KDE de las métricas de las redes completas comparando sujetos sanos (ctrl) con sujetos desnutridos (desn).



valores y una distribución menos diferenciable que cuando usamos únicamente las regiones corticales seleccionadas. Aunque ninguna de las características de redes obtenida tiene la suficiente significancia estadística para ser diferenciable por algún algoritmo, la asortatividad da indicios de ser la de mayor potencial en cuanto a diferenciar un sujeto Control y uno de Desnutrición. Esta propiedad, es la tendencia de los vértices de conectarse con otros vértices con características similares, por lo que enfocar el conectograma únicamente en aquellos vértices con características similares, tomando tanto morfométricas como de difusión, se desarrolló y se encuentra en el Apéndice A, donde los resultados fueron visualmente diferenciables.

CAPÍTULO 6

CONCLUSIÓN

Al comparar atributos de un grupo de sujetos sanos con otro de sujetos con algún padecimiento, se espera encontrar diferencias significativas entre ellos, sin embargo, dependerá del atributo a comparar si existe o no esa separación entre los valores de un grupo con el otro. El enfoque que siguió este trabajo fue el de comparar la conectividad física de las regiones cerebrales por medio de los tractos generados por el tensor de difusión, brinda la información del cerebro completo a un nivel de complejidad menor al que si se estudiara a una micro escala o nano escala, pero con menores tiempos de procesamiento, útil para un primer acercamiento.

Existen diversas formas de analizar la conectividad, así como de visualizarla, el usar el conectograma fue de utilidad para observar qué vértices (regiones cerebrales) nos brindan más información que otras, en cuanto a su conectividad con el resto de regiones y los patrones que forman. Aunque analizar todas las regiones brinda información del cerebro completo, nos fue de mayor utilidad reducir el número de vértices en el grafo utilizando la relevancia estadística de sus características morfométricas, de difusión y de red enfocadas en el vértice únicamente. Como era de esperarse, las características que incluían únicamente regiones corticales superaron al resto de conjuntos, dado que en la corteza se encuentran aquellas zonas funcionales de mayor complejidad en los humanos como la percepción, la imaginación, el juicio o la toma de decisiones, funciones en las que regularmente suelen diferir los sujetos sanos de aquellos con algún padecimiento neurodegenerativo.

El análisis de redes del grafo formado por las conexiones físicas entre los tractos y las regiones del cerebro, brinda información sobre las redes de los sujetos, más no siempre brinda una diferencia apreciable a simple vista. Esto mismo ocu-

rió en [Chung et al., 2017], donde comparó la conectividad de 32 regiones entre sujetos control y sujeto pediátrico víctima de maltrato infantil, por medio del conteo de tractos y la anisotropía fraccional, y tampoco presentó una matriz o grafo lo suficientemente distinto uno del otro, por lo que se obtuvieron métricas de redes de las redes completas para cada cerebro, esto fue de gran ayuda, ya que por medio de KDE es posible observar a simple vista, cuanto difieren las funciones de probabilidad entre la totalidad de sujetos sanos de los sujetos con desnutrición.

Un enfoque para la obtención de conectogramas a partir de sus características morfométricas, de difusión y de red, se muestra en el Apéndice A. El conectograma obtenido por medio de esta técnica, brinda una diferencia a simple vista en el análisis grupal de los sujetos. Dado que la importancia de tener certeza si el sujeto se encuentra o no desnutrido por medio de las características de su cerebro, es que se consideró tomar como base de grafo completo a todos aquellos conectogramas cuyas características se encontraran en el rango de los sujetos desnutridos, siendo así un grafo completo el sujeto desnutrido prototipo y un grafo sin vértices el del sujeto control prototipo. Esta técnica no es la estándar en cuanto a la obtención del conectoma o conectogramas, sin embargo es de utilidad para distinguir clases de “un vistazo”.

Dadas las limitaciones que se tienen cuando la muestra es pequeña, sería de utilidad realizar el procedimiento con más sujetos, y a su vez con un mayor número de características tanto morfométricas como de difusión y de redes. Otro enfoque que se ha demostrado ser de gran relevancia es la MRI funcional, con la que se puede complementar más información y realizar un enfoque que combine la modalidad estructural y la funcional.

REFERENCIAS

- [Andy Brain Blog, 2018] Andy Brain Blog (2018). Introduction to FreeSurfer, Part 1: Overview & Installation., <https://www.andysbrainblog.com/andysbrainblog/2018/11/27/introduction-to-freesurfer>.
- [Anoop et al., 2004] Anoop, S., Saravanan, B., Joseph, A., Cherian, A., and Jacob, K. S. (2004). Maternal depression and low maternal intelligence as risk factors for malnutrition in children: a community based case-control study from South India. *Arch. Dis. Child.*, 89(4):325–329.
- [Bardin, 2012] Bardin, J. (2012). Making Connections. *Nature*, 483:394–396.
- [Basser et al., 2000] Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.*, 44(4):625–632.
- [Bassett and Bullmore, 2006] Bassett, D. S. and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, 12(6):512–523.
- [Bennett et al., 2018] Bennett, S. H., Kirby, A. J., and Finnerty, G. T. (2018). Rewiring the connectome: Evidence and effects. *Neurosci. Biobehav. Rev.*, 88(March):51–62.
- [Bribiesca, 2000] Bribiesca, E. (2000). Measure of compactness for 3D shapes. *Comput. Math. with Appl.*, 40(10):1275–1284.
- [Bribiesca, 2008] Bribiesca, E. (2008). An easy measure of compactness for 2D and 3D shapes. *Pattern Recognit.*, 41(2):543–554.
- [Bullmore and Sporns, 2009] Bullmore, E. and Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186–198.

