



POLITÉCNICA



UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica Industrial y Automática

TRABAJO FIN DE GRADO

PREDICCIÓN DE LA CONVERSIÓN DE SUJETOS CON DETERIORO COGNITIVO LEVE A ENFERMEDAD DE ALZHEIMER MEDIANTE MARCADORES LONGITUDINALES DE IMÁGENES DE RESONANCIA MAGNÉTICA Y MEDIDAS COGNITIVAS

Javier Rojo Muñoz

Tutor: Carlos Platero Dueñas

Departamento de Ingeniería
Eléctrica, Electrónica,
Automática y Física Aplicada

Madrid, junio, 2018



POLITÉCNICA



UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Y
DISEÑO INDUSTRIAL

Grado en Ingeniería Electrónica Industrial y Automática

TRABAJO FIN DE GRADO

PREDICCIÓN DE LA CONVERSIÓN DE
SUJETOS CON DETERIORO COGNITIVO
LEVE A ENFERMEDAD DE ALZHEIMER
MEDIANTE MARCADORES
LONGITUDINALES DE IMÁGENES DE
RESONANCIA MAGNÉTICA Y MEDIDAS
COGNITIVAS

Firma Autor

Firma Tutor

Copyright ©2018. Javier Rojo Muñoz

Esta obra está licenciada bajo la licencia Creative Commons

Atribución-NoComercial-SinDerivadas 3.0 Unported (CC BY-NC-ND 3.0). Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.es> o envíe una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, EE.UU.

Todas las opiniones aquí expresadas son del autor, y no reflejan necesariamente las opiniones de la Universidad Politécnica de Madrid.

Título: Predicción de la conversión de sujetos con deterioro cognitivo leve a enfermedad de Alzheimer mediante marcadores longitudinales de imágenes de resonancia magnética y medidas cognitivas.

Autor: Javier Rojo Muñoz

Tutor: Carlos Platero Dueñas

Cotutor: No procede

EL TRIBUNAL

Presidente:

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Trabajo Fin de Grado el día de de en, en la Escuela Técnica Superior de Ingeniería y Diseño Industrial de la Universidad Politécnica de Madrid, acuerda otorgarle la CALIFICACIÓN de:

VOCAL

SECRETARIO

PRESIDENTE

Agradecimientos

Agradezco a mi tutor su infinita paciencia, su interés y sus inestimables sugerencias y orientaciones.

A mis padres, por su apoyo incondicional y todo lo que me han enseñado. Siguen siendo mis héroes y sin ellos no habría llegado hasta aquí. Les agradezco su empeño en que tanto mi hermana como yo aprendiésemos idiomas y los valores con los que nos han educado. Agradezco a mi hermana el estar siempre dispuesta a escuchar y enseñarme con su propio ejemplo a no rendirme ante nada.

A todos mis amigos de la universidad, con los que he compartido tantas experiencias y buenos momentos. Hemos sufrido y disfrutado juntos y es, sin duda, un período de mi vida que jamás olvidaré. A todos los buenos profesores que me contagiaron la pasión por aprender y me transmitieron la ilusión de seguir creciendo.

Gracias a todos los que directa o indirectamente han dejado algo de sí en mí, a quienes me han apoyado en los momentos más bajos y no han dejado que abandonase.

A todos los aquí mencionados y a todas las personas que no he citado expresamente,

gracias.

Resumen

El Alzheimer es una enfermedad neuro-degenerativa que se encuentra tras el 60 %-70 % de los casos de demencia, según un informe de la Organización Mundial de la Salud [6]. La atrofia sufrida en las distintas estructuras corticales y subcorticales se puede observar mediante neuroimagen. En este estudio se utilizarán imágenes de resonancia magnética de 1.5 Teslas (MRI-T1.5) para establecer biomarcadores en la evolución del grosor del córtex que permitan discernir entre sujetos con deterioro cognitivo leve (MCI) que progresan al Alzheimer (pMCI) y que se mantienen estables (sMCI).

MRI-T1.5 hace referencia a una Imagen de Resonancia Magnética que utiliza un campo magnético de 1.5 Tesla. Se trata de una técnica y una fuerza habitual en los centros especializados.

MCI hace referencia a sujetos que padecen Deterioro Cognitivo Leve (Mild Cognitive Impairment, en inglés). Se trata de sujetos que muestran indicios que apuntan al desarrollo de la enfermedad. Sin embargo, en este grupo heterogéneo se pueden distinguir casos que progresan hacia el Alzheimer (pMCI) y casos en los que el sujeto se mantiene estable (sMCI).

Palabras clave: Alzheimer, MRI, espesor cortical, MCI, estudio longitudinal.

Abstract

Alzheimer is a neuro-degenerative disease, cause of around 60 %-70 % of dementia cases, according to a World Health Organization's report [6]. Atrophy suffered in the different cortical and subcortical structures can be detected using neuroimage. In this study, images coming from magnetic resonances of 1.5 T (MRI-T1.5) will be used in order to establish biomarkers which are able to differentiate between subjects with mild cognitive impairment (MCI) that progress to Alzheimer Disease (pMCI) and those that stay stable (sMCI).

MRI-T1.5 stands for a Magnetic Resonance Image using a magnetic field with a strength of 1.5 Tesla. This technique and this strength are commonly used in specialized centers.

MCI stands for Mild cognitive impairment. These are subjects showing signs pointing to the development of the disease. However, in this heterogeneous group two main types of subjects can be found: pMCI (progressive MCI), used for subjects who finally develop the disease, and sMCI (stable MCI), used for subjects who doesn't.

Keywords: Alzheimer, MRI, cortical thickness, MCI, longitudinal study.

Índice general

Agradecimientos	VII
Resumen	IX
Abstract	XI
Índice	XIII
Índice de figuras	XVII
Índice de tablas	XIX
1. Introducción	1
1.1. Motivación del proyecto	1
1.1.1. Introducción a la enfermedad de Alzheimer	1
1.1.2. Naturaleza del presente estudio	3
1.2. Objetivos	4
1.3. Estructura del documento	5
2. Estado de la técnica	7
2.1. Introducción del capítulo	7
2.2. El problema del diagnóstico	7
2.3. Trabajos previos	8
2.3.1. Querbes2009	8
2.3.2. Li2012	9
2.3.3. Eskilsen2015	10
2.3.4. Korolev2016	10
2.3.5. Rathore2017	13
2.3.6. Landin2017	14
2.3.7. Sorensen2017	15
2.4. Conclusión del capítulo	16
3. Materiales y herramientas	17
3.1. Introducción del capítulo	17
3.2. Datos sobre la población estudiada	17
3.3. ADNIMERGE y tests neuropsicológicos	18
3.4. ADNI	19
3.5. MATLAB	21
3.6. FreeSurfer	22

3.7. Conclusión del capítulo	25
4. Métodos empleados	27
4.1. Introducción del capítulo	27
4.2. Extracción de datos	27
4.2.1. Adquisición de datos desde ADNI	27
4.2.2. Flujo de trabajo general en la extracción de características	29
4.2.3. getNIIFiles	30
4.2.4. FS_recon_all_cross	31
4.2.5. FS_recon_all-base	41
4.2.6. FS_recon_all-long	43
4.2.7. getAsegAparcADNI	46
4.2.8. asegAparc2table	49
4.2.9. Control de calidad	50
4.2.10. Análisis masivo	54
4.3. Análisis estadístico	65
4.3.1. Algoritmo de selección mRMR	65
4.3.2. Análisis longitudinal y modelo lineal de efectos mixtos	67
4.3.3. Modelo de K-Folds anidados	72
5. Resultados	79
5.1. Introducción del capítulo	79
5.2. Experimentos	79
5.2.1. Obtención de los datos de la población	79
5.2.2. Cálculo de la dimensión óptima	80
5.2.3. Cálculo de las características más repetidas	80
5.2.4. Cálculo de otros efectos fijos	81
5.3. Resultados con la edad como efecto fijo	81
5.3.1. Características más repetidas	81
5.3.2. Dimensión óptima	94
5.3.3. Mejora de la predicción con todas las visitas	101
5.4. Resultados con la edad y APOE4 como efectos fijos	102
5.4.1. Dimensión óptima	102
5.5. Resultados con la edad y la educación como efectos fijos	103
5.5.1. Dimensión óptima	103
5.6. Resultados con la edad, la educación y APOE4 como efectos fijos	104
5.6.1. Dimensión óptima	104
5.7. Comparativa entre distintas matrices de efectos fijos	105
5.8. Comparativa con otros estudios	105
5.9. Discusión	106
6. Conclusiones	107
6.1. Conclusión	107
6.2. Desarrollos futuros	108
Bibliografía	109
Anexos	113

A. Conceptos básicos de estadística	115
A.1. Introducción al contraste de hipótesis	115
A.1.1. Hipótesis nula e hipótesis alternativa	115
A.1.2. Test de hipótesis	115
A.1.3. Tipos de errores	117
A.2. La comparativa múltiple: FDR y q-valores	119
A.3. Tipos de grupos en algoritmos de aprendizaje	121
B. Adquisición y procesamiento de imágenes MRI	125
B.1. Imágenes de resonancia magnética	125
B.2. Procesamiento de imágenes	126
C. Diccionario de siglas	129
C.1. Ordenación alfabética	129
C.2. Ordenación temática	131
C.2.1. Diagnóstico	131
C.2.2. Tests neuropsicológicos	131
C.2.3. Segmentación	131
C.2.4. Medidas de error	132
C.2.5. Miscelánea	132

Índice de figuras

1.1.	Evolución de la demencia	1
1.2.	Esquema comparativo entre cerebro sano y AD	2
2.1.	Esquema anatómico de los gyri	10
2.2.	Esquema de la validación k-folds	12
2.3.	Región temporoparietal	13
2.4.	Esquema anatómico del cingulate	15
3.1.	Acceso a los datos de ADNI	20
3.2.	Logotipo de MATLAB	21
3.3.	Logotipo de FreeSurfer	22
3.4.	FreeView	24
4.1.	Pantalla de búsqueda avanzada de ADNI.	28
4.2.	Flujo de trabajo con FreeSurfer (I)	29
4.3.	Flujo de trabajo con FreeSurfer (II)	30
4.4.	Etapas generales de la función recon-all.	32
4.5.	Flujo de trabajo del método de normalización N3.	33
4.6.	Esquema de los componentes de una transformación afín.	34
4.7.	Figura del hipotálamo	36
4.8.	Esquema de la representación inflada del cerebro.	37
4.9.	Esquema de la representación esférica del cerebro.	37
4.10.	Etiquetado a partir del ATLAS	38
4.11.	Primera parte del flujo de trabajo de la función recon-all.	39
4.12.	Segunda parte del flujo de trabajo de la función recon-all.	39
4.13.	Control de errores cross	41
4.14.	Volumen brainmask.mgz	44
4.15.	Esquema de getAsegAparcADNI	47
4.16.	Esquema de AsegAparc2Table	50
4.17.	Control de calidad: atrofia de la corteza entorrinal	51
4.18.	Control de calidad: atrofia del volumen hipocampal	51
4.19.	Control de calidad: histograma de la corteza entorrinal	52
4.20.	Control de calidad: histograma del volumen hipocampal	53
4.21.	Control de calidad: correlación	54
4.22.	Flujo de trabajo del análisis masivo	55
4.23.	Esquema de get_Thickness_mgh	56
4.24.	Esquema de buildMassLmeModels	58
4.25.	Esquema de save_Thickness	60
4.26.	Mapa de significancia sMCI-pMCI	61

4.27. Esquema de entradas y salidas de generate_Clusters	62
4.28. Muestra de los clusters obtenidos del análisis masivo	63
4.29. Esquema del flujo de trabajo de join_Tables	64
4.30. Esquema de mRMR con selección frecuencial	67
4.31. Ejemplo de datos de un estudio longitudinal	68
4.32. Esquema de la implementación LME	71
4.33. Esquema de la validación K-folds original	74
4.34. Esquema de la variante de K-Folds con selección por frecuencia	76
4.35. Esquema de los folds internos	76
4.36. Esquema de la implementación del cálculo paralelo en KFSF	77
5.1. Clusters 1, 9, 10 y 13 del hemisferio izquierdo	87
5.2. Clúster 1 del hemisferio derecho	87
5.3. Dimensión óptima para ROIs (X = edad)	94
5.4. Dimensión óptima para tests (X = edad)	95
5.5. Dimensión óptima para clusters (X = edad)	96
5.6. Dimensión óptima para ROIs y tests (X = edad)	97
5.7. Dimensión óptima para ROIs y clusters (X = edad)	98
5.8. Dimensión óptima para clusters y tests (X = edad)	99
5.9. Dimensión óptima para todas las características (X = edad)	100
5.10. Dimensión óptima para todas las cc. (X = edad+APOE4)	102
5.11. Dimensión óptima para todas las cc. (X = edad+educación)	103
5.12. Dimensión óptima para todas las cc. (X = edad+educación+APOE4)	104
A.1. Esquema de la distribución de la hipótesis nula	116
A.2. Curvas ROC	118
A.3. Curvas ROC ante un grupo reducido de test	119
A.4. Esquema del histograma de p-valores	120
A.5. Ejemplo de p-valor en análisis masivo	121
A.6. Esquema de la problemática del overfitting	122
B.1. Esquema del equipo de resonancia magnética	125
B.2. Planos principales de adquisición	126
B.3. Esquema de giros cerebrales y surcos	127

Índice de tablas

2.1. Índices de deterioro	9
3.1. Tabla con los datos del grupo de estudio	17
3.2. Tabla con los datos de los scans estudiados	17
3.3. Tabla con los datos del grupo procesado en FreeSurfer	18
3.4. Tabla con los datos de los scans procesados	18
4.1. criterios de cladiificación de ADNI	29
5.1. Frecuencia de características partiendo de ROIs (X = edad)	82
5.2. Frecuencia de características partiendo de tests (X = edad)	84
5.3. Frecuencia de características partiendo de clusters (X = edad)	86
5.4. Frecuencia de características partiendo de ROIs y tests (X = edad)	88
5.5. Frecuencia de características partiendo de ROIs y clusters (X = edad)	89
5.6. Frecuencia de características partiendo de clusters y tests (X = edad)	91
5.7. Frecuencia de características con todas las cc. (X = edad)	92
5.8. Mejora de resultados con todas las visitas (X = edad)	101
5.9. Mejor medida de error para X=edad+APOE4	103
5.10. Mejor medida de error para X=edad+educación	104
5.11. Mejor medida de error para X=edad+educación+APOE4	105
5.12. Comparativa de errores según las matrices de efectos fijos	105
5.13. Tabla de resultados	106
A.1. Anexo: Tabla con los tipos de errores	117

Capítulo 1

Introducción

1.1. Motivación del proyecto

1.1.1. Introducción a la enfermedad de Alzheimer

La demencia llegó a afectar a 46 millones de personas en 2015, siendo el Alzheimer la principal causa según un reporte de la *Alzheimer's Disease International* [29]. Se trata de un trastorno neurológico caracterizado por una demencia progresiva, normalmente manifestada en la edad senil [23]. Este deterioro se manifiesta especialmente en forma de pérdida de memoria y cambios de comportamiento [17].

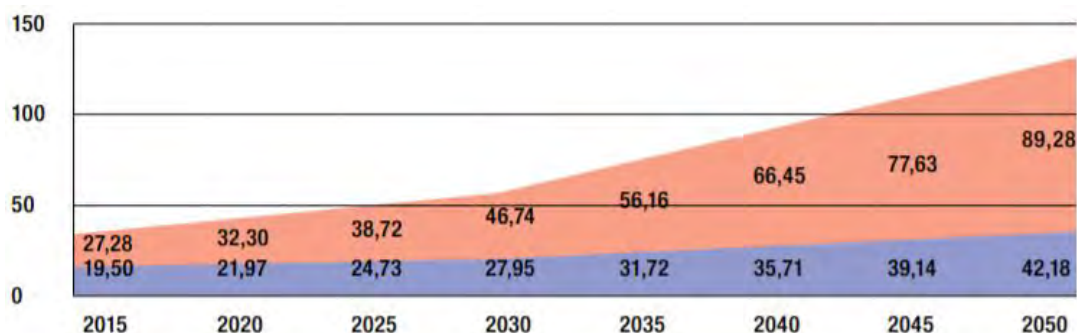


Figura 1.1: Estudio y predicción de la evolución del número de personas que padecen demencia (en millones) en países con rentas altas (en azul) y países con rentas medias y bajas (en rojo), según el informe de la *Alzheimer's Disease International* [29].

Como ya se ha mencionado, el Alzheimer es una enfermedad neurodegenerativa. Esto significa que el área del cuerpo afectada es el cerebro. La enfermedad comienza con la muerte de neuronas, impidiendo así la sinapsis. En etapas tempranas el cerebro es capaz de reestructurarse para compensar las pérdidas. A esta capacidad se le llama plasticidad cerebral. Sin embargo, llega un punto en el que el cerebro no es capaz de contrarrestar la pérdida y los síntomas comienzan a aparecer.

La recomendación de la lectura, la educación y la actividad mental en general como medios para paliar los síntomas se basa en el aumento de la plasticidad cerebral como consecuencia de estas actividades.

La mencionada muerte de neuronas es producida por la acumulación de placas de proteínas β -amiloide y proteínas τ . Estas proteínas interfieren en la sinapsis y en el transporte de nutrientes de las neuronas, lo que provoca el deterioro de las mismas[35].

Respecto a las zonas más afectadas, el hipocampo y la corteza entorrinal suelen ser las primeras en mostrar deterioro, y las más dañadas en etapas tardías. Precisamente el hipocampo está relacionado con numerosos procesos de memoria y aprendizaje. Es comprensible, por tanto, que la pérdida de memoria sea uno de los síntomas más comunes en esta enfermedad.

Debe recordarse, sin embargo, que estas no son las únicas estructuras que se ven afectadas. Además del deterioro natural del cerebro se pueden encontrar otras áreas aparte de las dos expuestas. Para más detalles sobre éstas, véase el capítulo del estado de la técnica, donde se estudian los resultados de distintos trabajos previos.