- [Carrillo-Bermejo et al., 2018] Carrillo-Bermejo, A., Hevia-montiel, N., Molinero re, E., and Aut, U. (2018). Características morfoétricas en dominio discreto para reconocimiento de tumores cerebrales. *147(7)*:319–333.
- [Chung, 2019] Chung, M. K. (2019). *Brain Network Analysis*. Cambridge University Press.
- [Chung et al., 2017] Chung, M. K., Hanson, J. L., Adluru, N., Alexander, A. L., Davidson, R. J., and Pollak, S. D. (2017). Integrative Structural Brain Network Analysis in Diffusion Tensor Imaging. *Brain Connect.*, *7(6)*:331–346.
- [Cook et al., 2019] Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., Nguyen, K. C., Tang, L. T., Bayer, E. A., Duerr, J. S., Bülow, H. E., Hobert, O., Hall, D. H., and Emmons, S. W. (2019). Whole-animal connectomes of both *Caenorhabditis elegans* sexes. *Nature*, *571(7763)*:63–71.
- [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-Vector Networks. *Mach. Learn.*, *20(3)*:273–297.
- [Craddock et al., 2013] Craddock, R. C., Jbabdi, S., Yan, C. G., Vogelstein, J. T., Castellanos, F. X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., and Milham, M. P. (2013). Imaging human connectomes at the macroscale. *Nat. Methods*, *10(6)*:524–539.
- [Dale et al., 1999] Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction.
- [DeFelipe, 2010] DeFelipe, J. (2010). From the connectome to the synaptome: An epic love story. *Science (80-.)*, *330(6008)*:1198–1201.
- [Desikan et al., 2006] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.
- [Divakar, 2023] Divakar (2023). Support Vector Machine (SVM).
- [Fischl et al., 2002] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33(3)*:341–355.

- [Fischl et al., 2004] Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cortex*, 14:11–22.
- [Fonov et al., 2011] Fonov, V., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., and Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage*, 54(1):313–327.
- [Fornito et al., 2013] Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *Neuroimage*, 80:426–444.
- [Freitas et al., 2019] Freitas, C. G., Aquino, A. L., Ramos, H. S., Frery, A. C., and Rosso, O. A. (2019). A detailed characterization of complex networks using Information Theory. *Sci. Reports 2019 91*, 9(1):1–12.
- [Georgieff, 2007] Georgieff, M. K. (2007). Nutrition and the developing brain: nutrient priorities and measurement. *Am. J. Clin. Nutr.*, 85(2).
- [Gutiérrez-Gómez et al., 2019] Gutiérrez-Gómez, L., Vohryzek, J., Chiêm, B., Baumann, P. S., Conus, P., Cuenod, K. D., Hagmann, P., and Delvenne, J. C. (2019). Stable biomarker identification for predicting schizophrenia in the human connectome. *NeuroImage Clin.*, 27(July 2019):102316.
- [Hangmann, 2005] Hangmann, P. (2005). *From diffusion MRI to brain connectomics*. PhD thesis.
- [Jenkinson and Chappell, 2018] Jenkinson, M. and Chappell, M. (2018). *Introduction to Neuroimaging Analysis*, volume 1.
- [Johnsonbaugh, 2005] Johnsonbaugh, R. . (2005). Teoría de Gráficas. In Education, P., editor, *Matemáticas Discret.*, chapter 8. México, 6 edition.
- [Joshi et al., 2012] Joshi, A. A., Shattuck, D. W., and Leahy, R. M. (2012). A Method for Automated Cortical Surface Registration and Labeling. *Biomed. image Regist. ... proceedings. WBIR (Workshop 2006-)*, author, 7359:180.
- [Kar et al., 2008] Kar, B. R., Rao, S. L., and Chandramouli, B. A. (2008). Cognitive development in children with chronic protein energy malnutrition. *Behav. Brain Funct.*, 4(1):1–12.

- [Klein and Tourville, 2012] Klein, A. and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.*, 0(DEC):171.
- [Kuncheva et al., 2020] Kuncheva, L. I., Matthews, C. E., Arnaiz-González, Á., and Rodríguez, J. J. (2020). Feature Selection from High-Dimensional Data with Very Low Sample Size: A Cautionary Tale. pages 1–25.
- [Larivière et al., 2019] Larivière, S., Vos De Wael, R., Paquola, C., Hong, S. J., Mišić, B., Bernasconi, N., Bernasconi, A., Bonilha, L., and Bernhardt, B. C. (2019). Microstructure-informed connectomics: Enriching large-scale descriptions of healthy and diseased brains. *Brain Connect.*, 9(2):113–127.
- [Larson-Nath and Goday, 2019] Larson-Nath, C. and Goday, P. (2019). Malnutrition in Children With Chronic Disease. *Nutr. Clin. Pract.*, 34(3):349–358.
- [Le Bihan et al., 2001] Le Bihan, D., Mangin, J. F., Poupon, C., Clark, C. A., Pappata, S., Molko, N., and Chabriat, H. (2001). Diffusion tensor imaging: Concepts and applications. *J. Magn. Reson. Imaging*, 13(4):534–546.
- [Lelijveld et al., 2019] Lelijveld, N., Jalloh, A. A., Kampondeni, S. D., Seal, A., Wells, J. C., Goyheneix, M., Chimwezi, E., Mallewa, M., Nyirenda, M. J., Heyderman, R. S., and Kerac, M. (2019). Brain MRI and cognitive function seven years after surviving an episode of severe acute malnutrition in a cohort of Malawian children. *Public Health Nutr.*, 22(8):1406–1414.
- [Lichtman and Sanes, 2008] Lichtman, J. W. and Sanes, J. R. (2008). Ome sweet ome: what can the genome tell us about the connectome? *Curr. Opin. Neurobiol.*, 18(3):346–353.
- [Livingston et al., 2020] Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., Orgeta, V., Ritchie, K., Rockwood, K., Sampson, E. L., Samus, Q., Schneider, L. S., Selbæk, G., Teri, L., and Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*, 396(10248):413–446.
- [Logothetis, 2008] Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878.

- [Maestú; et al., 2008] Maestú, F., Ríos Lago, A., and Alonso, Raúl, C. (2008). *Neuroimagen: técnicas y procesos cognitivos*.
- [Maier-Hein et al., 2017] Maier-Hein, K. H., Neher, P. F., Houde, J. C., Côté, M. A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F. C., Lin, Y. C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., Renjie, H., Li, Q., Westin, C. F., Deslauriers-Gauthier, S., González, J. O. O., Paquette, M., St-Jean, S., Girard, G., Rheault, F., Sidhu, J., Tax, C. M., Guo, F., Mesri, H. Y., Dávid, S., Froeling, M., Heemskerk, A. M., Leemans, A., Boré, A., Pinsard, B., Bedetti, C., Desrosiers, M., Brambati, S., Doyon, J., Sarica, A., Vasta, R., Cerasa, A., Quattrone, A., Yeatman, J., Khan, A. R., Hodges, W., Alexander, S., Romascano, D., Barakovic, M., Auría, A., Esteban, O., Lemkaddem, A., Thiran, J. P., Cetingul, H. E., Odry, B. L., Mailhe, B., Nadar, M. S., Pizzagalli, F., Prasad, G., Villalon-Reina, J. E., Galvis, J., Thompson, P. M., Requejo, F. D. S., Laguna, P. L., Lacerda, L. M., Barrett, R., Dell'Acqua, F., Catani, M., Petit, L., Caruyer, E., Daducci, A., Dyrby, T. B., Holland-Letz, T., Hilgetag, C. C., Stieltjes, B., and Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.*, 8(1).
- [Maniega et al., 2007] Maniega, S. M., Bastin, M. E., and Armitage, P. A. (2007). A quantitative comparison of two methods to correct eddy current-induced distortions in DT-MRI. *Magn. Reson. Imaging*, 25(3):341–349.
- [Manjón et al., 2010] Manjón, J. V., Coupé, P., Martí-Bonmatí, L., Collins, D. L., and Robles, M. (2010). Adaptive non-local means denoising of MR images with spatially varying noise levels. *J. Magn. Reson. Imaging*, 31(1):192–203.
- [Margulies et al., 2013] Margulies, D. S., Böttger, J., Watanabe, A., and Gorgolewski, K. J. (2013). Visualizing the human connectome. *Neuroimage*, 80:445–461.
- [Martín Nogueroles and Martínez Barbero, 2017] Martín Nogueroles, T. and Martínez Barbero, J. P. (2017). Advanced diffusion MRI and biomarkers in the central nervous system: a new approach. *Radiologia*, 59(4):273–285.
- [Meoded et al., 2020] Meoded, A., Goldenberg, N. A., and Huisman, T. A. (2020). Structural Connectomics: State of the Art and Applications in Pediatric Neurodevelopmental Disorders, Neuro-Oncology, and Arterial Ischemic Stroke. *J. Pediatr.*, 221:S37–S42.

- [Meoded et al., 2017] Meoded, A., Huisman, T. A., Casamassima, M. G. S., Jallo, G. I., and Poretti, A. (2017). The structural connectome in children: basic concepts, how to build it, and synopsis of challenges for the developing pediatric brain. *Neuroradiology*, 59(5):445–460.
- [Meoded et al., 2015] Meoded, A., Katipally, R., Bosemani, T., Huisman, T. A., and Poretti, A. (2015). Structural connectivity analysis reveals abnormal brain connections in agenesis of the corpus callosum in children. *Eur. Radiol.*, 25(5):1471–1478.
- [Modat et al., 2014] Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., and Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *J. Med. Imaging*, 1(2):024003.
- [Morgan and Lichtman, 2013] Morgan, J. L. and Lichtman, J. W. (2013). Why not connectomics? *Nat. Methods*, 10(6):494–500.
- [Moseley et al., 1990] Moseley, M. E., Cohen, Y., Kucharczyk, J., Mintorovitch, J., Asgari, H. S., Wendland, M. F., Tsuruda, J., and Norman, D. (1990). Diffusion-weighted MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology*, 176(2):439–445.
- [Moura et al., 2019] Moura, L. M., Luccas, R., De Paiva, J. P., Amaro, E., Leemans, A., Leite, C. d. C., Otaduy, M. C., and Conforto, A. B. (2019). Diffusion tensor imaging biomarkers to predict motor outcomes in stroke: A narrative review.
- [Pandya et al., 2017] Pandya, S., Kuceyeski, A., and Raj, A. (2017). The Brain’s Structural Connectome Mediates the Relationship between Regional Neuroimaging Biomarkers in Alzheimer’s Disease. *J. Alzheimers. Dis.*, 55(4):1639–1657.
- [Payabvash et al., 2019] Payabvash, S., Palacios, E. M., Owen, J. P., Wang, M. B., Tavassoli, T., Gerdes, M., Brandes-Aitken, A., Cuneo, D., Marco, E. J., and Mukherjee, P. (2019). White matter connectome edge density in children with autism spectrum disorders: Potential imaging biomarkers using machine-learning models. *Brain Connect.*, 9(2):209–220.
- [Scheffer and Xu, 2020] Scheffer, L. K. and Xu, C. S. (2020). A connectome and analysis of the adult drosophila central brain. *Elife*, 9:1–74.
- [Seung, 2012] Seung, S. (2012). *Connectome: How the Brain’s Wiring Make Us Who We Are*, volume 50. Mariner Books.