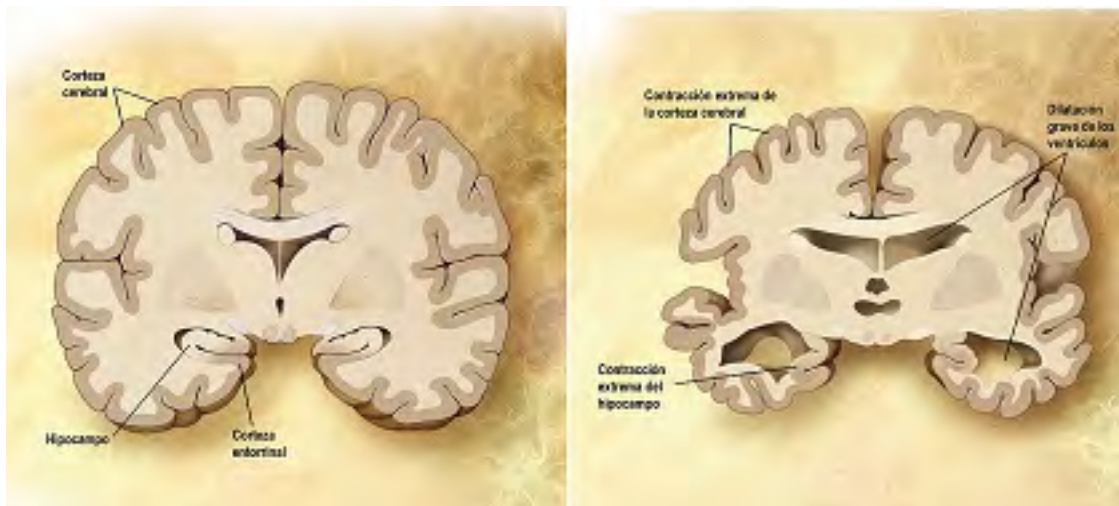


Figura 1.2: Esquema comparativo de un cerebro sano y un sujeto con Alzheimer.

Además, existe un grupo de apolipoproteínas E que incrementan notablemente el riesgo de padecer Alzheimer [33]. Por lo general, las apolipoproteínas se encargan de mover compuestos grasos en los mamíferos y existen tres tipos distintos en los humanos: APOE2, APOE3 y APOE4 [35].

El alelo (tipo) APOE3 es el más común en los humanos (50%-70%) y se encarga de la eliminación de las placas de beta-amiloide, causantes del deterioro neuronal. Sin embargo, el alelo APOE4 parece impedir que esta función se desempeñe con la eficacia necesaria, favoreciendo el desarrollo de la enfermedad. Aunque respecto al total de la población no es muy común (en torno al 10%), se calcula que en torno al 30% ó 40% de los pacientes con Alzheimer poseen este alelo.

1.1.2. Naturaleza del presente estudio

La atrofia producida por el Alzheimer se puede observar mediante el análisis de neuroimagen. Este trabajo presenta un estudio longitudinal, en el que se pretende realizar un diagnóstico automatizado entre los grupos con deterioro cognitivo leve progresivo (pMCI) y deterioro cognitivo leve estable (sMCI). Para ello se obtendrá información a partir de imágenes de resonancia magnética de 1.5 Teslas (MRI-T1.5).

- Un análisis longitudinal consiste en un estudio que tiene en cuenta no sólo el valor de distintas variables sino su evolución en el tiempo. Este concepto se contrapone a un estudio transversal, que sólo tiene en cuenta el valor de las variables en un instante de tiempo dado. Como se verá, los modelos estadísticos son distintos, y ofrece mejores resultados a medida que se dispone de más datos del paciente (más visitas).
- MRI-T1.5 hace referencia a una Imagen de Resonancia Magnética estructural que utiliza un campo magnético de 1.5 Teslas. Se trata de una técnica y una fuerza habitual en los centros especializados que permite obtener la estructura y composición del cráneo. Además, la resonancia magnética no presenta efectos negativos para el paciente, a diferencia de otros métodos de adquisición de imagen. Para una explicación más profunda, véase el anexo B.
- MCI hace referencia a sujetos que padecen Deterioro Cognitivo Leve (Mild Cognitive Impairment en inglés). Se trata de sujetos que se encuentran en un estado transitorio entre el envejecimiento natural y el diagnóstico probable de la enfermedad [27]. Sin embargo, en este grupo heterogéneo se pueden distinguir casos que progresan hacia el Alzheimer (pMCI) y casos en los que el sujeto se mantiene estable (sMCI). Se discuten más detalles sobre esta diferenciación en la sección 4.2.1 del capítulo Métodos.

Respecto a los datos utilizados para predecir el diagnóstico del paciente, se van a utilizar distintas combinaciones y se compararán los resultados en cada caso:

1. Regiones de interés (ROIs): se trata de estructuras delimitadas del cerebro. Esta división ha sido consensuada por la comunidad científica respecto a una serie de criterios de estructura y funcionalidad. Como referencia se suele utilizar algún valor estadístico, que normalmente consiste en el valor medio de la región. Estos datos provienen de la información contenida en las imágenes MRI.
2. Tests neuropsicológicos: se trata de una serie de pruebas estandarizadas de memoria y comportamiento, con los que se adquiere información funcional del paciente. Cada tipo de test se evalúa de una manera distinta y mide parámetros distintos. Los tests evaluados se encuentran en la sección de Materiales.
3. Clusters: se trata de grupos de puntos en la corteza cerebral que han sido obtenidos mediante un análisis masivo (véase la sección correspondiente en Métodos utilizados). Funcionan de una manera similar a las ROIs, sólo que en este caso se corresponden de manera más precisa con zonas significativamente diferentes para los dos grupos (pMCI y sMCI). De nuevo, la información del espesor cortical en cada punto se obtiene a partir de las imágenes MRI.

La importancia de centrarse en este grupo MCI reside en tres aspectos básicos:

1. Por un lado, estos sujetos presentan una acelerada progresión hacia la enfermedad de Alzheimer y la demencia senil [25].
2. Por otro lado, se considera que en esta etapa se pueden aplicar ciertos tratamientos con resultados satisfactorios [28].
3. Por último, esta diferenciación permitiría evaluar la efectividad de un medicamento o tratamiento sin tener que alcanzar la etapa de Alzheimer durante los ensayos clínicos.

1.2. Objetivos

Los objetivos principales que se persiguen en este estudio son los siguientes:

1. Mostrar los resultados de asumir un acercamiento longitudinal basado en un modelo de Efectos Lineales Mixtos (LME). Este modelo se explica con más detalle en el correspondiente apartado de Métodos.
2. Demostrar que un acercamiento longitudinal a la problemática del diagnóstico pMCI-sMCI puede mejorar los resultados respecto a un estudio transversal. Para ello se comparará la predicción, contando sólo con la primera visita (*baseline*), con los resultados del estudio de Eskilsen2015 [10].
3. Estudiar cómo mejora el pronóstico a medida que se dispone de más datos (visitas) del paciente.
4. Estudiar cómo mejoran los resultados al partir de distintos tipos de datos. En concreto: ROIs, tests neuropsicológicos y clusters obtenidos a partir de las diferencias en la atrofia del espesor cortical.
5. Proporcionar una estructura ordenada y documentada del código utilizado para agilizar futuros trabajos. A esto se le suman pequeñas explicaciones sobre el manejo del software empleado.

Además, el presente estudio se ha presentado como una oportunidad personal única de adentrarse en el mundo de la investigación y adquirir una serie de conocimientos transversales: contraste de hipótesis, manejo de LaTeX, soltura en el manejo de MATLAB, etc.

1.3. Estructura del documento

A continuación y para facilitar la lectura del documento, se detalla el contenido de cada capítulo:

- En el **capítulo 1** se realiza una introducción al proyecto y sus motivaciones.
- En el **capítulo 2** se hace un repaso del estado de la cuestión, indicando las principales aportaciones relacionadas con este proyecto de los artículos consultados.
- En el **capítulo 3** se hace una recopilación de los materiales y herramientas utilizados.
- En el **capítulo 4** se explica en detalle el procedimiento seguido para la obtención de los resultados.
- En el **capítulo 5** se muestran los resultados, realizando una comparativa con otros artículos y entre distintos resultados dentro de este estudio.
- En el **capítulo 6** se discuten los resultados del presente estudio y se proponen vías de mejora y desarrollo para futuros proyectos.

Capítulo 2

Estado de la técnica

2.1. Introducción del capítulo

En este capítulo se pretende dar una visión general del estudio del Alzheimer, y en concreto del estudio a partir de imágenes de resonancia magnética. Es aconsejable que el lector poco acostumbrado a los conceptos de estadística y contraste de hipótesis acuda previamente al anexo A. Por último se recuerda que si en algún momento no se especifica el significado de unas siglas en el anexo C se encuentra un diccionario de siglas.

2.2. El problema del diagnóstico

Desde la introducción del criterio para el diagnóstico del *National Institute of Neurological Disorders and Stroke Alzheimer's Disease and Related Disorders Association* (NINDS-ADRDA) en 1984 [23] los métodos clínicos para el diagnóstico de esta enfermedad no han variado en exceso. Sin embargo, los criterios para el diagnóstico neuropatológico han sufrido varias modificaciones [16][39][12].

Estos criterios sólo garantizan un diagnóstico probable. Para poder hacer un diagnóstico seguro es necesario recurrir a un estudio post mortem [14]. Esto hace que este estudio esté sujeto al error inherente del diagnóstico neuropatológico.

Respecto a los distintos subgrupos de diagnóstico se pueden distinguir:

- Sujetos control (NC): estos sujetos padecen únicamente un deterioro natural debido a la edad.
- Sujetos con deterioro cognitivo leve (MCI): se trata de sujetos que se encuentran en un estado transitorio entre el envejecimiento natural y el diagnóstico probable de la enfermedad [27]. Sin embargo, en este grupo heterogéneo se pueden distinguir dos grandes familias:
 - Sujetos MCI que acaban por desarrollar la enfermedad de Alzheimer, llamados por lo tanto MCI progresivos (pMCI).
 - Sujetos MCI que se mantienen estables (sMCI) y que en algunos casos incluso vuelven a un estado en el que son considerados como sujetos control.

- Sujetos diagnosticados de Alzheimer (AD): estos sujetos han sido diagnosticados con Alzheimer probable según los criterios de ADNI (véase sección 4.2.1).

2.3. Trabajos previos

Se describen en esta sección los conocimientos adquiridos a través de diversos artículos de estudios previos relacionados con el aquí presentado. Cada uno de estos estudios ha aportado algo de información que, puesta en conjunto, ha permitido alcanzar una mejor comprensión del proceso.

Son varios los estudios que buscan discernir entre sujetos pertenecientes a distintos grupos clínicos (pMCI, sMCI, AD, CN, demencia, etc.). Este estudio, como ya se ha dicho, se centra en la diferenciación entre los grupos pMCI y sMCI. Esto suele tener peores resultados que la comparativa entre otros grupos [20], y esto se debe principalmente a dos motivos:

1. Se trata de grupos muy heterogéneos. Ambas familias, pero especialmente sMCI, están formadas por sujetos muy distintos. Esto dificulta el centrarse en una serie de características concretas. Se trata además de diagnósticos cercanos. La diferencia entre uno y otro la define el tiempo de estudio. Esto quiere decir que si en un estudio se han considerado 36 meses para comprobar que un sujeto MCI no convierte a AD (sMCI), nada impide que a los 38 meses haya convertido y resulte ser en realidad un sujeto pMCI.
2. La evolución del sujeto depende en parte de factores no reflejados en el análisis. Se han encontrado sujetos sMCI que tras empeorar sus biomarcadores se han recuperado, se supone que gracias al ejercicio mental, e incluso han vuelto a medidas propias de CN (sujeto sano).

2.3.1. Querbes2009

En su estudio, Querbes [30] utilizó 444 pacientes de ADNI divididos en partes más o menos iguales entre los grupos CN,AD, MCI amnésico. Tras un estudio de 2 años, 72 de los 132 MCI habían convertido (pMCI) mientras que 50 habían permanecido estables (sMCI). Para entrenar al algoritmo se escogieron 30 sujetos control y 30 sujetos AD. En el mencionado estudio se utilizó un índice normalizado de espesor cortical (NTI) para establecer un continuo desde sujetos control hasta AD. Este índice tiene en cuenta la disminución del espesor del córtex con la edad y de esta manera agrupa distintas medidas de diferentes zonas del cerebro en una sola medida.

Las zonas escogidas cuyas medidas forman este índice se calcularon mediante un algoritmo de aprendizaje supervisado, incluyendo aquellas zonas que mejor discriminasen entre sMCI y pMCI. El proceso de asignación de índices a los sujetos del grupo de entrenamiento se repitió 100 veces con subconjuntos aleatorios para hacer un cálculo estadístico y evitar que se adecuara solamente a ese grupo de entrenamiento.

Para el análisis de covarianza entre los grupos se incluyeron en el estudio de Querbes tres factores intergrupales (género, diagnóstico y ApoE) y dos covariantes (edad y educación) para corregir sus efectos significativos.

Querbes pone de manifiesto ciertos resultados que deben tenerse en cuenta a la hora de abordar cualquier estudio:

- No se encontraron indicios de que el nivel de educación se reflejase en la puntuación MMSE. Además, el nivel cognitivo era independiente del grupo.
- Sujetos pMCI y AD no pudieron ser diferenciados a través del mencionado índice NTI.
- Respecto a la distinción entre pMCI y sMCI, Querbes informa de que una atrofia cortical generalizada fue observada en el grupo pMCI. Esto es un indicativo de que los sujetos se encontraban, anatómicamente hablando, en un estado algo avanzado de la enfermedad, habiendo perdido la posibilidad de detectar los cambios locales en regiones concretas.

2.3.2. Li2012

En su estudio Li [19] utiliza un acercamiento distinto. Introduce el coeficiente de deterioro como una medida más exacta del estado de la enfermedad ($TR = (1 - \text{fracendlinebaseline})$). Además, realiza una parcelación en regiones de interés (ROIs) basada en la muestra en lugar de las etiquetas anatómicas. Esto permite que los valores estadísticos como el espesor medio tengan una capacidad mayor de discriminación.

Los datos se almacenaron en una red para realizar la comparativa de las relaciones dentro del mismo sujeto y entre la población. Para cada ROI se extrajeron 5 grupos de características:

- espesor cortical en el baseline, en mm (estática).
- espesor cortical al final, en mm (estática).
- coeficiente de deterioro, como porcentaje (dinámica).
- Velocidad de deterioro, en mm/año (dinámica).
- Coeficiente de agrupación (clustering coefficient) de la red (red).

Los resultados del estudio demuestran que el coeficiente de deterioro (thinning ratio) es distinto para los diferentes grupos en un periodo de 2 años:

Grupo	NC	sMCI	pMCI	AD
TR	< 1,5 %	< 2 %	1 % – 4 %	2 % – 6 %

Tabla 2.1: Tabla con los índices de deterioro (TR) para los distintos grupos.

Este estudio muestra la mejora de la precisión al añadir estas características dinámicas como el índice o la velocidad de deterioro, llegando a mejorar de un

76.7% (contando sólo con las características estáticas) a un 80.3% (contando con ambas). Además, se comprobó que las características dinámicas eran más discriminantes para la diferenciación entre sMCI-pMCI mientras que las estáticas eran más discriminantes para la diferenciación entre CN-AD. También se sugiere que la distribución espacial es similar y que las zonas que sufren deterioro no difieren entre los grupos, pero sí cambia la velocidad del deterioro, siendo más rápido cuanto más cerca se está de padecer la enfermedad.

2.3.3. Eskilsen2015

Eskilsen llevó a cabo un estudio transversal (cross) en 2015 [10]. En este estudio se pretendía conseguir una clasificación entre sujetos NC, AD, sMCI y pMCI. Este estudio aporta una visión general ya recogida en los anteriores, por lo que simplemente se comentarán aquí las características que se revelaron como importantes a raíz de este estudio para la discriminación entre estos grupos de diagnóstico:

- Espesor del precúneo izquierdo
- El espesor del giro parahipocampal (véase figura 2.1)
- Los gradientes del hipocampo izquierdo y derecho
- El sulcus temporal superior, en combinación con otras características.

Respecto a los resultados, serán vistos en profundidad en la sección correspondiente de este trabajo al compararlos con los obtenidos.

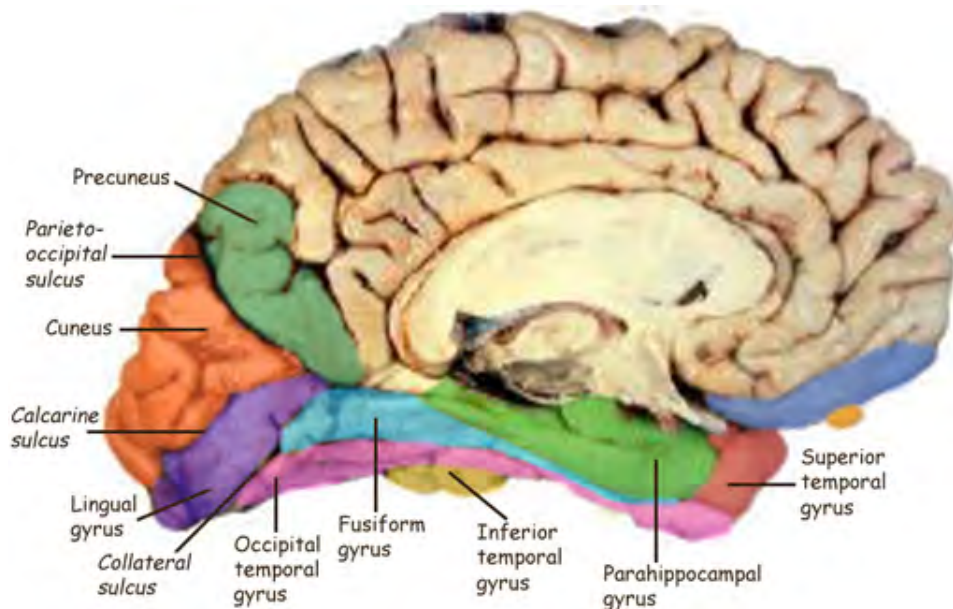


Figura 2.1: Esquema anatómico de los gyri.

2.3.4. Korolev2016

Otro estudio dedicado a diferenciar sujetos sMCI de sujetos pMCI es el estudio de Korolev [17], hecho sobre 259 pacientes MCI pertenecientes a la base de datos

de ADNI. Como fuentes de características que pudieran predecir esta división se incluyeron factores de riesgo, test neurodegenerativos, imágenes MRI y por último datos proteómicos (referentes al estudio de proteínas) en el plasma.