- [Shao et al., 2012] Shao, J., Yang, Q., Wohlschläger, A., and Sorg, C. (2012). Discovering aberrant patterns of human connectome in Alzheimer’s disease via sub-graph mining. *Proc. - 12th IEEE Int. Conf. Data Min. Work. ICDMW 2012*, pages 86–93.
- [Sotiropoulos et al., 2013] Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., Glasser, M. F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D. A., Yacoub, E., Lenglet, C., Van Essen, D. C., Ugurbil, K., and Behrens, T. E. (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *Neuroimage*, 80:125–143.
- [Sporns, 2018] Sporns, O. (2018). *Networks of the Brain*. MIT Press.
- [Sporns et al., 2005] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Comput. Biol.*, 1(4):0245–0251.
- [Sun et al., 2015] Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., Li, K., Mok, V. C., Chu, W. C., and Wang, D. (2015). Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *Biomed. Eng. Online*, 14(1):1–17.
- [Tax et al., 2022] Tax, C. M., Bastiani, M., Veraart, J., Garyfallidis, E., and Okan Irfanoglu, M. (2022). What’s new and what’s next in diffusion MRI preprocessing. *Neuroimage*, 249(December 2021):118830.
- [Timm et al., 2009] Timm, F., Klement, S., Martinetz, T., and Barth, E. (2009). Welding inspection using novel specularly features and a one-class SVM. *VISAPP 2009 - Proc. 4th Int. Conf. Comput. Vis. Theory Appl.*, 1(January):146–153.
- [Tony Yiu, 2019] Tony Yiu (2019). Understanding Random Forest. How the Algorithm Works and Why it Is... | by Tony Yiu | Towards Data Science.
- [Tournier et al., 2011] Tournier, J. D., Mori, S., and Leemans, A. (2011). Diffusion tensor imaging and beyond.
- [Vanderah and Nolte, 2018] Vanderah, T. W. and Nolte, J. (2018). *Nolte’s the human brain in photographs and diagrams*.
- [Walker et al., 2007] Walker, S. P., Wachs, T. D., Meeks Gardner, J., Lozoff, B., Wasserman, G. A., Pollitt, E., and Carter, J. A. (2007). Child development: risk factors for adverse outcomes in developing countries. *Lancet (London, England)*, 369(9556):145–157.

-
- [Wang et al., 2019] Wang, Y., Xu, C., Park, J. H., Lee, S., Stern, Y., Yoo, S., Kim, J. H., Kim, H. S., and Cha, J. (2019). Diagnosis and prognosis of Alzheimer’s disease using brain morphometry and white matter connectomes. *NeuroImage Clin.*, 23(September 2018):101859.
- [Widjaja et al., 2015] Widjaja, E., Zamyadi, M., Raybaud, C., Snead, O. C., Doesburg, S. M., and Smith, M. L. (2015). Disrupted Global and Regional Structural Networks and Subnetworks in Children with Localization-Related Epilepsy. *AJNR Am. J. Neuroradiol.*, 36(7):1362.
- [Wilcox, 2016] Wilcox, R. R. (2016). *Understanding and applying basic statistical methods using R*.
- [Xie et al., 2020] Xie, L., Li, Z., Zhou, Y., He, Y., and Zhu, J. (2020). Computational diagnostic techniques for electrocardiogram signal analysis. *Sensors (Switzerland)*, 20(21):1–32.

APÉNDICE A

CONECTOGRAMA CON ENFOQUE EN GRAFOS COMPLETOS POR CARACTERÍSTICAS SIMILARES

A.1. Construcción del conectoma

Para construir el conectoma bajo este enfoque fue necesario definir qué atributos se requieren mostrar del conectograma que permitan diferenciar un grafo proveniente de un sujeto del grupo control del de uno del grupo desnutrición. Para ello, el objetivo constará en caracterizar el conectograma de los sujetos del grupo Desnutrición, que permita distinguirlos en este caso, del grupo Control, esto por medio del grado de conectividad, el cual mide cuántas conexiones en total tienen todos los vértices del grafo, que en nuestro conectograma son regiones.

En la construcción del conectoma individual para cada sujeto, para considerar una unión de vértices por medio de una arista, las regiones que representan tienen que tener en común el mismo descriptor; y para considerar la región (vértice), este tiene que tener un valor dentro del rango de valores según las funciones de probabilidad de los KDE de la característica a evaluar. Para representar dichas uniones y dado que cada descriptor puede ser del tipo morfométrico, difusión o de red, el color de los arcos o aristas del conectograma serán rojo, amarillo y azul,

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS COMPLETOS POR CARACTERÍSTICAS SIMILARES

respectivamente.

El enfoque estimación por densidad de kernel (KDE), nos permite discretizar si se considera o no un vértice en el conectograma, donde el valor será clasificado como 0 si pertenece al grupo Control y el 1 si pertenece al de Desnutrición. Véase en la Figura A.1, el estudio por KDE realizado a una característica en específico, los valores inferiores a 0.6, la mayor densidad o probabilidad, la tiene el grupo control; y a partir de ese mismo valor, se considera que esa característica forma parte del grupo de desnutrición, por lo que se considera en el total de regiones que la comparten.

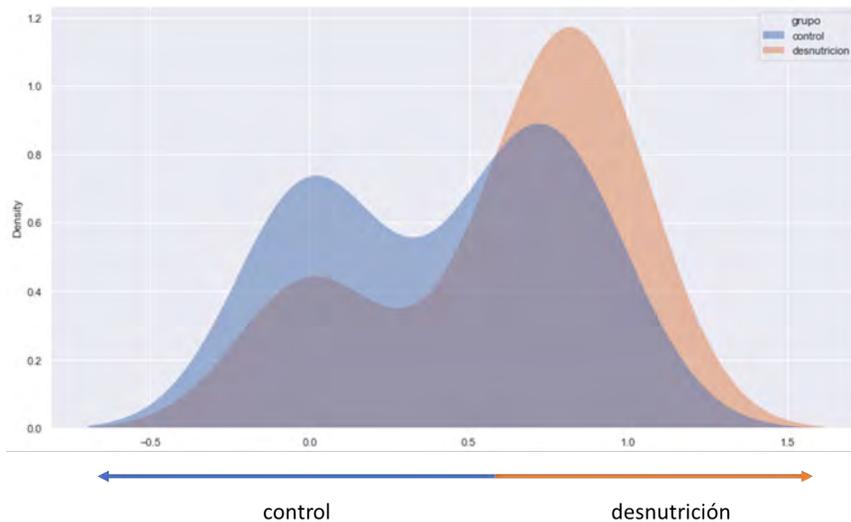


Figura A.1: KDE de dos grupos para la asignación del valor más probable.

Veamos el siguiente ejemplo, se tienen las siguientes 10 características seleccionadas:

descriptor1 + region1	descriptor1 + region2
descriptor1 + region3	descriptor2 + region1
descriptor2 + region3	descriptor3 + region2
descriptor3 + region4	descriptor3 + region5
descriptor4 + region3	descriptor5 + region4

Como se observan en las características, se tienen 5 descriptores y 5 regiones, para fines explicativos los descriptores 1 y 2 serán de tipo morfométrico; los descriptores 3 y 4 serán de difusión y el descriptor 5 será de redes enfocadas en vértices. Dado que las regiones 1, 2 y 3 comparten el descriptor 1, entre ellas se forma un grafo completo, como se observa en la Tabla A.1 y en la Figura A.2, dado que solo estamos considerando el grafo formado por una característica, el color de los

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS
COMPLETOS POR CARACTERÍSTICAS SIMILARES

arcos es negro.

Tabla A.1: Tabla para formar el grafo completo de la conexión del descriptor1 con las regiones 1,2 y 3.

VÉRTICE 1	VÉRTICE 2	Descriptor	Tipo de Descriptor	Color
region1	region2	descriptor1	morfométrico	rojo
region1	region2	descriptor1	morfométrico	rojo
region2	region3	descriptor1	morfométrico	rojo

Haciendo los mismos pasos que se realizaron para del descriptor1 en todos los demás obtenemos la Tabla A.1 y su respectivo conectograma en Figura A.2b.

Tabla A.2: Tabla de ejemplo con el que se construyó el conectograma de la Figura A.2

VÉRTICE 1	VÉRTICE 2	Descriptor	Tipo de Descriptor	Color
region1	region2	descriptor1	morfométrico	rojo
region1	region2	descriptor1	morfométrico	rojo
region2	region3	descriptor1	morfométrico	rojo
region1	region2	descriptor2	morfométrico	rojo
region2	region4	descriptor3	difusión	amarillo
region2	region5	descriptor3	difusión	amarillo
region4	region5	descriptor3	difusión	amarillo
region3	region3	descriptor4	difusión	amarillo
region4	region4	descriptor5	red	azul

Mientras que en la Figura A.2a observamos un conectograma cuyos vértices (r_1 , r_2 y r_3) se encuentran en el perímetro del círculo y cómo están unidos mediante arcos que representan las aristas del grafo. Dado que en este solo existe un tipo de características, los arcos son de un solo color. Por otro lado, la Figura A.2 muestra el conectograma de las regiones seleccionadas y cuya tabla (Tabla A.1), muestra también los colores que representan los tipos de descriptores.

Cabe destacar que aquellas regiones cuyo descriptor no lo comparten con ningún otra región, se considerará como una conexión consigo misma. Como podemos ver, la región 3 con el descriptor 4 y también la región 4 con el descriptor 5 en la Tabla A.1.

A.2. Visualización del conectoma

Para la obtención del conectoma fue necesario implementar un sistema donde sea posible identificar con la vista el resultado de un sujeto control al de uno desnutrición. Por ende, se optó por el uso del conectograma, el cual es un grafo

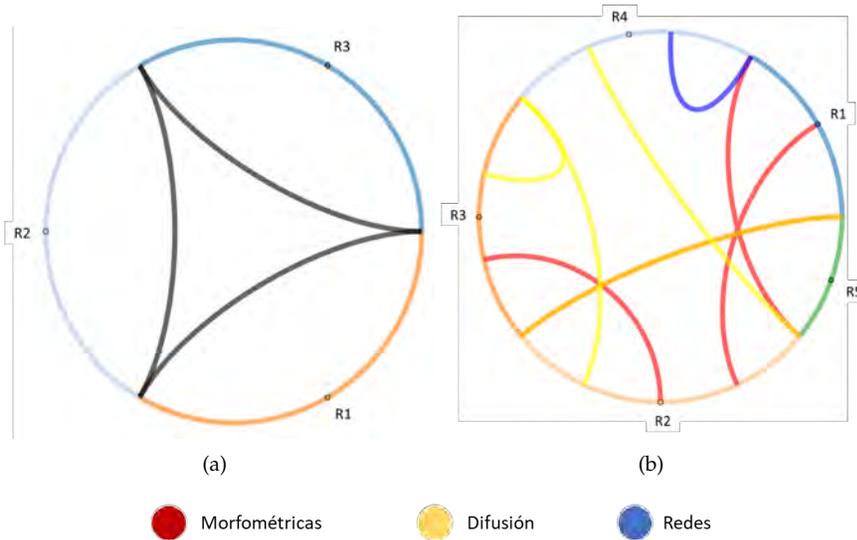


Figura A.2: Ejemplos de conectogramas (a) con una sola característica para 3 regiones y (b) con tres tipos de características de 5 regiones. En la figura (a) se muestran los tres vértices representados por una sección del perímetro del círculo equitativamente y sus tres conexiones de la misma clase. En la figura (b) se muestra como las secciones de los vértices varía dependiendo su grado, también destaca el color de las conexiones, ya que en este se tienen tres tipos de características.