En este caso se utilizó un clasificador basado en Kernels para codificar el grado de similitud entre sujetos de un conjunto. En concreto se utilizó una extensión conocida como *multiple kernel learning* (MKL) que permite integrar información complementaria, resultando más eficaz para discernir entre AD y MCI.

Se plantea una opción interesante que consiste en la posibilidad de incluir un grupo de “poca confianza” para los casos en que la probabilidad de pertenecer a uno de los dos grupos no es muy alta. A este acercamiento se le conoce como clasificación con “opción a descartar” (*classification with “reject option”*).

Los factores de riesgo medidos incluyen la edad, sexo, educación, el genotipo APOE e historial familiar, entre otros. Respecto a los test neuropsicológicos, se incluyó una gran cantidad de ellos como candidatos. A continuación se expone la lista con los nombres originales para evitar malentendidos debido a la traducción:

- Mini-Mental State Examination (MMSE)
- ClinicalDementia Rating scale (CDR)
- Functional Activities Questionnaire (FAQ)
- Geriatric Depression Scale (GDS)
- Neuropsychiatric Inventory Questionnaire (NIQ)
- Modified Hachinski Ischemic Scale (MHIS)
- American National Adult Reading Test (ANART)
- WMS-III Logical Memory
- Alzheimer’s Disease Assessment Scale-Cognitive sub-scale (ADAS-Cog)
- Rey Auditory Verbal Learning Test (RAVLT)
- Verbal fluency test (VFT)
- Boston Naming Test (BNT)
- Digit Span Test (DST)
- Trail Making Test (TMT)
- Digit-Symbol Coding Test (DSCT)
- Clock-Drawing Test (CDT)

Esto es importante ya que los resultados de este estudio pueden ser utilizados para comprobar qué test neuropsicológicos tienen mayor poder de discriminación y, por lo tanto, no sería raro que apareciesen en los resultados. Para más información sobre los tests neuropsicológicos utilizados en el presente estudio véase la sección 3.3.

Para la selección de características se utilizó una combinación filtro-*wrapper*. Un método de selección del tipo filtro es independiente del algoritmo de clasificación y suele ser un paso previo al procesamiento. Aunque tiene la ventaja de ser muy rápido, corre el riesgo de no escoger al subconjunto adecuado de características. Un método *wrapper* corrobora la eficacia de la selección acudiendo al algoritmo de clasificación. La desventaja de este método es que por su naturaleza no permite utilizar algoritmos de clasificación demasiado pesados.

En el caso del estudio de Korolev, la parte de filtro se encarga de definir subconjuntos de distintos tamaños favoreciendo la capacidad de discriminación. En la etapa *wrapper* estos conjuntos son evaluados en términos de precisión en el grupo de validación. Precisamente esta estructura se ha establecido como referencia para el presente estudio, y aparece mejor explicada en la sección *K-folds Validation*, en el capítulo de Métodos.

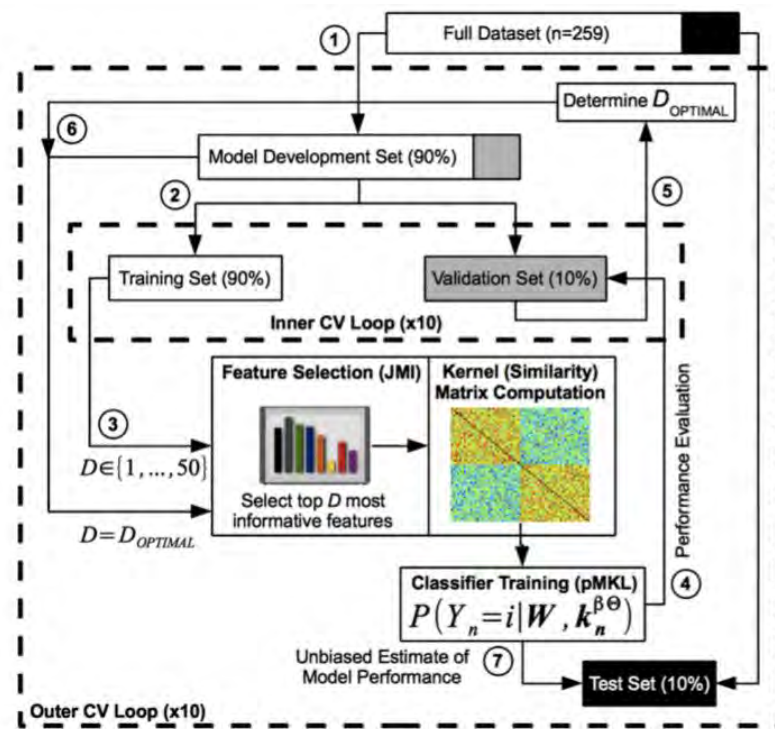


Figura 2.2: Esquema de la validación mediante k-folds llevada a cabo en el estudio de Korolev[17]. Imagen del artículo original.

El estudio de Korolev se ha dividido en 10 experimentos, cada uno de ellos agrupando distintas características:

1. Factores de riesgo (CFR)
2. Test neuropsicológicos+marcadores clínicos (CAM)
3. Marcadores MRI (MRI)
4. Marcadores proteómicos del plasma (PPM)

5. Modelo conjunto con las cuatro fuentes anteriormente mencionadas(CONCAT)

Los otros cinco experimentos se realizaron con la misma base para comprobar si el algoritmo de aprendizaje con múltiples kernels representa una mejora respecto al algoritmo con un único kernel

Como conclusión, se extrae que los conjuntos MRI y CAM dieron resultados aceptables y robustos (en torno al 78 % y 71 % de precisión respectivamente), y que el modelo conjunto CONCAT y el experimento con el uso de varios kernels dieron los mejores resultados (en torno al 80 % de precisión). Sin embargo, de estos se hablará más en profundidad en la sección correspondiente, al compararlos con el método utilizado en el presente trabajo.

Respecto a las características finalmente utilizadas del grupo CFR, sólo el APOE4 fue seleccionado con cierta consistencia. Respecto al modelo CAM, las características que resultaron más relevantes fueron ADAS-Cog, FAQ y RAVLT. En el modelo MRI, las características más relevantes fueron el volumen y el espesor cortical para varias regiones temporoparietales del cerebro (véase figura 4.33).

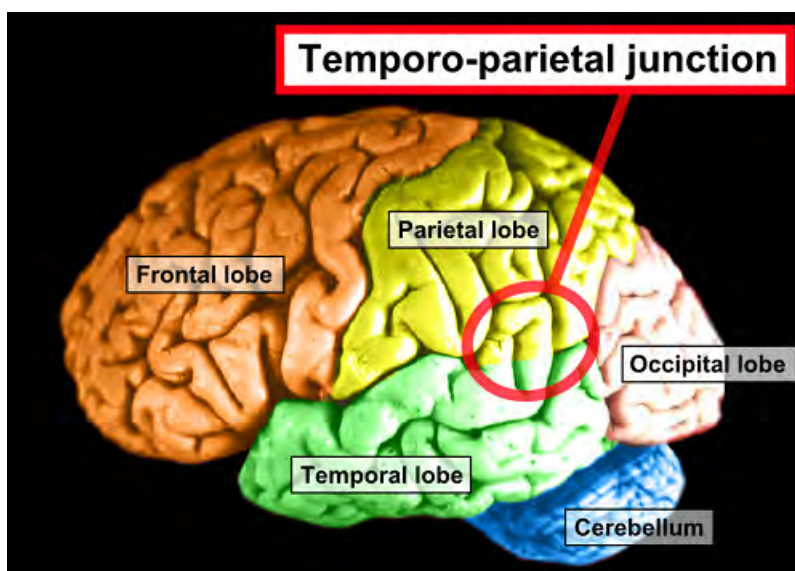


Figura 2.3: Esquema de la región temporoparietal. Imagen de John A Beal.

2.3.5. Rathore2017

Por la naturaleza de su estudio, se destaca respecto al resto el proyecto del equipo de Rathore [31]. Rathore elaboró en 2017 una recopilación de varios estudios relacionados con el Alzheimer. Este documento se muestra indispensable para la fase de documentación y estudios previos de cualquier proyecto relacionado con esta enfermedad. Además, en este estudio se ofrecen ciertas recomendaciones:

1. Se recomienda utilizar datos de pacientes que se encuentren en bases de datos como ADNI. Esto facilita la reproducibilidad del estudio y garantiza que el proceso de adquisición se ha realizado correctamente.

2. Velando por la reproducibilidad del estudio, se insta igualmente a ofrecer una versión asequible del código utilizado. Esto se hace especialmente crítico en caso de utilizar algún algoritmo de clasificación más complejo.
3. Respecto a las características escogidas biológicamente, Rathore presenta las siguientes:
 - a) El volumen hipocampal, así como el de otras estructuras subcorticales próximas (amígdala, putamen, tálamo, caudate, etc).
 - b) El espesor de la corteza entorrinal

Estas características, además de un sentido anatómico, han probado tener un gran poder de clasificación, tal y como se verá en otros estudios.

4. Respecto a los estudios basados en la superficie cortical, Rathore argumenta que los sutiles cambios producidos en etapas tempranas sólo se pueden detectar mediante el uso del análisis masivo (medidas de todos los vértices). Esto corresponde al aumento de la precisión que se verá en los estudios descritos al introducir este tipo de análisis.

2.3.6. Landin2017

En el caso de Landin [18], el estudio se centra en la diferenciación entre pacientes de Alzheimer y demencia frontotemporal. Esta última afecta a los lóbulos frontal y temporal, relacionados con diversas tareas del lenguaje y el comportamiento y planificación [9] [15].

En lo que atañe al estudio aquí presentado, lo realmente importante son las guías o pistas sobre posibles biomarcadores que puedan servir para una mejor identificación. También es importante el hecho de que en este estudio se utilizó el mismo modelo de clasificación que se pretende aplicar: el modelo de efectos lineales mixtos (LME). Sobre este modelo y su aplicación se hablará en detalle en la sección 4.3.2, dentro del capítulo de Métodos. No obstante, el estudio de Landin sirve para comprobar la eficacia de este modelo.

Por lo tanto, en este artículo también se introducen características que son candidatas a ser incluidas como efectos fijos o aleatorios. En el caso de Landin, el diagnóstico, el tiempo de seguimiento y la interacción entre ambos fueron considerados efectos fijos.

Landin también resalta la importancia que puede tener un estudio combinado entre regiones de interés (ROIs) y estadísticos globales. Se calcula, además, que tan solo un tercio de los estudios longitudinales basados en imágenes MRI ofrecen resultados con ambos tipos de características.

2.3.7. Sorensen2017

El estudio llevado a cabo por Sorensen en 2017 [38] combinó por primera vez medidas de volumen y espesor cortical con medidas de forma y textura del hipocampo para intentar dar un diagnóstico entre CN, MCI y AD simultáneamente. De nuevo la fuente de los datos fue ADNI, con imágenes MRI de 1.5T. El procesamiento de la imagen para obtener las medidas se hizo a través de Freesurfer. Respecto a la textura y forma del hipocampo, se calcularon tomando como punto de partida la segmentación realizada por Freesurfer, pues se trata de un método ya establecido.

Respecto a las medidas de volumen y espesor cortical, se tomaron en cuenta los valores medios en las regiones de interés (ROIs) que calcula FreeSurfer. Y aunque los resultados se comenten en mayor profundidad en su correspondiente apartado, es importante ver las medidas que resultaron finalmente eficaces para discriminar los tres grupos. Las características más frecuentes, según señala Sorensen, fueron el volumen hipocampal, el volumen ventricular, la textura hipocampal, el espesor de los lóbulos parietal y occipital y, por último, el espesor del córtex cingulate (véase figura 2.4). Además, Sorensen informa de las combinaciones más habituales que se han dado. Entre éstas encontramos:

- Textura hipocampal + Volumen hipocampal + Volumen ventricular
- Textura hipocampal + Volumen hipocampal + Volumen ventricular + Espesor del lóbulo parietal
- Textura hipocampal + Volumen hipocampal + Volumen ventricular + Espesor del lóbulo occipital

Esto puede dar una idea para estudios posteriores, como el presente, de medidas candidatas a formar parte del vector de características utilizado.

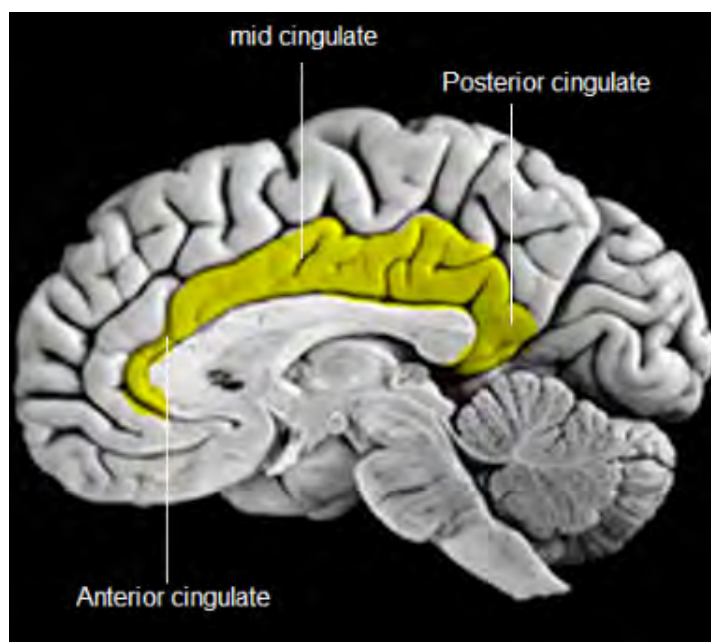


Figura 2.4: Esquema de la región cingulate.

2.4. Conclusión del capítulo

Aunque como se ha indicado estas lecturas han aportado una idea general de la problemática, hay ciertos aspectos concretos que se han tenido más en cuenta a la hora de llevar a cabo el presente estudio. Entre ellos destacamos:

- De Querbes2009 se extrae una advertencia respecto al análisis de los resultados. Esto es: que los sujetos pMCI estudiados probablemente sufran igualmente de un deterioro ya avanzado que dificulte el estudio del deterioro local.
- De Li2012 se extrae que la inclusión de medidas dinámicas (deterioro anual, p.ej.) pueden aportar bastante información a la hora de clasificar los dos grupos.
- De Eskilsen2015 se extraen algunas características que es probable que aparezcan a la hora de realizar el análisis estadístico. Entre ellas encontramos características relacionadas con el hipocampo (espesor del giro parahipocampal, volumen hipocampal, etc.) y otras medidas como el espesor del precúneo izquierdo.
- De Korolev2016 se extraen dos conceptos importantes: el primero es que la inclusión de tests neuropsicológicos puede mejorar en gran medida los resultados obtenidos. El segundo es el flujo de trabajo mediante dos k-folds anidados, que se verá más detalladamente en la sección correspondiente de Métodos.
- De Rathore2017 obtenemos una recopilación de estudios con los que se pueden contrastar los resultados obtenidos en la sección de Resultados. Además se han tenido en cuenta los consejos dados por el autor.
- De Landin2017 se extraen distintos elementos que pueden ser considerados a la hora de elaborar la matriz de efectos fijos (modelo LME). En concreto encontramos el diagnóstico y el tiempo de seguimiento. Además, se confirma la utilidad del modelo LME y su eficacia en este tipo de estudios.
- De Sorensen2017 se nos introducen nuevas ROIs con gran poder de discriminación. Aparte de las ya mencionadas medidas hipocampales se muestran el espesor del lóbulo parietal, occipital y el volumen ventricular como otras características candidatas.

Capítulo 3

Materiales y herramientas

3.1. Introducción del capítulo

En este capítulo se presentarán los materiales utilizados en el presente estudio. Se va a hacer especial hincapié en la base de datos (ADNI) y los datos de la población estudiada. Asimismo se comenta el software utilizado, los motivos de esta elección y cómo empezar a utilizarlo.

3.2. Datos sobre la población estudiada

Para este estudio se han utilizado datos obtenidos a través de ADNI. Se han realizado los experimentos con un set de pacientes que se habían procesado durante cursos anteriores.

En las tablas 3.1 y 3.2 se muestran los datos de este grupo estudiado. Se muestra la información clínica en el baseline así como el número de visitas del que se dispone.

Grupo	N	Edad (años)	MMSE	Género(% Masc)
pMCI	81	74.08(6.65)	26.84(1.73)	59.26 %
sMCI	101	75.33(7.20)	27.39(1.75)	66.34 %

Tabla 3.1: Tabla con los datos clínicos y demográficos del grupo utilizado para el estudio. Todos los datos se corresponden al baseline (primera visita).

Grupo	N_scans	avg_scans	bl	m06	m12	m24
pMCI	298	3.68(0.50)	0	1	24	56
sMCI	370	3.66(0.47)	0	0	34	67
Total	668	3.67(0.48)	0	1	58	123

Tabla 3.2: Tabla con los datos sobre el número de scans estudiados. N_scans indica el número total de visitas por cada grupo, mientras que avg_scans indica la media de visitas por sujeto para cada grupo y en total. Las últimas columnas indican el número de sujetos que sólo tienen la visita de base (*baseline*, *bl*), cuya última visita fue al medio año (m06), al año (m12), o a los dos años (m24).

Sobre el protocolo de clasificación utilizado se hablará en la sección correspondiente en el capítulo de Métodos (4.2.1).

De forma paralela se están procesando con FreeSurfer imágenes de nuevos pacientes que permiten aumentar el número de muestras. Con esto se pueden plantear nuevos acercamientos en futuros trabajos. De todos los sujetos que se pretenden procesar e incluir, se han formado tres grupos:

1. Un grupo de sujetos sMCI.
2. Dos grupos de sujetos pMCI.

De estos dos últimos grupos, uno de ellos se ha procesado a la par que se realizaba este trabajo. Los datos demográficos del susodicho grupo son los siguientes:

Grupo	N	Edad	MMSE	Género(% Masc)
pMCI	42	75.35(7.67)	26.45(1.67)	71.43 %
sMCI	0	-	-	-

Tabla 3.3: Tabla con los datos clínicos y demográficos del grupo procesado mediante FreeSurfer.