circular cuyos vértices forman parte del perímetro del círculo y los arcos del círculo representan las aristas que unen un vértice con otro o con sí mismo. Dado que se utilizó la distancia entre las medianas de las distintas poblaciones, se consideró una conexión entre regiones cuando comparten un mismo biomarcador siempre y cuando dichas características se encuentren con una función de probabilidad más cercano al valor del grupo desnutrición que del grupo control, y se coloreó de rojo aquellas aristas que representan las características morfométricas, amarillo las que representan las características de difusión y de azul las de redes.

Sujetos control

Como podemos ver en la Figura A.3, tenemos un conectograma formado por 9 regiones, donde las características que predominan son las morfométricas, habiendo una única región con características de difusión correspondiente a la población Desnutrición, así como una sola región con varios biomarcadores de redes.

Sujetos desnutrición

A diferencia de los sujetos control, el número de descriptores para cada región genera una visualización con mayor grado de conectividad (número de arcos en

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS
COMPLETOS POR CARACTERÍSTICAS SIMILARES

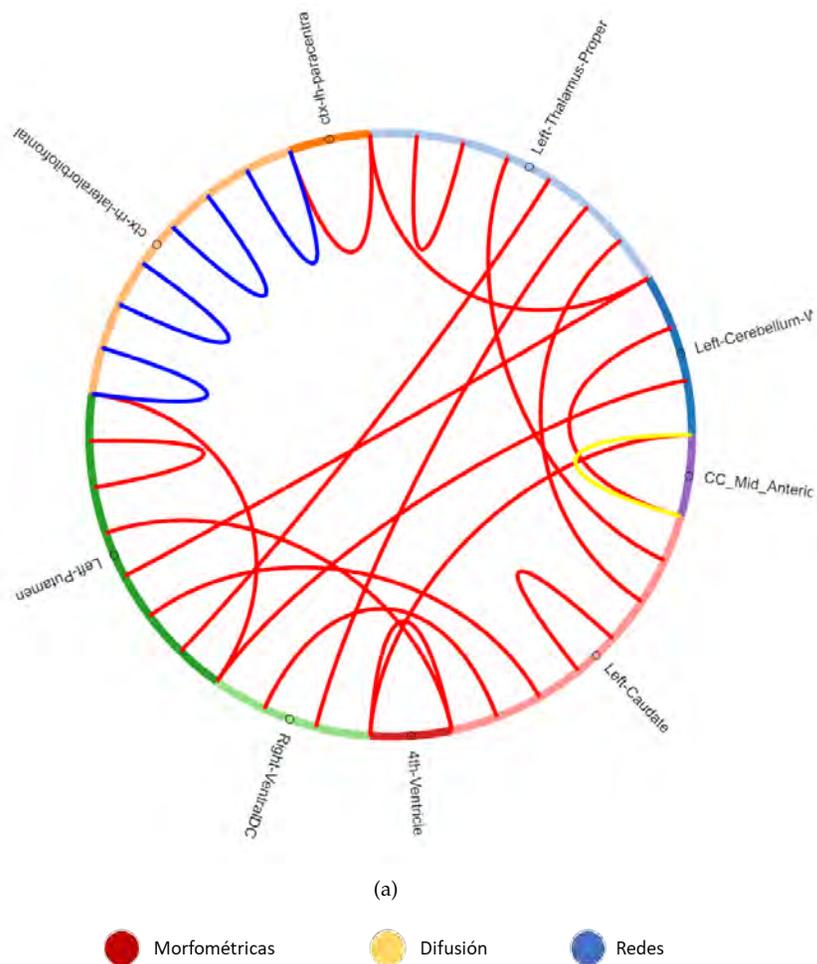


Figura A.3: Conectograma de sujeto Control 1 utilizando el enfoque de grafos completos por características en común. Aquí se observa una mayor densidad en las conexiones entre regiones con características morfométricas en común.