Grupo	N_scans	avg_scans	bl	m06	m12	m24	m36
pMCI	209	4.98(1.49)	3	0	4	13	22
sMCI	0	-	-	-	-	-	-
Total	209	4.98(1.49)	3	0	4	13	22

Tabla 3.4: Tabla con los datos sobre el número de scans estudiados. N_scans indica el número total de visitas por cada grupo, mientras que avg_scans indica la media de visitas por sujeto para cada grupo y en total. Las últimas columnas indican el número de sujetos que sólo tienen la visita de base (*baseline*, *bl*), cuya última visita fue al medio año (m06), al año (m12), o a los dos años (m24).

3.3. ADNIMERGE y tests neuropsicológicos

ADNIMERGE es una inmensa tabla que ofrece ADNI con una gran cantidad de información, tanto demográfica como de tests neuropsicológicos. Su importancia es capital en este estudio, ya que ha permitido ampliar la cantidad de tests neuropsicológicos estudiados. En trabajos previos de este mismo proyecto se habían utilizado los tests CDR, GDS y MMSE. En este estudio esta variedad se ha podido expandir a:

1. **MMSE:** Mini-Mental State Examination. Se trata de un test de cribado de demencias, útil también en el seguimiento evolutivo de las mismas. Puntúa como máximo un total de 30 puntos y hay 5 apartados que comprueban orientación, memoria inmediata, atención y cálculo, recuerdo diferido y lenguaje y construcción. En definitiva, se trata de medidas de la capacidad mental de un individuo, y existen distintos valores esperables dependiendo de los años de educación y la edad.

2. **CDR:** *Clinical Dementia Rating scale* se trata de una escala que va de 0 (sin demencia) a 3 (demencia severa) y que caracteriza precisamente la gravedad de la demencia [34]. Existen varias secciones y, en concreto, la del área de memoria tiene un gran peso a la hora de determinar el diagnóstico general.
3. **GDS:** *Geriatric Depression Scale* se trata de un sencillo test desarrollado por el equipo del doctor Brink[5] que trata de dar una medida de la depresión.
4. **ADAS-Cog:** *Alzheimer's Disease Assessment Scale - Cognitive sub-scale* es una medida de la capacidad cognitiva del paciente. Dentro de las pruebas hay distintas categorías que miden distintos efectos (desorientación, memoria léxica, reconocimiento de palabras, etc.) [37]. Con ADNIMERGE se han podido añadir las pruebas ADAS11, ADAS13 y ADASQ4.
5. **RAVLT:** *Rey Auditory Verbal Learning Test* consiste en una serie de pruebas de memoria lingüística [4]. Dentro de estas pruebas encontramos distintas variaciones. Concretamente, en ADNIMERGE se encuentran RAVLT_immediate, RAVLT_learning y RAVLT_perc_forgetting.
6. **FAQ:** *Functional Activities Questionnaire* es un cuestionario realizado a alguna persona conocida o de confianza (normalmente pareja, hijos o familiar) y que está dividido en diez preguntas sobre actividades diarias. La puntuación varía desde 0 (total independencia) hasta 3 (dependencia total) [22]. La puntuación final es la suma de las puntuaciones asignadas a cada actividad.

3.4. ADNI

Alzheimer's Disease Neuroimaging Initiative (ADNI) es una plataforma pública creada en 2003 para albergar un banco de neuroimágenes y datos de distintos pacientes. El fin de esta iniciativa es que los datos puedan ser utilizados en distintos estudios estadísticos en la investigación en la prevención del Alzheimer. Esta iniciativa está englobada dentro del conjunto de *Image and Data Archive* (IDA).

En este estudio en concreto se ha recurrido al uso de imágenes de resonancia magnética estructural (MRI) de un campo magnético de 1.5T de fuerza. Se trata de una técnica que permite obtener la estructura y composición del cerebro. Esta técnica con este campo magnético es frecuente en los centros especializados en neurociencia, y suele formar parte de cualquier procedimiento de diagnóstico.

En esta sección se pretende ofrecer información suficiente como para tener acceso y empezar a utilizar ADNI. Para obtener acceso a los datos existen dos opciones:

1. registrarse directamente en ADNI (<http://adni.loni.usc.edu/>), con lo que automáticamente se creará una cuenta con ese mismo nombre de usuario en IDA.
2. Registrarse primero en IDA, con lo que luego será necesario rellenar otro formulario para ADNI. En concreto, desde <https://ida.loni.usc.edu> se permite

registrarse en IDA y posteriormente navegar hasta la página de ADNI. Esto se consigue a partir de la referencia que aparece en la lista de bancos de datos mostrada en la parte inferior.

Lo primero que se debe hacer es entrar en la página de ADNI y se pulsa en la pestaña *Data and Samples*, y después en el submenú *access data* (véase figura 3.1). Se deberán aceptar las condiciones de uso y posteriormente ofrecer un correo electrónico al que llegará un código de confirmación.

Tras escribir el código se tendrá que rellenar un formulario donde se pedirá información sobre la investigación que se pretende llevar a cabo. Por último se debe esperar una semana aproximadamente para que un comité decida si falta información.

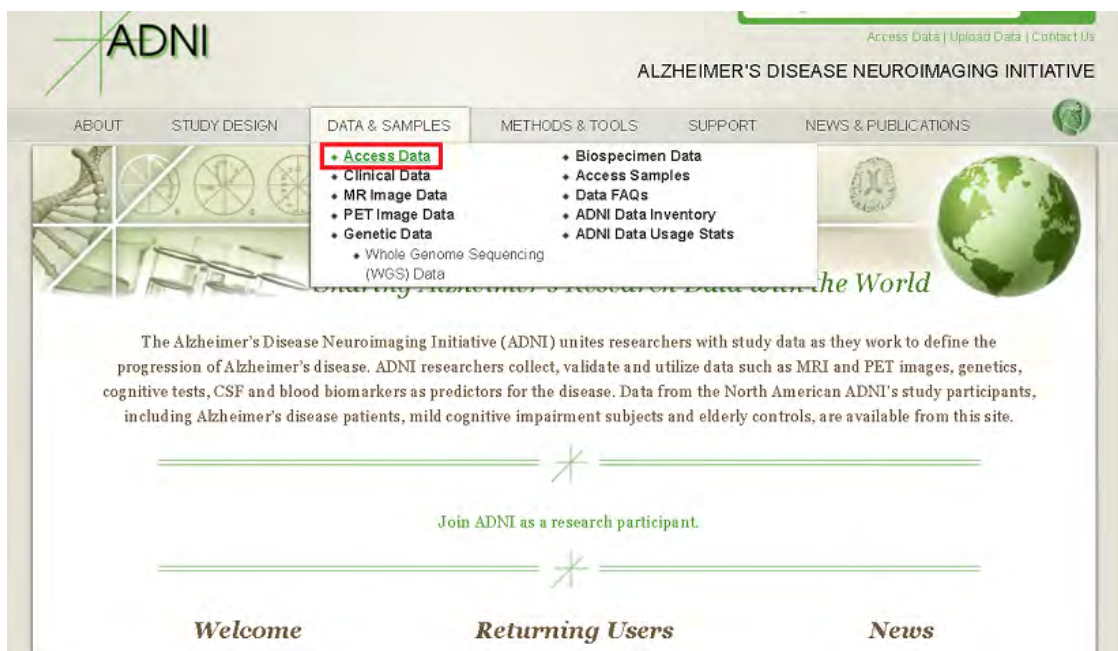


Figura 3.1: Captura de la web de ADNI. En rojo se resalta la pestaña de acceso a los datos

Una vez finalizado el registro se podrán llevar a cabo las mismas búsquedas con opción a descargar los datos. Es importante señalar que se dispone de dos tipos de datos distintos por cada conjunto descargado:

- Archivos *.nii* que contienen la información de las imágenes MRI recogidas. Se trata de archivos que ocupan un espacio relativamente grande en memoria, por lo que normalmente se guardarán como archivos comprimidos en formato *.nii.gz*.
- Archivos *.xml* que contienen la metainformación del archivo y del paciente. Se trata de datos como la edad, el tipo de paciente, el sexo, etc. así como de los resultados de distintos test neuropsicológicos: CDR, GDS, MMSE, etc.

El conjunto de estos dos archivos será el punto de partida para cualquier estudio estadístico o de simple análisis de neuroimagen. El proceso de descarga, así como su posterior tratamiento, se describe en una sección aparte.

3.5. MATLAB

Matrix Laboratory (MATLAB) es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M).

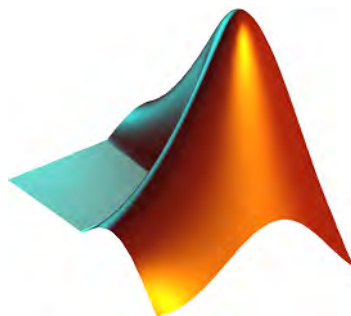


Figura 3.2: Logotipo de MATLAB.

Se trata de una herramienta muy utilizada en estudios de ciencia e ingeniería. En este estudio se han aprovechado cuatro características importantes de este software:

- Su capacidad para hacer llamadas al sistema. Esta capacidad se ha utilizado para automatizar las llamadas a FreeSurfer y otros softwares de terceros.
- Su capacidad para incluir librerías. En esta línea se han utilizado las librerías mRMR y LME, tal y como se verá en la sección de Métodos. Esta característica facilita la implementación de ciertos algoritmos mediante una interfaz relativamente sencilla.
- Su programación orientada a vectores, esencial a la hora de acelerar los procesos relacionados con el análisis de grandes cantidades de datos.
- Su amplia documentación. Dado que es un programa y un lenguaje con un largo recorrido, la información respecto a su uso o la implementación de funciones concretas está convenientemente documentada.

En concreto, en este estudio se ha utilizado la versión MATLAB2017. En los primeros pasos no tiene mayor importancia que el cambio en la sintaxis de algunos mandatos respecto a versiones anteriores. Sin embargo, es de vital importancia el cambio introducido de las cadenas de caracteres como objetos de tipo *string*. En etapas posteriores de procesamiento de datos se utiliza este nuevo comportamiento para estructurar los datos.

3.6. FreeSurfer

FreeSurfer es una herramienta de software libre que nos permite hacer operaciones, visualizar y marcar zonas concretas en un cerebro a partir de los datos de neuroimagen MRI. Este programa va a ser el encargado de transformar las imágenes iniciales en datos de entrada para el posterior estudio estadístico.



Figura 3.3: Logotipo de FreeSurfer.

Antes de descargar este software debemos tener en cuenta que ocupa un gran espacio dentro del disco duro. Por lo tanto, una recomendación sensata en caso de estar trabajando en una máquina virtual es definir el disco duro como dinámico para poder ampliarlo.

Para descargar el programa, lo primero que se debe hacer es acudir a la página oficial <https://surfer.nmr.mgh.harvard.edu/fswiki/DownloadAndInstall>. Una vez allí se deberá descargar un fichero comprimido atendiendo al sistema operativo. Tras esto, se seguirán las instrucciones de la propia página, resumidas en este documento para el caso de Linux con un intérprete bash:

1. Se descomprime el fichero descargado, recomendablemente en `/usr/local` con el comando

```
$> tar -C /usr/local -xzvf nombre_del_archivo_de_instalacion.tar.  
↪ gz
```

Es importante señalar que esta línea de código debe ser modificada para que el nombre del archivo sea el nombre del que ha sido descargado.

2. Lo siguiente que se necesita es la licencia, consistente en un archivo `.txt`. Para obtenerla sólo es necesario seguir el hipervínculo ofrecido en la misma página, en la sección `LICENSE`. Tras rellenar un formulario la licencia será enviada al

correo electrónico proporcionado. Una vez en posesión de la licencia, se la debe mover al directorio donde está instalada la aplicación (normalmente /usr/local/freesurfer) para que sea efectiva.

3. En un terminal deben escribirse las siguientes líneas, definiendo el lugar de trabajo de FreeSurfer, que es el directorio donde ha sido descomprimido:

```
$> export FREESURFER_HOME=/usr/local/freesurfer
$> source $FREESURFER_HOME/SetupFreeSurfer.sh
```

4. Por último se debe definir una variable con el nombre SUBJECTS_DIR que indique el directorio donde FreeSurfer buscará los datos de los pacientes para su uso dentro del programa:

```
$> export SUBJECTS_DIR=<path to subject data>
```

Este directorio, sin embargo, puede cambiar. En ese caso simplemente tendremos que redefinir la variable SUBJECTS_DIR de la misma forma. Se recurrirá a esto a la hora de hacer la extracción de datos procedentes del estudio longitudinal. Esto se verá más en detalle en la sección de Métodos.

Estos dos últimos pasos serán obligatorios cada vez que se inicie una sesión de terminal. Por ello, es aconsejable incluirlos en el archivo .profile (o similares) para que se ejecuten de manera automática al iniciar la terminal. Este fichero suele ser hijo del directorio del usuario en los sistemas linux. En todo caso, es un paso imprescindible antes de comenzar a trabajar en cualquier ordenador.

Además de las operaciones y procesamiento que permite realizar, FreeSurfer incluye una serie de herramientas adicionales. La más importante para este estudio es freeview: se trata de una aplicación gráfica que permite mostrar los distintos archivos que procesa FreeSurfer.

Principalmente permite mostrar volúmenes (imágenes basadas en vóxeles) y superficies (estructuras formadas por vértices). En esta visualización se pueden modificar aspectos como la opacidad o incluir una segmentación o etiquetado.

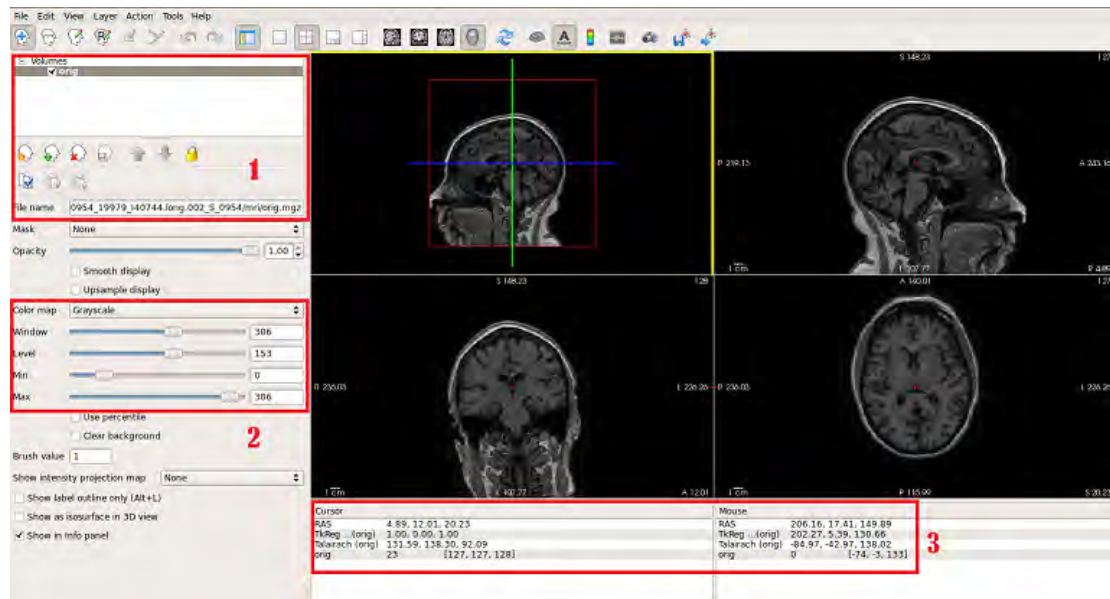


Figura 3.4: Captura de pantalla de FreeView. En ella se pueden apreciar las tres zonas principales además de la representación de un archivo .mgz: La selección de archivos mostrados (1), las opciones de visualización (2) y coordenadas y/o características de la zona seleccionada (3).

Dentro del ámbito de Freesurfer/Freeview se pueden destacar distintos tipos de archivos:

- **Archivos .mgh/.mgz:** se trata de archivos con información volumétrica (vóxeles). Los archivos .mgz no son más que un archivo .mgh comprimido (.mgh.gz). Esto es conveniente ya que suelen ser archivos pesados, por lo que se almacenan como .mgz. Pueden importarse desde FreeView, al igual que las superficies.
- **Archivos .label:** se trata de archivos de etiquetado. Normalmente se pueden encontrar en dos contextos: en la segmentación del Atlas, donde cada etiqueta corresponde a una ROI, o al generar clusters para el análisis masivo, donde cada cluster se ve representado por una etiqueta. También es posible generar etiquetas personalizadas para estudiar una zona concreta.
- **Archivos .stats:** se trata con archivos con información estadística (normalmente la media) sobre un volumen o superficie. Estos archivos generados durante el procesamiento permiten acceder a estos datos rápidamente a través de las funciones de FreeSurfer para generar las tablas de datos.
- **Archivos overlay:** no poseen una terminación concreta, pero se caracterizan por dar un valor a cada vértice de una malla. En este contexto encontramos archivos que almacenan el espesor en cada punto del córtex (?h.thickness) o el mapa de los q-valores (véase anexo A). Pueden ser cargados en FreeView para su visualización.

3.7. Conclusión del capítulo

Como se ha visto, los sujetos utilizados se han descargado de ADNI y han sido procesados mediante FreeSurfer. El flujo completo del programa ha sido programado en MATLAB (Véase la sección de Métodos) por su capacidad de llamar al sistema y su carácter vectorial. Además, se ha utilizado FreeView para la representación gráfica. Por último debe destacarse la aportación de ADNIMERGE así como de los distintos paquetes de MATLAB que han agilizado todo el trabajo.

Capítulo 4

Métodos empleados

4.1. Introducción del capítulo

Se procede en esta sección a describir los métodos empleados a la hora de procesar y analizar las imágenes MRI para la obtención de los resultados descritos en el presente estudio.

Se ha considerado organizar el capítulo en varias partes, siguiendo la división propia del flujo de trabajo:

- Una sección dedicada a la extracción de datos. Todo lo contenido en esta sección es previo al trabajo estadístico. No sólo se incluye una explicación detallada de cada paso, sino que se hace referencia a su implementación en el código adjunto a este trabajo.

En este contexto se incluye el procesamiento con FreeSurfer, el análisis masivo y la construcción de la tabla de características que se toma como punto de partida para la parte estadística.

- De manera análoga, en la parte dedicada al análisis estadístico se muestran los procesos y el flujo de trabajo una vez obtenidos los datos. Aquí se incluye la parte de selección de características mediante mRMR, el modelo LME así como la aplicación de los k-folds anidados.

Este trabajo comparte una parte importante del procesamiento de datos con otros estudios [20], por lo que partes del código son propias y otras han sido reutilizadas. En cualquier caso, en todas las etapas se ha adaptado el código para ofrecerlo en una versión autocontenida y documentada adjunta a este trabajo.

4.2. Extracción de datos

4.2.1. Adquisición de datos desde ADNI

Búsqueda de datos

Para este estudio, como ya se ha mencionado, se han utilizado datos provenientes de ADNI. Parte de estas imágenes habían sido ya procesadas en años anteriores. Sin embargo, se han añadido sujetos nuevos con el fin de aumentar el tamaño de la

población N.

Esto no sólo garantiza una mayor validez del resultado estadístico. Una mayor N también permite utilizar otros acercamientos al problema. Se puede encontrar un mayor detalle sobre los datos demográficos de esta población en el capítulo de Materiales y Herramientas.

A la hora de buscar un paciente, un método recomendable es acudir a la pestaña de búsqueda avanzada de ADNI. Una vez aquí se especifica el ID del sujeto y con la opción de *Image Processing* seleccionada se activa la búsqueda. Este método permite asegurarse de no omitir archivos disponibles de alguna de las visitas.

Una vez realizada la búsqueda se añaden los archivos deseados a una colección personal. Esta colección puede crearse antes o a la hora de añadir los archivos. Tras esto se navega de manera habitual por la pestaña de colecciones tal y como se ha descrito previamente en la sección de Materiales.

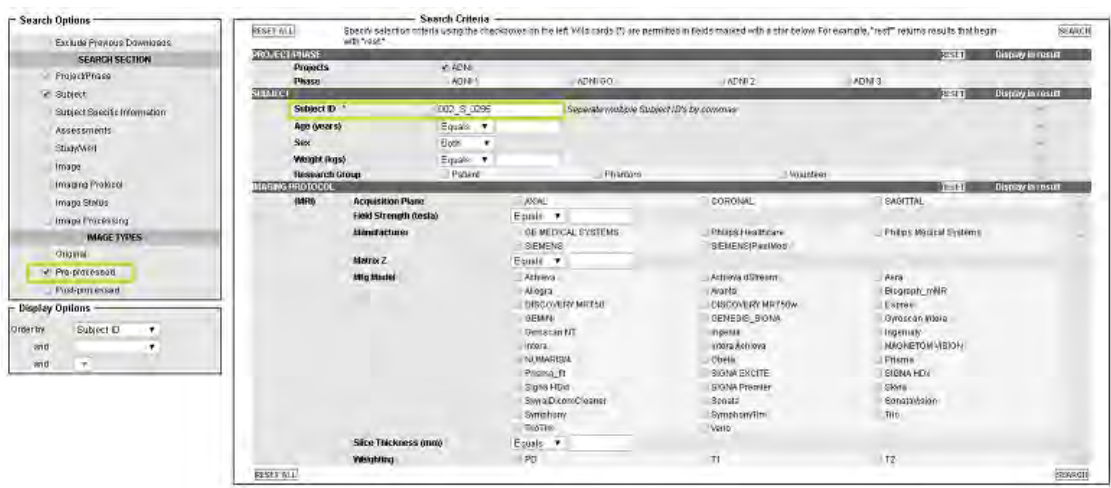


Figura 4.1: Pantalla de búsqueda avanzada de ADNI.

Criterio de clasificación de ADNI

El criterio de inclusión de ADNI (<https://clinicaltrials.gov/ct2/show/NCT00106899>) diagnostica Alzheimer si el sujeto obtiene una puntuación MMSE entre 20 y 26, una puntuación CDR de 0.5 ó 1, y cumple con los criterios NINCD-S/ADRDA para el diagnóstico probable de Alzheimer [23]. Como ya se ha explicado anteriormente, este criterio presenta un error inherente respecto al diagnóstico post mortem. Sin embargo para este estudio se va a considerar como la verdad con la que se van a comparar los diagnósticos del algoritmo (*the ground truth*, en inglés). A continuación se especifica más detalladamente el criterio de inclusión de ADNI para los distintos grupos:

Por último, para este estudio se ha hecho necesaria la distinción entre pMCI y SMCI. Para ello lo único que se ha hecho es filtrar los sujetos procesados por aquellos

Grupo		CN	MCI	AD
Logical Memory (II)	+16 AE	≥ 9	≥ 8	≥ 8
	8-15 AE	≥ 5	≥ 4	≥ 4
	0-7 AE	≥ 3	≥ 2	≥ 2
MMSE		24-30	24-30	20-26
CDR		0	0.5	0.5 ó 1

Tabla 4.1: Tabla con los criterios de clasificación de ADNI. AE = años de educación.

que presentasen un diagnóstico MCI en el baseline. Tras esto, si el sujeto llega a presentar un diagnóstico AD en alguna de sus visitas basándose en el criterio mostrado en la tabla 4.1, entonces el sujeto en cuestión es etiquetado como pMCI. En caso contrario es etiquetado como sMCI.

4.2.2. Flujo de trabajo general en la extracción de características

A continuación se describen en líneas generales las tres etapas principales de la extracción de características:

1. **Procesamiento mediante FreeSurfer de las imágenes de ADNI:** en esta fase se pasa de archivos .nii de las imágenes MRI a una segmentación longitudinal del cerebro del sujeto. El flujo completo queda recogido en la figura 4.2.

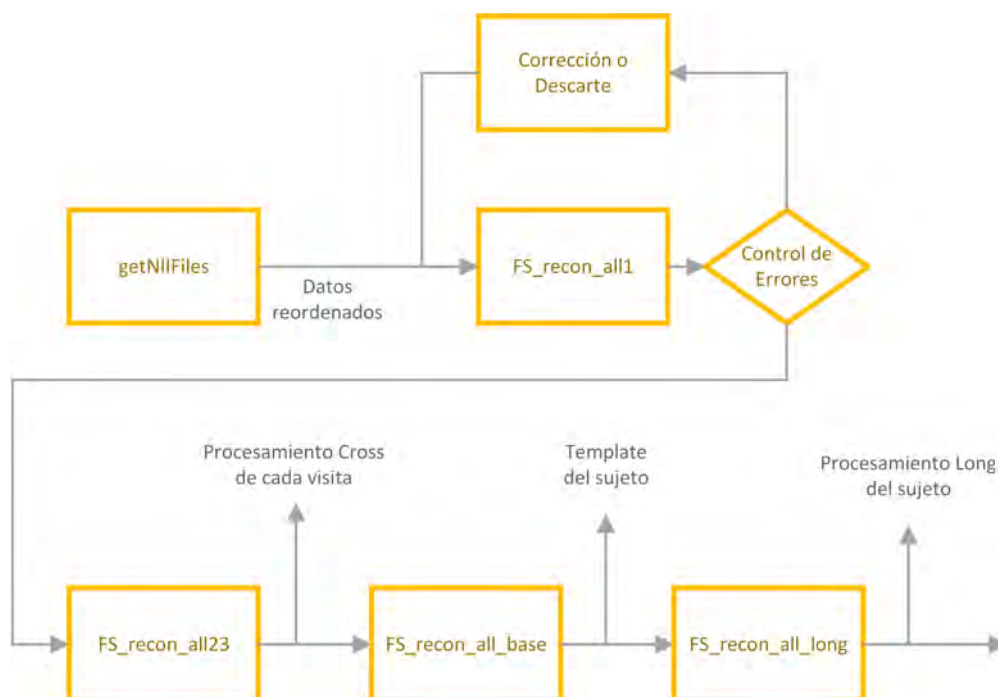


Figura 4.2: Flujo de trabajo con FreeSurfer. A partir de los archivos Long se obtendrán todas las características.

2. **Extracción de ROIs y Tests:** en esta fase se extraen los valores estadísticos (valor medio) de las ROIs y se añade el valor de los tests neuropsicológicos ex-

traído desde ADNIMERGE. El flujo de trabajo puede apreciarse en el diagrama 4.3

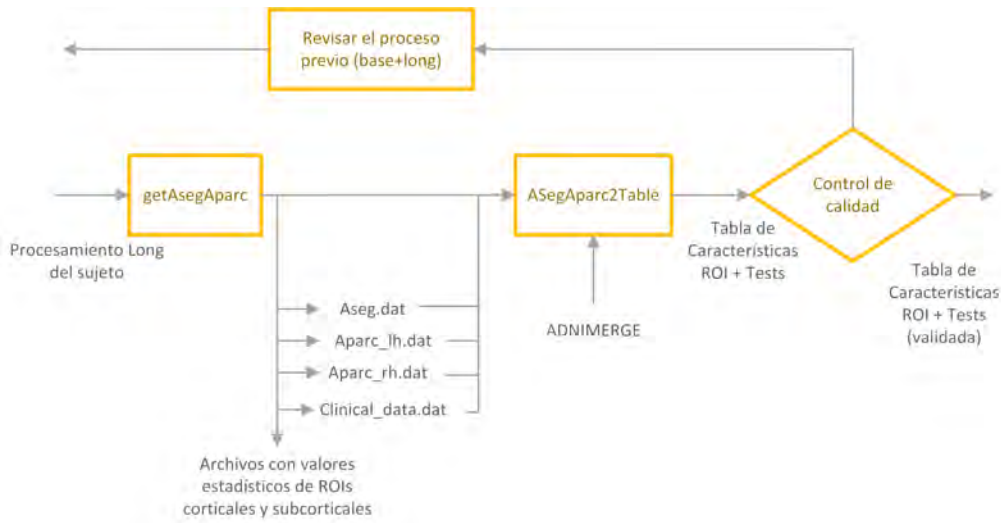


Figura 4.3: Flujo de trabajo con FreeSurfer. A partir de los archivos Long se obtendrán todas las características.

3. **Procesamiento masivo:** este análisis se encuentra a medio camino entre el procesamiento y el análisis estadístico. Se trata del análisis vértice a vértice, definiendo nuevas zonas, guardando la máscara para cada una de ellas y añadiendo el valor medio en cada zona a la tabla de características.

4.2.3. getNIIFiles

En esta fase simplemente se redistribuyen los archivos .nii y .xml en un árbol de directorios más manejable que servirá de base para los pasos posteriores. Por último, se comprimirán los archivos .nii a un formato .nii.gz.

Deben crearse previamente dos directorios donde se guardarán los archivos .nii y .xml, y debe localizarse la carpeta proveniente de la descarga de ADNI donde se encuentran los susodichos archivos.

Al inicio del código se cuenta con una zona donde se introducen las carpetas necesarias para el correcto funcionamiento del script. Esto será una constante en todos los códigos desarrollados o retocados.

Tras esto, se procede a contar el número de parejas de archivos xml+nii de los que se dispone. La función searchNIIFiles se encarga de determinar el número, localización y nombre de todos los archivos .nii, así como de juntar todos los archivos xml en una lista para iterar más adelante sobre ella en el código principal. Esta función está preparada para el formato ADNI, por lo que debe ser cambiada en caso de trabajar con otras fuentes. A continuación se muestra un pseudocódigo con los pasos seguidos:

Algoritmo 1 pseudocódigo de getNIIFiles

Entrada:

pathADNI: carpeta donde se guardan los archivos según la estructura de ADNI.
 pathNIIDest: carpeta donde se guardarán los archivos nii.gz.
 pathXMLDest: carpeta donde se guardarán los archivos xml.

Salida:

Nuevo árbol de directorios.
 1: xmlList = lista de archivos pathADNI/*.xml
 2: count = 1
 3: **para** todos los subdirectorios de pathADNI **hacer**
 4: **si** es archivo válido **entonces**
 5: listFiles(count) = camino/nombre_original.nii
 6: names(count) = IDsujeto_visita.nii
 7: count = count + 1
 8: **fin si**
 9: **fin para**
 10: **para** i = 1 hasta Número de archivos en xmlList **hacer**
 11: copiar archivo listFiles(i) a pathNIIDest/names(i)
 12: comprimir archivo pathNIIDest/names(i)
 13: copiar archivo pathADNI/xmlList(i) a pathXMLDest/names(i).xml
 14: **fin para**

4.2.4. FS_recon_all_cross

Explicación

En esta fase se procesan individualmente las imágenes. Es decir, no se tiene en cuenta la evolución temporal, sino simplemente el estado para una visita concreta. A este tipo de análisis se le llama transversal o *cross*. Se trata de una operación con un alto coste computacional (en el rango de horas) y a ser posible es deseable ejecutar el bucle principal en paralelo (véanse los algoritmos 2 y 3).

Para entender totalmente la función de FreeSurfer que da nombre a esta etapa de procesamiento (recon-all) se debe acudir a la sección correspondiente de la página oficial de FreeSurfer: <https://surfer.nmr.mgh.harvard.edu/fswiki/recon-all#StepDescriptionSummaries>.

FreeSurfer ofrece la posibilidad de ejecutar todas las etapas del proceso de una sola vez con el flag -all. Sin embargo, se ha implementado una separación entre las secciones -autorecon1 (hasta la obtención de la máscara del cerebro) y -autorecon2/-autorecon3. Estas dos últimas se ejecutan en el mismo script.

Esto permite ejecutar un pequeño control de calidad manual en una etapa temprana (véase figura 4.13). Además, la mayor parte de errores aparecen en la primera etapa, que no tarda tanto en ejecutarse, permitiendo no perder horas de trabajo. En concreto, en este trabajo se ha incluido el flag -cw256, que permite corregir errores de rotación o posición que dan lugar a errores en el campo de visión (*Field of View*).

En la figura 4.4 se muestran las etapas generales que ayudan a entender el objetivo de cada paso dentro de la función recon-all, eje central del script.



Figura 4.4: Etapas generales de la función recon-all.

Al iniciarse la ejecución, se crean una serie de carpetas de sujetos mediante el comando `mksubdirs`. De la carpeta de cada sujeto cuelgan distintos directorios donde se guardará la información procesada. Se procede a continuación a describir brevemente todos los pasos internos dentro de este procesamiento de imágenes:

1. **Corrección de movimiento:** si se dispone de varios volúmenes (archivos `.mgz`) corrige pequeñas desviaciones calculando una media entre las distintas medidas de una misma sesión. Estos archivos de entrada se encuentran en la subcarpeta del sujeto llamada `mri/orig/XXX.mgz`, siendo `XXX=001,002,003,etc.` La salida será el volumen `mri/orig.mgz`, y si sólo hubiese un archivo `001.mgz` será éste el que se considerará como `mri/orig.mgz`
2. **Corrección de intensidad no uniforme:** se realiza una normalización no paramétrica y no uniforme de la intensidad (N3) [36]. Esto corrige defectos de uniformidad en los datos de la resonancia magnética. Por defecto ejecuta cuatro iteraciones (véase figura 4.5).

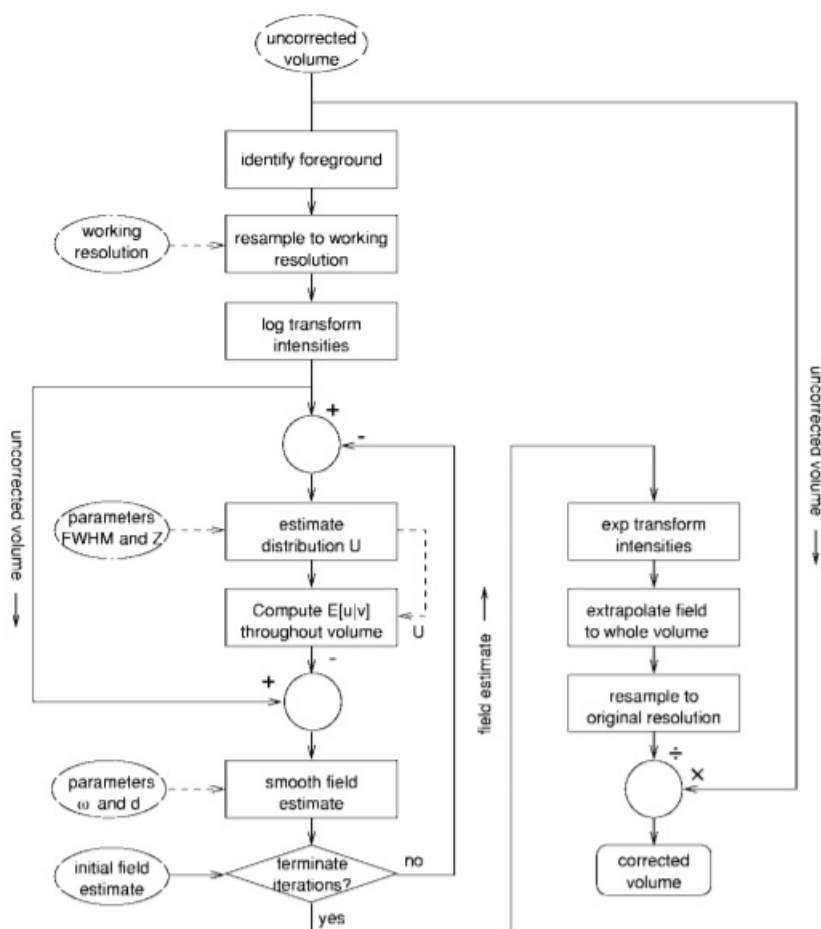


Figura 4.5: Flujo de trabajo del método de normalización N3.

3. **Talairach:** en esta etapa se produce una transformación afín desde el volumen inicial para cuadrar con el Atlas MNI305. Una transformación afín incluye 12 grados de libertad: 3 por cada eje en las operaciones de traslación, rotación, escalado y orientación (véase la figura 4.6).

Debe recordarse que esta transformación, así como todas las que aparezcan posteriormente, es necesaria para trabajar en un mismo espacio y poder comparar los cerebros. A través de la transformación, cada punto del espacio original tiene su correspondiente en el espacio común o ATLAS. Gracias a esto también se puede realizar una segmentación o etiquetado de zonas.

Una ventaja decisiva de las transformaciones afines es que pueden representarse en forma matricial, lo que acelera los cálculos. Un ejemplo cotidiano y visual de estas transformaciones se puede encontrar en cualquier programa de diseño gráfico, donde se permite al usuario realizar estas manipulaciones a una imagen [24].