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS
COMPLETOS POR CARACTERÍSTICAS SIMILARES

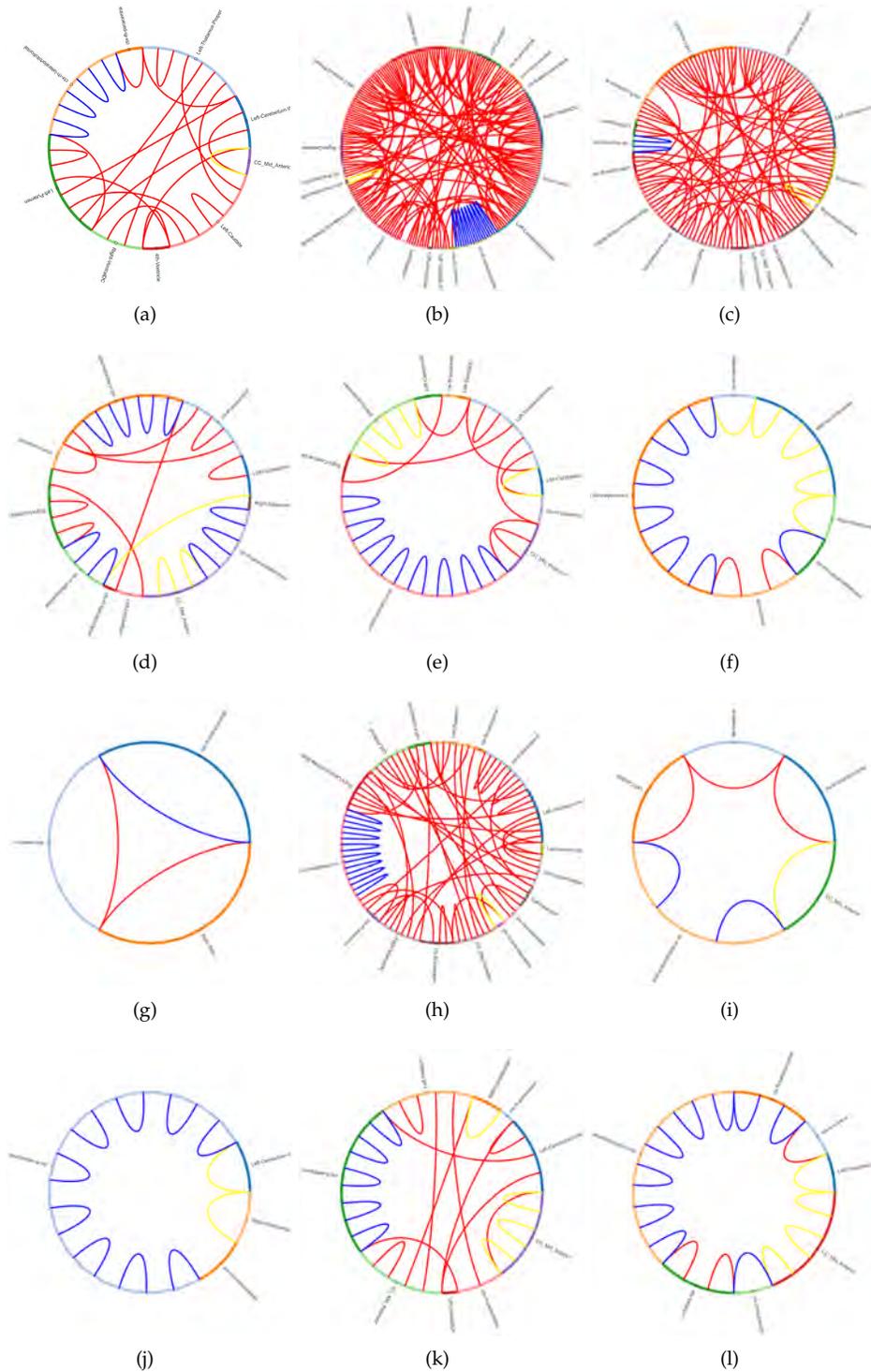


Figura A.4: Resultados de los conectogramas de sujetos control. Las figuras (b) y (c) se distinguen del resto de la población, dado que su grado de conectividad es mayor que el resto. Esto podría significar que los sujetos cuyos conectogramas son dichas figuras están dentro de los rangos esperados en un sujeto con desnutrición.

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS COMPLETOS POR CARACTERÍSTICAS SIMILARES

total), por lo que al ver la Figura A.5, y compararla con el conectograma de la Figura 5.13, podemos observar una diferencia tanto en el número de regiones como en la cantidad de arcos en la visualización.