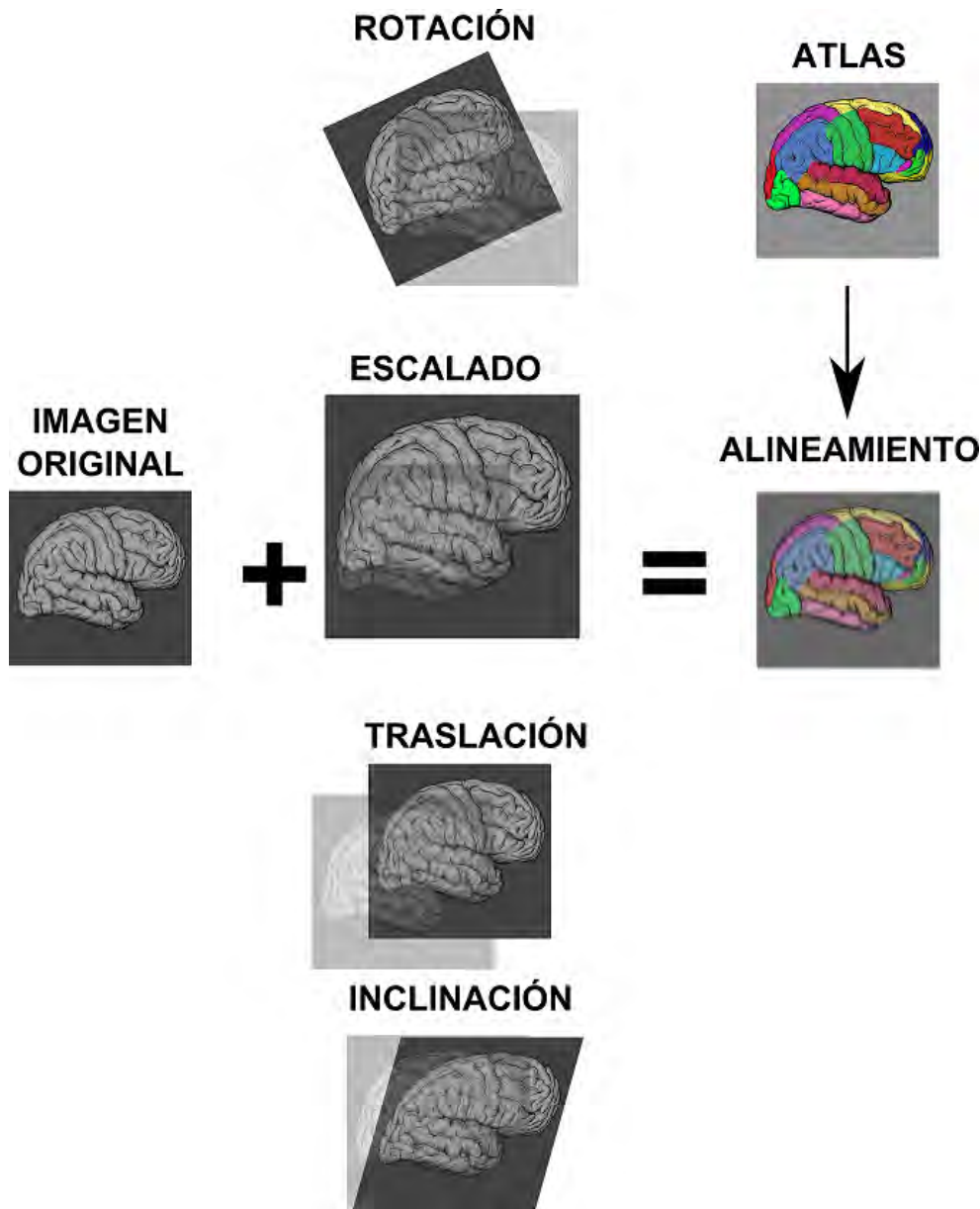


Figura 4.6: Esquema de los componentes de una transformación afín.

Gran parte de los pasos siguientes utilizan las coordenadas talairach generadas como punto de partida. Esta etapa crea el archivo *mri/transform/talairach.auto.xfm* y *talairach.xfm*.

4. **Normalización:** se trata de una segunda normalización de intensidad del volumen *orig.mgz*. El resultado se guarda en *mri/T1.mgz*. Esta etapa es fundamental para la comparación entre distintos pacientes y la identificación de estructuras corticales y subcorticales, ya que de esta forma los puntos o zonas etiquetados en el Atlas pueden traducirse a su correspondiente en el volumen del sujeto.

5. **Eliminación del cráneo:** en esta etapa se elimina el cráneo y se almacena el resultado en *mri/brainmask.auto.mgz* y *mri/brainmask.mgz*.
6. **Segmentación subcortical automática:** Esta etapa se desarrolla en seis fases:
 - a) Registro EM(CGA): se encarga de la transformación para alinear el volumen *mri/nu.mgz* con el volumen del atlas GCA por defecto (Gaussian Classifier Atlas) almacenado en *FreeSurfer_HOME/average/RB_all_2008-03-26.gca*.
 - b) Normalización CA: una tercera normalización basada en el modelo GCA. En este paso se crea el volumen *mri/norm.mgz*. También se genera el archivo *mri/transforms/talairach.lta*.
 - c) Registro CA: calcula la transformación no lineal que permite alinear de nuevo el volumen con el atlas GCA tras haberse producido la normalización. Es el llamado registro canónico.
 - d) Eliminar cuello: se elimina la región del cuello del volumen resultante de la corrección NU: *mri/nu.mgz*. Se sirve de la transformación calculada en el anterior paso (Registro CA). Este paso da como resultado el archivo *mri/nu_noneck.mgz*.
 - e) Registro EM con cráneo: calcula la transformación para alinear el volumen obtenido en el paso anterior (*mri/nu_noneck.mgz*) con el volumen GCA procesando el cráneo. Este paso genera el archivo *mri/transforms/talairach_with_skull.lta*.
 - f) Etiquetado CA: etiqueta las estructuras subcorticales basándose en el modelo GCA con el que previamente se ha alineado para permitir esta comparación. Esto genera los archivos *mri/aseg.auto.mgz* y *mri/aseg.mgz*.
7. **Aseg Stats:** calcula la estadística sobre las estructuras subcorticales segmentadas (*mri/aseg.mgz*). El resultado se guarda en el archivo *stats/aseg.stats*. Este archivo es de gran importancia, pues de él se pueden extraer los datos para un estudio transversal. En estudios longitudinales como el presente todavía serán necesarios otros dos tipos de procesamiento: *Base* y *Long*.
8. **Normalización 2:** se produce una segunda corrección de intensidad utilizando únicamente el volumen del cerebro. El haber quitado el cráneo provoca que esta normalización funcione mejor, y se crea un nuevo volumen *brain.mgz volume*.
9. **Segmentación de materia blanca:** se intenta separar la materia blanca de todo lo demás. Se toma como punto de partida el volumen anteriormente calculado *mri/brain.mgz*. Para ello utiliza restricciones de intensidad, cercanía y suavizado. En este paso se genera el archivo *mri/wm.mgz*.
10. **Corte/relleno:** se crea la masa subcortical para posteriormente calcular la superficie. El hipotálamo (véase figura 4.7) es extraído y los hemisferios se separan. El hemisferio izquierdo se binariza a 255 y el derecho a 127. El volumen de entrada es *mri/wm.mgz* y el volumen generado es *mri/filled.mgz*. Es importante señalar que este es el último paso dentro del procesamiento volumétrico.

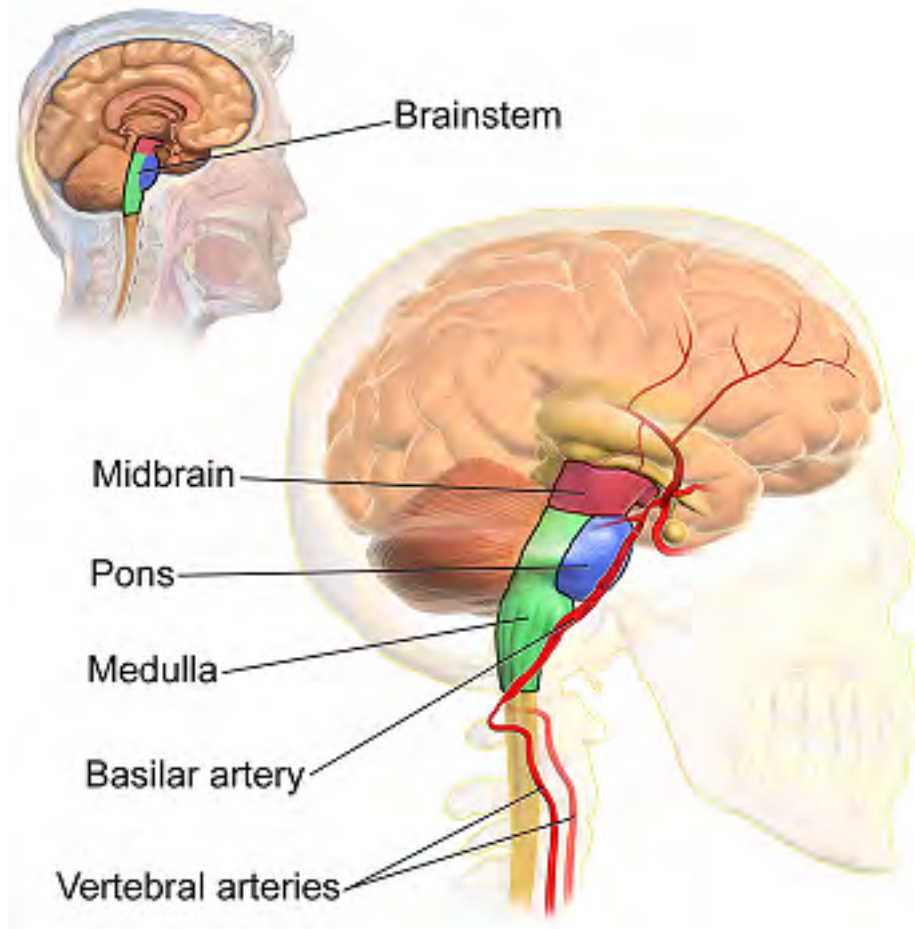


Figura 4.7: Hipotálamo, o midbrain en inglés.

11. **Teselación:** en este paso se crea la superficie orig (i.e. *surf/?h.orig.nofix*). Esto se hace cubriendo el hemisferio con triángulos. Es decir, que se ha pasado de vóxeles a vértices y caras (de volumen a superficie).
12. **Suavizado de superficie Orig:** tras la teselación la forma obtenida es demasiado angular ya que los triángulos se forman en las caras de los vóxeles. Esto es necesario para los procesos de inflado que se verán más adelante. Como resultado se crea la superficie *surf/?h.smoothwm(.nofix)*.
13. **Inflado:** se aplica sobre la superficie *surf/?h.smoothwm(.nofix)* creada anteriormente. Con esto se crea una nueva superficie: *surf/?h.inflated*. Este proceso intenta minimizar la distorsión métrica para que las distancias y áreas se preserven. Esta superficie se utiliza muchas veces para ofrecer resultados. Sin llegar a abstraer tanto la información como la transformación esférica sigue permitiendo librarse de los surcos y valles (véase figura 4.8).

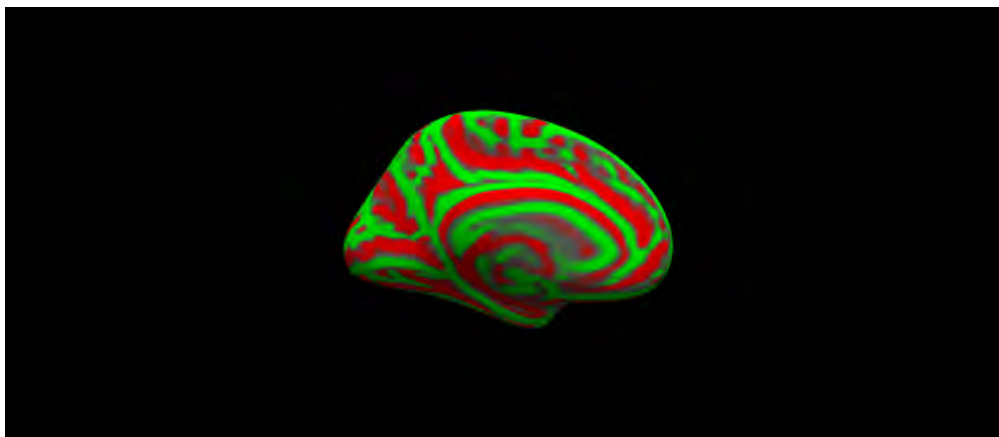


Figura 4.8: Esquema de la representación inflada del cerebro.

14. **Qesfera:** este es el primero de los pasos orientado a la corrección automática de la topología. Es una transformación esférica quasi-homeomórfica de la superficie inflada (véase figura 4.9). Está diseñada para localizar defectos topológicos con el fin de preparar la superficie para la siguiente etapa. Como resultado se crea el archivo *surf/?h.qsphere.nofix* como resultado.

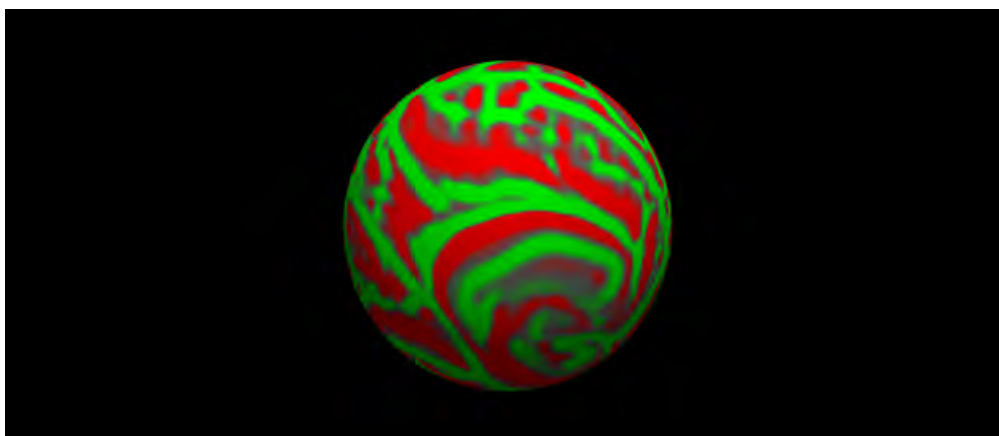


Figura 4.9: Esquema de la representación esférica del cerebro.

15. **Corrector automático de topología:** utilizando la superficie esférica original *surf/?h.qsphere.nofix* encuentra defectos de forma y los elimina. Esto cambia el número de vértices creando la superficie *surf/?h.orig*.
16. **Superficies finales:** crea las superficies *?h.white* y *?h.pial*, el archivo de grosor *?h.thickness* y el archivo de curvatura *?h.curv*. La superficie de materia blanca se crea modificando la superficie original para que siga el gradiente de intensidad entre materia blanca y gris. La superficie pial se crea expandiendo la superficie de materia blanca para que siga el gradiente de intensidad entre la materia gris y los fluidos cerebrospinales (CSF).
17. **Máscara del ribete cortical:** crea una máscara binaria del volumen del ribete cortical. Si el voxel forma parte del ribete cortical se codifica con un 1, y en caso contrario con un 0. El resultado se guarda en *?h.ribbon.mgz*.

18. **Inflado esférico:** la superficie *orig* se infla mientras se mantiene al mínimo la distorsión métrica. Este paso es necesario para registrar la superficie en un atlas esférico. El resultado se guarda en *surf/?h.sphere*.
19. **Registro ipsilateral de superficie (morfología esférica):** registra la superficie *orig* en el atlas esférico a través del recientemente creado archivo *surf/?h.sphere*.
20. **Registro contralateral de superficie (morfología esférica):** paso análogo al anterior, solo que esta vez se toma como referencia el atlas controlateral. Genera los archivos *lh.rh.sphere.reg* y *rh.lh.sphere.reg*.
21. **Curvatura media:** recombina la curvatura media del atlas para que coincida con la del sujeto. Permite mostrar la actividad en la superficie de un sujeto con el patrón de pliegues de un grupo. Genera la superficie *surf/?h.avg_curv*.
22. **Parcelación cortical:** etiqueta cada punto de la superficie cortical (Regiones de interés) a partir de la información del ATLAS. Genera el archivo *label/?h.aparc.annot* (véase figura 4.10).

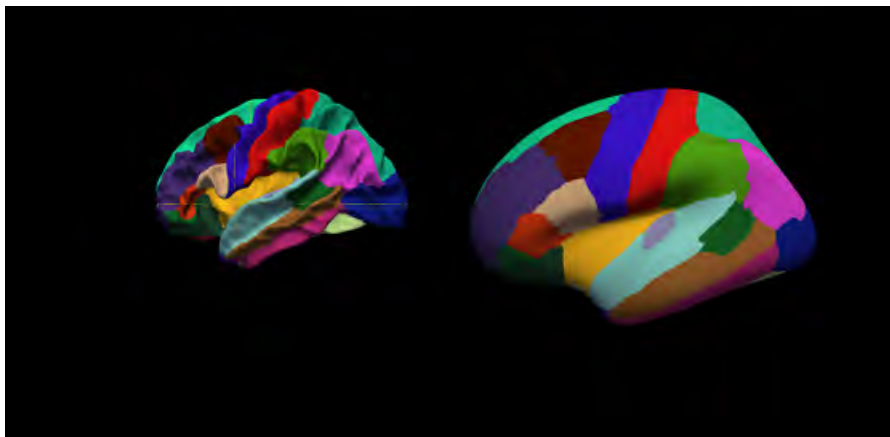
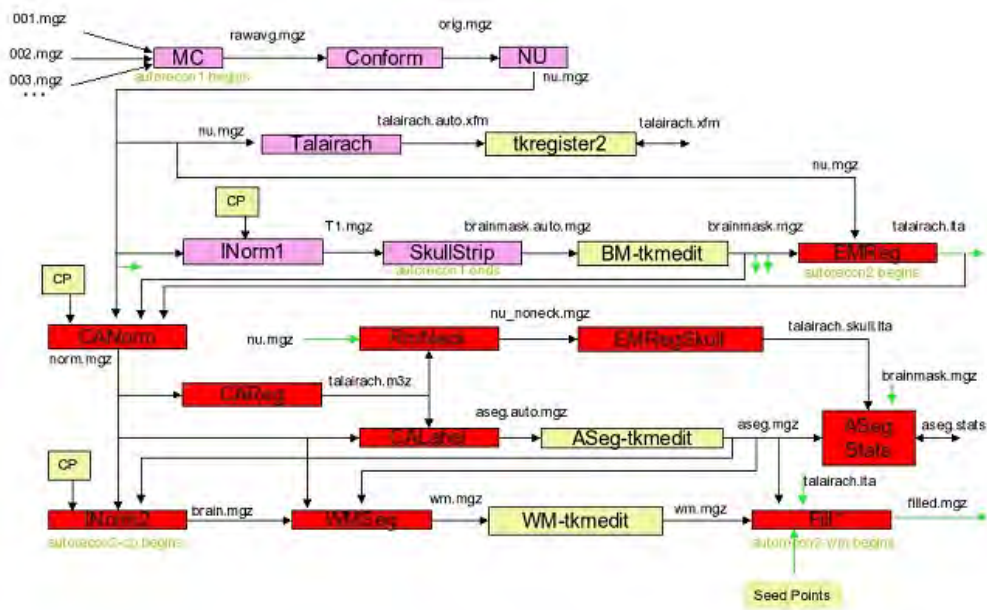


Figura 4.10: Etiquetado del cerebro en ROIs. El archivo *label/?h.aparc.annot* puede utilizarse sobre la superficie original (izquierda) o sobre la superficie inflada (derecha).

23. **Parcelación estadística:** genera un resumen estadístico para cada superficie etiquetada. Estos datos estadísticos incluyen:
 - a) Nombre de la estructura.
 - b) Número de vértices.
 - c) Área total de la superficie (en mm^2).
 - d) Volumen total de materia gris (en mm^3).
 - e) Curvatura media.
 - f) Curvatura gaussiana rectificada.
 - g) Índice de pliegue.
 - h) Índice de curvatura intrínseca.

Esta información se guarda en el archivo *stats/?h.aparc.stats*.



WMSeg includes mri_segment, edit_wm_with_aseg, and mri_pretext
 CP = Control Points
 Fill can have (aseg.mgz&tal.ita) or (tal.xfm,cutting planes) as input, but not both

Figura 4.11: Primera parte del flujo de trabajo de la función recon-all.

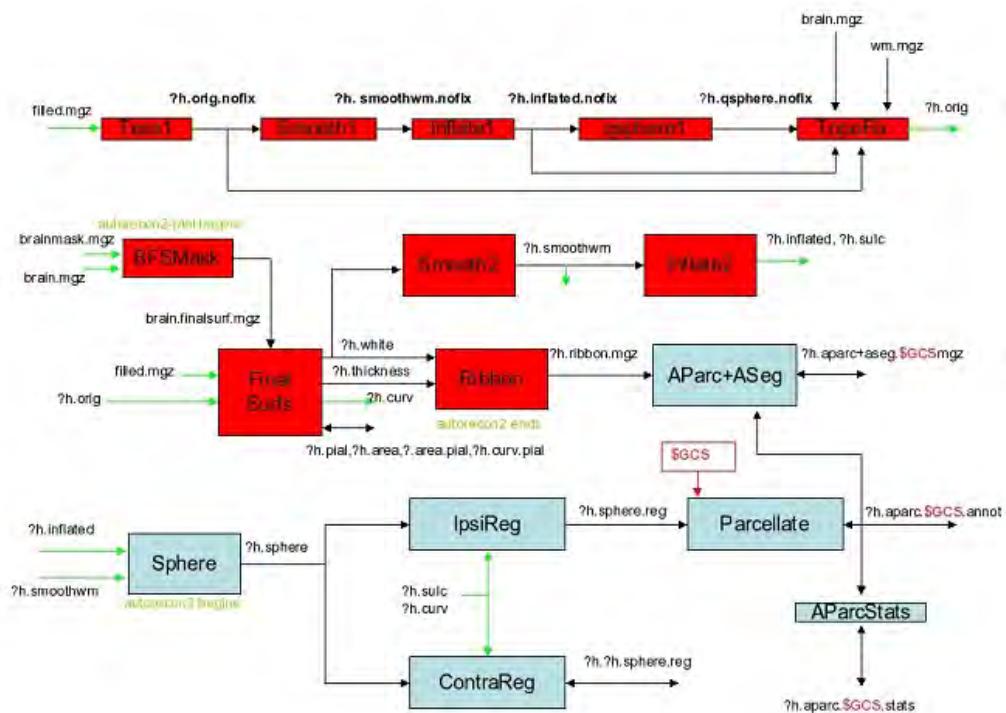


Figura 4.12: Segunda parte del flujo de trabajo de la función recon-all.

Implementación

Una vez entendido el comportamiento de la función recon-all de FreeSurfer, es hora de aplicar dicha función. La estructura del algoritmo utilizado es sencilla, y principalmente se va a encargar de iterar en los archivos .nii buscándolos a partir de los nombres de los archivos .xml.

Para la ejecución de -autorecon1 el archivo brainmask.mgz se utiliza para verificar que todo ha funcionado correctamente, así como para verificar si se ha procesado previamente. En la segunda parte (autorecon2 y autorecon3) se utiliza el archivo aseg.stats.

También es importante remarcar que la herramienta de procesamiento en paralelo ha sufrido cambios, por lo que debe revisarse su aplicación según la versión de MATLAB utilizada. Esto se aplica no sólo a este script, sino a la mayoría de funciones utilizadas.

Por último se ha añadido por defecto el flag -cw256 en la ejecución de recon-all -autorecon1 para garantizar que en caso de tener un *Field of View* excesivamente grande la imagen se corte a 256. Este problema suele ser generado por una mala posición de la imagen MRI.

Algoritmo 2 pseudocódigo de FS_recon_all_cross1

Entrada:

Definir variables de entorno de FreeSurfer.
 pathSubj: carpeta definida como SUBJECTS_DIR en FreeSurfer.
 pathError: carpeta donde se guardarán los archivos en caso de problemas.
 pathId: carpeta donde los archivos .nii.gz han sido guardados anteriormente.

Salida:

Archivos generados por la función recon-all -autorecon1 de FreeSurfer para cada visita de cada sujeto.

- 1: listFichImg = pathId/*.nii.gz
- 2: **para** i = 1 hasta Número de archivos en listFichImg (en paralelo si es posible) **hacer**
- 3: subjDir = pathSubj/listFichImg(i)
- 4: brainmask = subjDir/mri/brainmask.mgz
- 5: **si** no existe brainmask **entonces**
- 6: mksubjdirs subjDir (creación de la carpeta del sujeto)
- 7: niiFile = pathId/listFichImg(i)
- 8: mgzFile = subjDir/mri/orig/001.mgz
- 9: mri_convert -it nii niiFile -ot mgz mgzFile (convertir archivo)
- 10: recon-all -s subjDir -autorecon1 -cw256
- 11: **si** no existe brainmask **entonces**
- 12: mover carpeta del sujeto a pathError
- 13: **fin si**
- 14: **fin si**
- 15: **fin para**

Algoritmo 3 pseudocódigo de FS_recon_all_cross23**Entrada:**

Definir variables de entorno de FreeSurfer.

pathSubj: carpeta definida como SUBJECTS_DIR en FreeSurfer.

pathError: carpeta donde se guardarán los archivos en caso de problemas.

pathId: carpeta donde los archivos .nii.gz han sido guardados anteriormente.

Salida:

Archivos generados por la función recon-all de FreeSurfer para cada visita de cada sujeto.

- 1: listFichImg = pathId/*nii.gz
- 2: **para** i = 1 hasta Número de archivos en listFichImg (en paralelo si es posible) **hacer**
- 3: subjDir = pathSubj/listFichImg(i)
- 4: stats = subjDir/stats/aseg.stats
- 5: **si** no existe stats **entonces**
- 6: recon-all -s subjDir -all (operación de FreeSurfer)
- 7: **si** no existe stats **entonces**
- 8: mover carpeta del sujeto a pathError
- 9: **fin si**
- 10: **fin si**
- 11: **fin para**

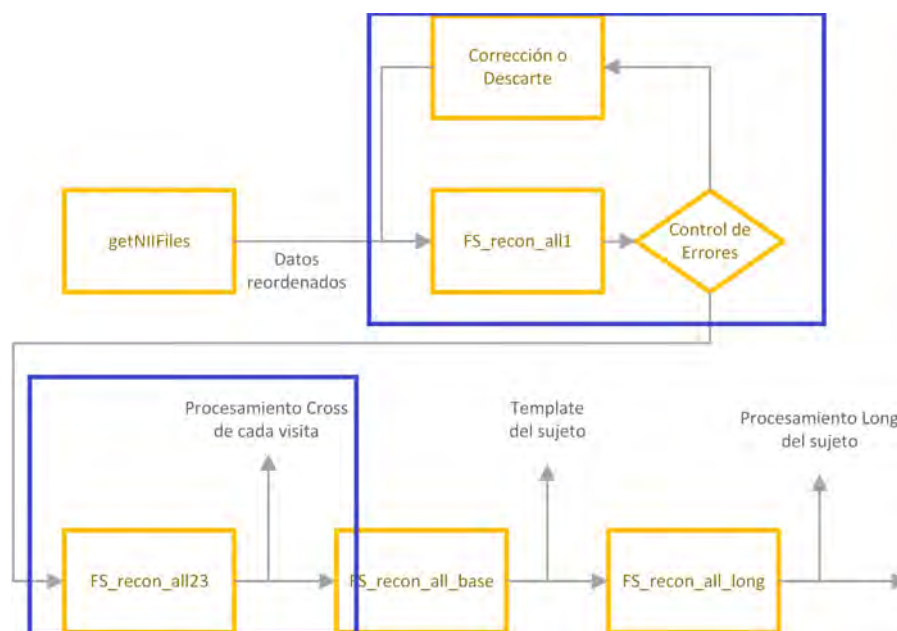


Figura 4.13: La separación en la ejecución entre -autorecon1 y (-autorecon2, -autorecon3) permite evaluar posibles fallos en un estadio temprano, ahorrando una cantidad considerable de tiempo al usuario.

4.2.5. FS_recon_all-base

Explicación

En esta fase se forma una plantilla del sujeto utilizando los distintos resultados obtenidos en el análisis Cross. Esto permite disponer de la anatomía media del sujeto. Tiene como punto de partida el fichero *norm.mgz* de todas las imágenes de un paciente. Es decir: el volumen sin cráneo y con la intensidad normalizada. De nuevo, antes de ver su implementación es conveniente tener en cuenta las distintas

operaciones que se llevan a cabo al ejecutar esta función de FreeSurfer:

1. **Inicialización de la plantilla:** se construye la plantilla utilizando la media o la mediana. Se obtiene `norm.template.mgz`. Tras esto se alinea `norm.mgz` de cada visita con la plantilla. Este registro y su inversa se usará para transferir información entre los distintos puntos a lo largo del tiempo.
2. **Normalización:** obtiene puntos de control (`ctrl_vol.mgz`) y un campo de variación (`bias_vol.mgz`).
3. **Eliminación del cráneo:** mapea el archivo de `brainmask.mgz` creado por `recon-all (cross)` en cada visita y sólo excluye los vóxeles que ninguna de las imágenes del paciente ha considerado como parte del cerebro. Esto permite un ahorro computacional considerable en los siguientes pasos.
4. **Registro EM(GCA):** utiliza el volumen `norm_template.mgz` en lugar de `nu.mgz` para el registro Talairach.
5. **Normalización CA:** utiliza el volumen `norm_template.mgz` en lugar de `nu.mgz` para la normalización.

Implementación

Respecto a la estructura de su implementación en este proyecto, tras una zona donde se definen los distintos directorios utilizados, se ejecutan tres bucles:

1. En el primero se almacena el índice de los subdirectorios que son válidos. Esto quiere decir que no se almacena el puntero al directorio padre (`.`) o al mismo directorio (`.`), así como los enlaces a carpetas dentro de `/usr/local/FreeSurfer`.
2. El segundo bucle se encarga de agrupar los ficheros y preparar todo para la creación de la plantilla o `template`. Esto incluye guardar el nombre de los archivos que se quieren procesar en una tabla.
3. El último bucle se encarga de, finalmente, crear la plantilla. Aquí también debe tenerse en cuenta el número de imágenes de las que se dispone.

Es importante mencionar que los resultados se guardarán en el mismo directorio que `recon-all`, que es el definido como `SUBJECTS_DIR` en FreeSurfer. Los archivos serán separados tras ejecutar el script correspondiente a `FSrecon_all.long`.

Algoritmo 4 pseudocódigo de FS_recon_all_base

Entrada:

Definir variables de entorno de FreeSurfer.

pathSubj: carpeta definida como SUBJECTS_DIR en FreeSurfer.

Salida:

Archivos generados por la función recon-all -base -all de FreeSurfer con la plantilla para cada sujeto.

- 1: listImg = cell(1,número máximo de imágenes planeado);
 - 2: listFichImg = ficheros en pathSubj
 - 3: Eliminar ficheros no válidos de listFichImg.
 - 4: **para** i=1 hasta número de ficheros válidos en listFichImg **hacer**
 - 5: Guardar en listImg las referencias a las carpetas de cada visita. Las filas definen distintos sujetos. Las columnas, distintas visitas del mismo sujeto.
 - 6: **fin para**
 - 7: **para** i = 1 hasta número de sujetos (en paralelo si es posible) **hacer**
 - 8: aux = listImg(i,1). El primer parámetro que se prepara para la función es el nombre del sujeto.
 - 9: **para** j = 1 hasta número máximo de imágenes planeado **hacer**
 - 10: **si** listImg(i,j) no está vacío **entonces**
 - 11: aux=[aux, ' -tp ' pathSubj,listImg(i,j)]. se concatenan las sucesivas visitas del sujeto a la cadena de parámetros para ejecutar después la función de FreeSurfer.
 - 12: **fin si**
 - 13: **fin para**
 - 14: recon-all -base aux -all. Se entiende como aux a la cadena entera de parámetros generada en el anterior bucle for.
 - 15: **fin para**
-

4.2.6. FS_recon_all-long**Explicación**

Esta es la última fase del proceso de tratamiento de imágenes antes de comenzar la obtención de datos. Se trata del tratamiento longitudinal (teniendo en cuenta la correspondencia en el tiempo de las imágenes de un mismo paciente).

Los pasos seguidos por esta función de FreeSurfer son similares a los del procedimiento recon-all. En este caso, se utiliza como punto de partida algunos datos calculados en la primera etapa (cross) y en la segunda (base):

1. **Imagen de entrada:** se copia el volumen 00?.mgz generado en recon-all.
2. **Corrección de movimiento:** se mapea el susodicho volumen al espacio del generado en la etapa base. De manera análoga al procedimiento cross, al generar la media de los dos se obtiene *orig.mgz*.
3. **Corrección de intensidad no uniforme:** se trata del mismo procedimiento seguido en el procesamiento cross.
4. **Talairach:** copia el *talairach.xfm* del template.
5. **Normalización:** mapea y usa los puntos de control en control.dat del procesamiento cross. Si ya existían estos puntos de una vez anterior no se sobrescribirán.
6. **Quitar el cráneo:** copia el *brainmask.mgz* del template (véase figura 4.14). Se utiliza para crear una máscara *T1.mgz* y obtener así la máscara final.

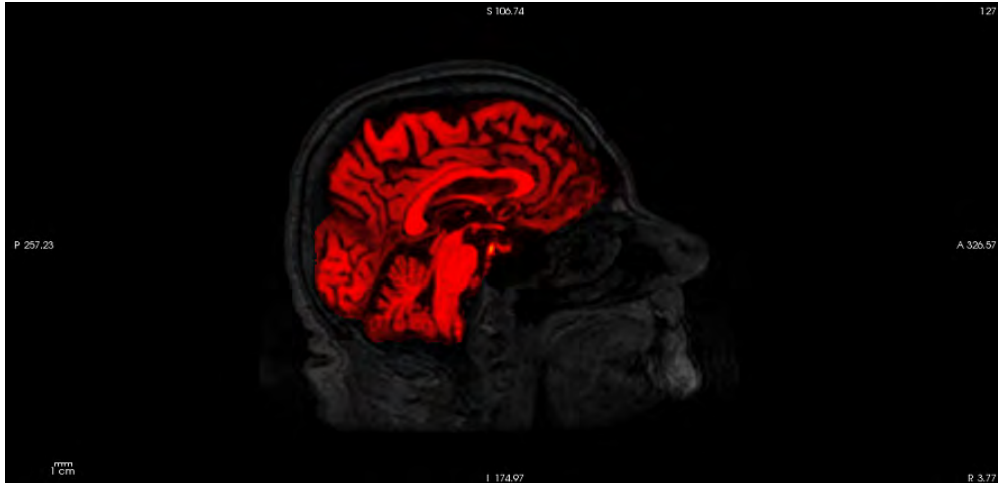


Figura 4.14: Volumen brainmask.mgz, resaltado en rojo, frente a la imagen original orig.mgz.

7. **Registro EM(GCA)**: copia el *talairach.lta* del template.
8. **Normalización CA**: se inicializa con el archivo *aseg.mgz* del template copiado en la etapa anterior.
9. **Registro CA no lineal**: utiliza *talairach.m3z* del template como punto de partida.
10. **Registro CA no lineal inverso**: proceso análogo al hecho en recon-all.
11. **Quitar cuello**: proceso análogo al hecho en recon-all.
12. **Registro EM con cráneo pero sin cuello**: proceso análogo al hecho en recon-all.
13. **Etiquetado CA**: copia la transformación lineal de la visita al template. Tras esto se crea el archivo *aseg.fused.mgz* mapeando e incorporando información de la segmentación de todas las visitas. De esta manera se reducen los errores. Finalmente utiliza esta combinación como punto de partida para el último etiquetado.
14. **Normalización 2**: proceso análogo al hecho en recon-all.
15. **Máscara de la superficie final del cerebro**: ajusta cambios realizados manualmente. Si no se han realizado cambios se copia de la etapa de cross.
16. **Segmentación de materia blanca**: se aplica únicamente a los cambios realizados manualmente. En caso contrario se copia de la etapa de cross.
17. **cortar/rellenar**: proceso análogo al hecho en recon-all.
18. **Superficies finales**: copia y utiliza *?h.white* y *?h.pial* del template para inicializar las superficies white, pial y orig. Esto asegura que todas las superficies están indirectamente registradas a lo largo de los puntos de tiempo. Esto quiere decir que el número de vértices concuerda.