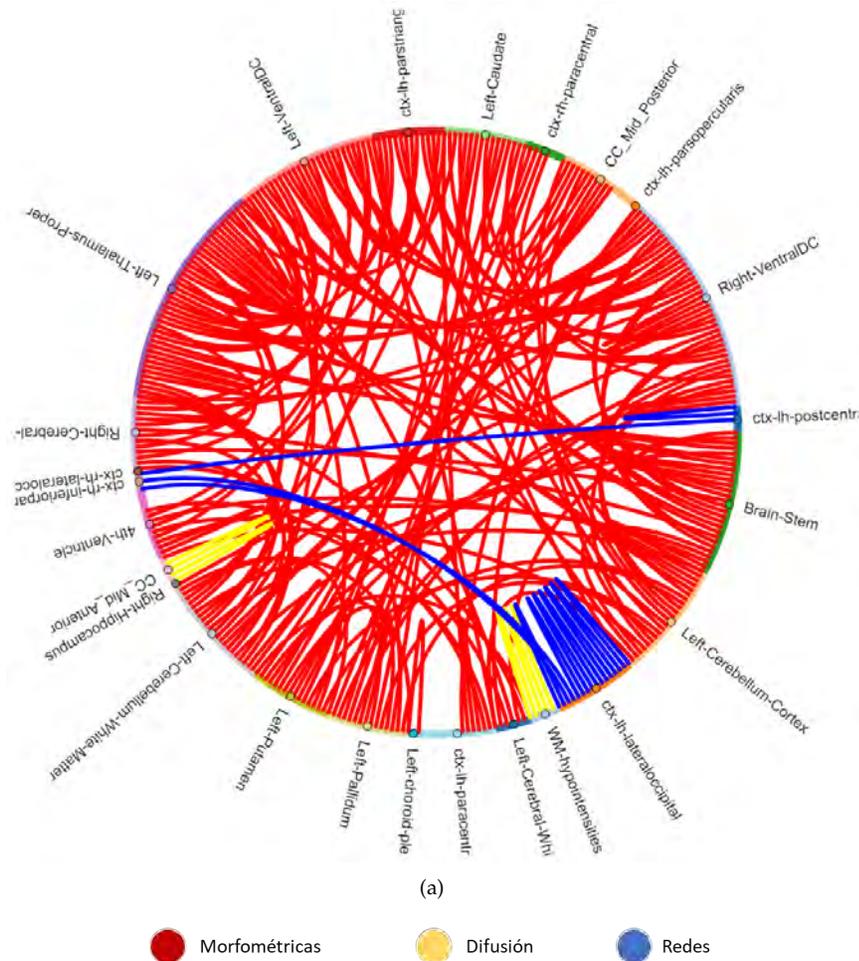


Figura A.5: Resultado del conectograma del sujeto 1 del grupo Desnutrición. En comparación con el sujeto 1 del grupo control, la densidad de conexiones es mayor, así como el número de vértices que lo conforman. La seguridad con la que se confirma que efectivamente se trata de un sujeto del grupo desnutrición es proporcional a la conectividad máxima, es decir, el grafo completamente conexo y con todas las regiones.

En la Figura A.4, podemos observar todos los conectogramas de los sujetos control, a simple vista los conectogramas b, c y h, denotan un grado de conectividad mayor, esto debido a que la cantidad de características que tienen más parecido a las de un sujeto control es mayor que la que no, y así como es notorio en los sujetos control, también son notorios los sujetos Desnutrición cuyo grado de conectividad es menor que el resto.

APÉNDICE A. CONECTOGRAMA CON ENFOQUE EN GRAFOS
COMPLETOS POR CARACTERÍSTICAS SIMILARES

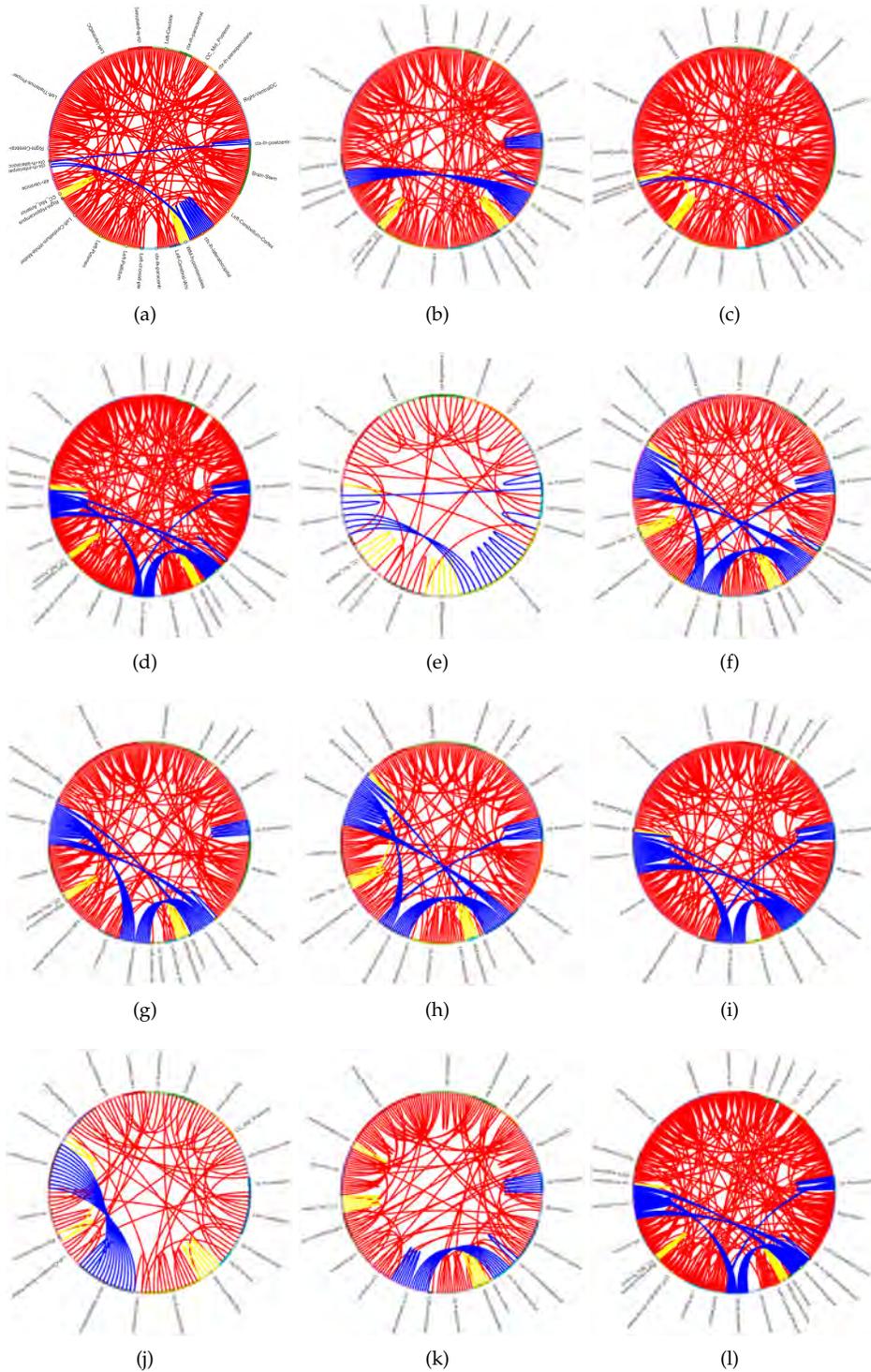


Figura A.6: Resultados de los conectogramas de sujetos desnutrición. Las figuras (e) y (j) se distinguen del resto de la población, dado que su grado de conectividad es menor que el resto. Esto podría significar que los sujetos cuyos conectogramas son dichas figuras no están del todo dentro de los rangos esperados en un sujeto con desnutrición.