19. **Volumen de la superficie:** proceso análogo al hecho en recon-all.
20. **Suavizado de superficie orig 2:** proceso análogo al hecho en recon-all.
21. **Inflado 2:** proceso análogo al hecho en recon-all.
22. **ASeg Stats:** proceso análogo al hecho en recon-all.
23. **Inflado esférico:** el archivo ?h.sphere es copiado del template.
24. **Registro de superficie Ipsilateral (morfología esférica):** proceso análogo al hecho en recon-all.
25. **Jacobiano:** proceso análogo al hecho en recon-all.
26. **Curvatura media:** proceso análogo al hecho en recon-all.
27. **Parcelación cortical:** proceso análogo al hecho en recon-all.
28. **Estadísticos de parcelación:** proceso análogo al hecho en recon-all.
29. **Parcelación cortical 2:** proceso análogo al hecho en recon-all.
30. **Máscara del ribete cortical:** proceso análogo al hecho en recon-all.
31. **Agregar la parcelación al ASeg:** proceso análogo al hecho en recon-all.
32. **Actualizar WMparc:** proceso análogo al hecho en recon-all.

Implementación

En la implementación de MATLAB utilizada en este estudio no sólo se aplica la función de FreeSurfer, sino que también se separan los archivos que se han obtenido en estas tres etapas en distintas carpetas. Para ello se vuelve a recurrir a tres bucles principales.

El primero se encarga de ejecutar el comando de FreeSurfer, el segundo se encarga de desvincular los ficheros no válidos de la carpeta principal. El último se encarga de distribuir los ficheros generados entre distintas carpetas, dependiendo de si se corresponden con el procesamiento *cross*, *base* o *long*.

Algoritmo 5 pseudocódigo de FS_recon_all_long

Entrada:

Definir variables de entorno de FreeSurfer.

pathSubj: carpeta definida como SUBJECTS_DIR en FreeSurfer, donde se encuentran los archivos provenientes de etapas anteriores de recon-all.

BaseDir: carpeta donde se moverán los resultados de recon-all -base al acabar. LongDir: carpeta donde se moverán los resultados de recon-all -long al acabar.

Salida:

Archivos generados por la función recon-all -long -all de FreeSurfer con la plantilla para cada sujeto.

```

1: listImg = directorios en pathSubj.
2: para i = 1 hasta número de subdirectorios de listImg hacer
3:   si número de caracteres del nombre de listImg(i) es mayor que 15 entonces
4:     template = pathSubj/listImg(i).name(1:10)
5:     img = pathSubj/listImg(i).name
6:     recon-all -long img template -all
7:   fin si
8: fin para
9: para i = 1 hasta el número de directorios en pathSubj hacer
10:  si es un fichero válido entonces
11:    unlink pathSubj listImg(i)
12:  fin si
13: fin para
14: listImg = directorios en pathSubj (ahora actualizado)
15: para i = 3 hasta número de ficheros en listImg hacer
16:  orig = pathSubj/listImg(i)
17:  si longitud del nombre listImg(i) == 10 entonces
18:    dest = BaseDir/listImg(i)
19:    mover archivo orig a dest
20:  si no, si listImg(i).name contiene 'long' entonces
21:    dest = LongDir/listImg(i)
22:    mover archivo orig a dest
23:  fin si
24: fin para

```

4.2.7. getAsegAparcADNI

Esta etapa se encarga de extraer la información estadística correspondiente a las estructuras corticales y subcorticales, además de la información demográfica (edad, sexo, tests neuropsicológicos, etc.).

El script se puede utilizar tanto para extraer los datos del procesamiento cross como los provenientes del procesamiento long. Lo único que hay que hacer es comentar o descomentar parte del código. Sin embargo, para este estudio se ha escogido el procesamiento Long como base para las medidas de las ROIs. En cualquier caso, queda descrito el algoritmo completo en el pseudocódigo 6.

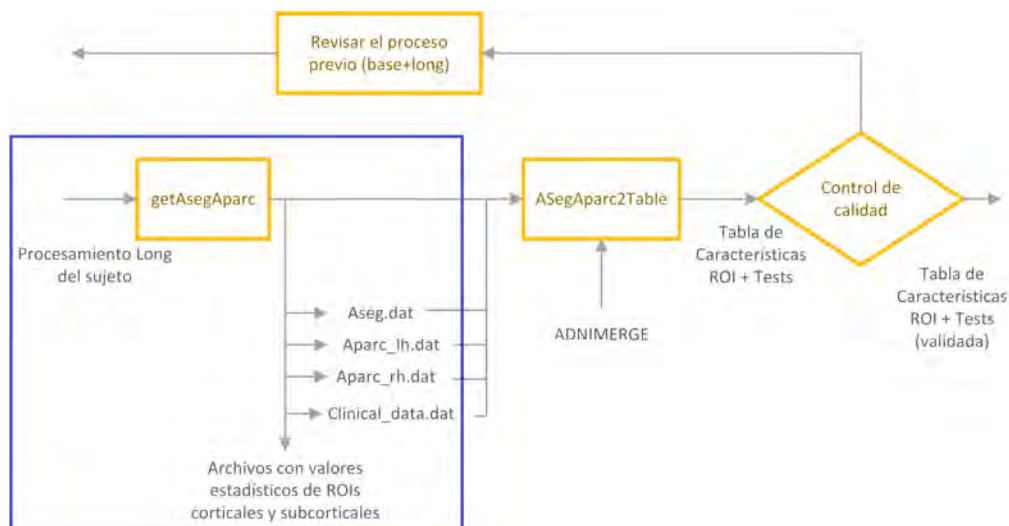


Figura 4.15: Esquema de los archivos de entrada y salida en el flujo de trabajo de getAsegAparcADNI.

El primer paso consiste en localizar los archivos xml (información demográfica y clínica), así como los directorios Long o Cross, dependiendo de la modalidad. Estos son los archivos básicos que se van a utilizar como entrada del algoritmo.

Respecto a la implementación en MATLAB, es necesario decir que se recurre a una serie de funciones auxiliares que son compatibles con distintas versiones de MATLAB y sistemas. Esto facilita la portabilidad. Sin embargo, también se recurre a una función de FreeSurfer llamada `asegstats2table()` que requiere tener instalado el susodicho software.

La más importante de estas funciones auxiliares portables es `QdecTable()`. Esta función genera el archivo con los datos demográficos que usará `stat2table()` para navegar por los directorios de los sujetos y extraer la información de las ROIs almacenada en distintos archivos. Es, por tanto, un paso previo a las tres llamadas a `stat2table()`.

Dentro de `QdecTable()` se hacen ciertas llamadas a otras funciones secundarias. De estas, las más importantes son `parseXML()`, `getPatientsInfo()`, `sortData()` y `replace()`:

- La función `parseXML()` se encarga de traducir un documento XML a una estructura de MATLAB. Esto permite una navegación sencilla a través de la estructura de etiquetas. Esta función proviene de la propia web de MathWorks (<https://es.mathworks.com/help/MATLAB/ref/xmlread.html>).
- `getPATientsInfo()` se encarga de obtener de esta estructura los datos correspondientes a las distintas características demográficas de interés.
- `sortData()` proviene de la Toolbox para MATLAB ofrecida por Bernal [2] y

permite ordenar los sujetos en el tiempo, asegurando así un orden establecido para el posterior análisis de datos.

- Finalmente la función `replace()` se encarga de sustituir 'fsidbase' por 'fsid-base' en el archivo de texto generado. Este cambio permite el correcto funcionamiento de `asegstats2table()`, que espera una cabecera llamada 'fsid-base'. Este cambio no se puede realizar en MATLAB directamente puesto que la sintaxis del lenguaje M no permite ese tipo de nombre ni para nombrar variables ni para cabeceras de una tabla.

Una vez obtenido el archivo con los datos clínicos se producen tres llamadas a `asegstats2table()` para obtener los valores estadísticos de las estructuras corticales (Aseg) y subcorticales de ambos hemisferios (Aparc-lh, Aparc-rh). Las llamadas a estas funciones quedan mejor descritas en el pseudocódigo 6.

Algoritmo 6 pseudocódigo de `getAsegAparcADNI`

Entrada:

Definir variables de entorno de FreeSurfer.

`xmlPath`: carpeta donde se encuentran los archivos xml.

`destination_path`: carpeta donde se almacenarán los resultados generados tanto desde el procesamiento cross como long.

`path_subj`: carpeta donde se encuentran los datos procesados (Cross o Long).

Salida:

Para cada tipo de procesamiento (Cross o Long):

`clinicaldata_tipoproc.dat`: archivo con información demográfica.

`aseg_tipoproc.clinicaldata.dat`: archivo con estadísticos de estructuras subcorticales.

`aparc_tipoproc_lh.clinicaldata.dat`: archivo con estadísticos de estructuras corticales del hemisferio izquierdo.

`aparc_tipoproc_rh.clinicaldata.dat`: archivo con estadísticos de estructuras corticales del hemisferio derecho.

→ Tomando como input el procesamiento Cross.

1: `SUBJECTS_DIR = path_subj`.

2: `name_dat = 'clinicaldata_cross.dat'`

3: `typeProc = 'Cross'`

4: `QdecTable(xmlPath, path_subj, typeProc, name_dat)`. Se interpreta la estructura XML, guardando así los datos demográficos.

5: `asegstats2table -qdec name_dat -t cross_dest_path aseg_cross.name_dat`

6: `aparcstats2table -qdec name_dat -t cross_dest_path aparc_cross_rh.name_dat -hemi rh -meas thickness`

7: `aparcstats2table -qdec name_dat -t cross_dest_path aparc_cross_rh.name_dat -hemi rh -meas thickness`

8: mover archivo `name_dat` (información demográfica) a `destination_path`

→ Tomando como input el procesamiento Long.

9: `SUBJECTS_DIR = path_subj`.

10: `typeProc = 'Long'`

11: `name_dat = 'clinicaldata_long.dat'`

12: `QdecTable(xmlPath, path_subj, typeProc, name_dat)`;

13: `asegstats2table -qdec-long name_dat -t long_dest_path aseg_long.name_dat`

14: `aparcstats2table -qdec-long name_dat -t long_dest_path aparc_long_lh.name_dat -hemi lh -meas thickness`

15: `aparcstats2table -qdec-long name_dat -t long_dest_path aparc_long_rh.name_dat -hemi rh -meas thickness`

16: mover archivo `name_dat` (información demográfica) a `destination_path`

4.2.8. asegAparc2table

Se trata de un paso intermedio que se encarga de traducir la estructura extraída desde getAsegAparc en una tabla en la que se condensa toda la información relevante para el aprendizaje automático. Esto incluye la adición en una tabla de las medidas estadísticas de espesor y volumen en las ROIs generadas por FreeSurfer.

Además, a partir de la tabla ADNIMERGE se extraen los valores de varios tests neuropsicológicos. Es por ello que requiere ejecutar previamente el código Take_ADNIMERGE. Este último es un script sencillo que se encarga de cambiar el archivo .csv original en una tabla a la que se puede acceder más fácilmente.

Respecto al funcionamiento interno, aparece descrito en el algoritmo 7. En esencia lee el archivo con los datos demográficos y las medidas tomadas en los archivos aseg y aparc.

Por último los une en una nueva tabla. Antes de guardar el resultado se pasa un filtro en el que se verifica que las visitas procesadas sean válidas. Esto quiere decir que su diagnóstico real está incluido en la tabla de ADNIMERGE y, en el caso de estudiar sMCI vs. pMCI, son rechazados todos los sujetos que no pertenezcan a estos dos grupos.

En este código se utilizan funciones de FreeSurfer para MATLAB. Es por ello que se ha creado una carpeta para alojar estas funciones. Este patrón se va a repetir en distintos scripts, y permite poder ejecutarlos incluso sin tener FreeSurfer instalado en la máquina.

Algoritmo 7 pseudocódigo de asegAparc2table

Entrada:

ADNIMERGE: Tabla con los datos proporcionados por ADNIMERGE.
 clinicaldata.dat: archivo con información demográfica.
 aseg.clinicaldata.dat: archivo con estadísticos de estructuras subcorticales.
 aparc_lh.clinicaldata.dat: archivo con estadísticos de estructuras corticales del hemisferio izquierdo.
 aparc_rh.clinicaldata.dat: archivo con estadísticos de estructuras corticales del hemisferio derecho.

Salida:

Tresults: Tabla con la información de ROIs+Tests neuropsicológicos

- 1: Tdec = almacena datos clínicos en una tabla.
- 2: TMarkers = almacena datos de aseg y aparc en una tabla
- 3: Tresults = [Tdec,TMarkers]
- 4: **si** estudio MCI **entonces**
- 5: Selección únicamente de sujetos que en ADNIMERGE estén reconocidos como MCI.
- 6: **si no**
- 7: Selección únicamente de sujetos que en ADNIMERGE estén reconocidos como AD o NC.
- 8: **fin si**
- 9: Se guarda la tabla Tresults en un archivo .mat
- 10: Se mueve Tresults a la carpeta de salida (DATA).

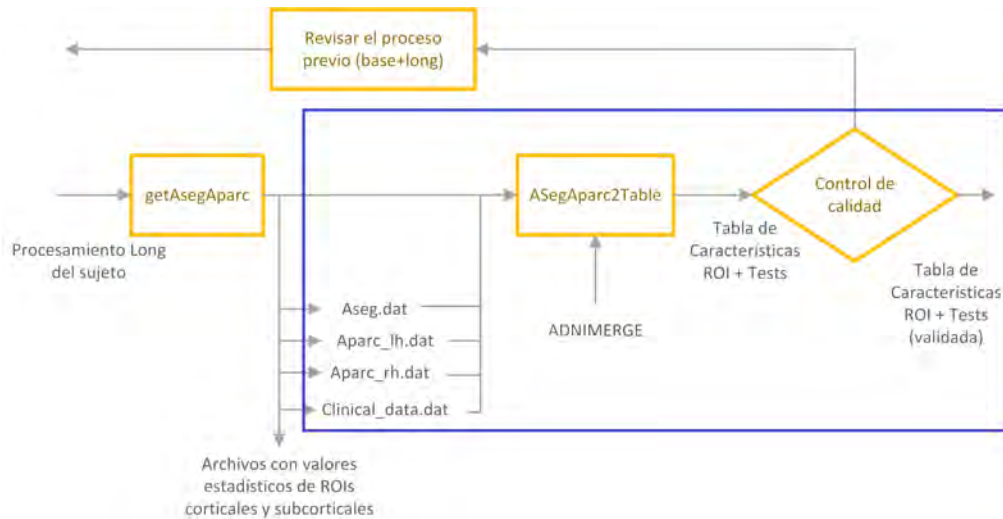


Figura 4.16: Esquema de los archivos de entrada y salida en el flujo de trabajo de AsegAparc2Table, donde se añaden los archivos provenientes de ADNIMERGE. En concreto, es interesante el añadido de los tests neuropsicológicos.

4.2.9. Control de calidad

Con el fin de comprobar que los datos son fiables y que se han procesado correctamente se procede a realizar un control de calidad con los datos de la tabla Tresults. Este control consiste esencialmente en una comparación entre los resultados del procesamiento y los resultados obtenidos por Bernal [2]. Al final del proceso se comparan dos características:

1. La atrofia del volumen hipocampal.
2. La atrofia de la corteza entorrinal.

Sin embargo, no sólo se comparan sus valores medios y desviaciones estándar para cada grupo (sMCI,pMCI) sino que se realizan una serie de representaciones gráficas.

A partir de estas representaciones es sencillo comprobar a simple vista si los resultados concuerdan con lo obtenido por Bernal hasta cierto punto o si hay un error inesperado.

Dentro de estas representaciones, tal vez una de las más importantes sea la ya mencionada atrofia para sMCI y pMCI. Este cambio en el ritmo de la atrofia ya ha sido descrito en estudios previos [19]. El resultado tras ejecutar el control de calidad es el mostrado en las figuras 4.17 y 4.18. Nótese que en este caso lo que más importa es que la pendiente (atrofia) sea similar.

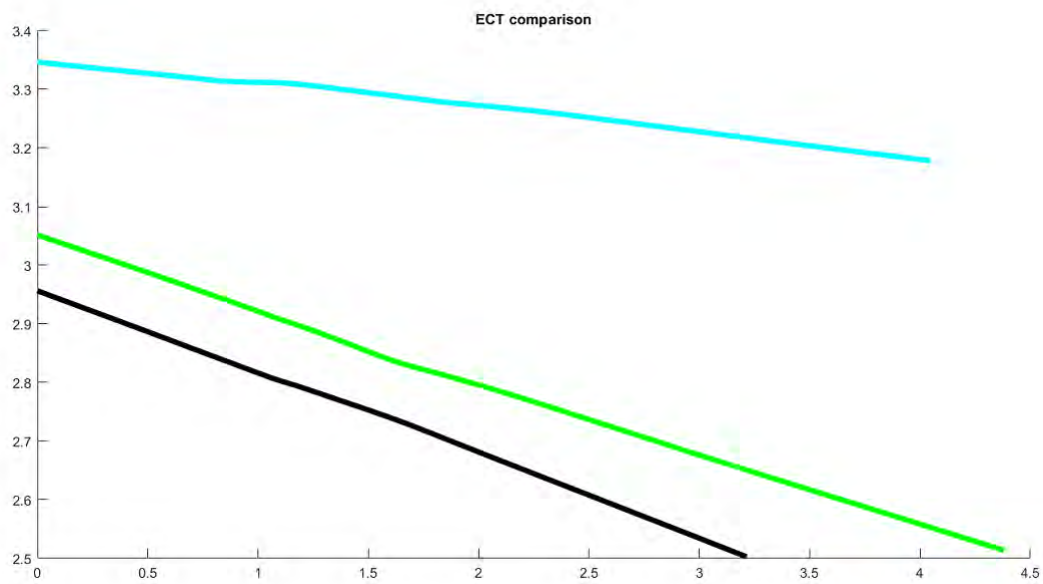


Figura 4.17: Gráfica del control de calidad de la atrofia de la corteza entorrinal comparando el estudio de Bernal con el presente. Los datos del estudio de Bernal se muestran en azul (sMCI) y en verde (pMCI). Los datos del estudio actual se muestran en rosa (sMCI, si hubiera) y en negro (pMCI).

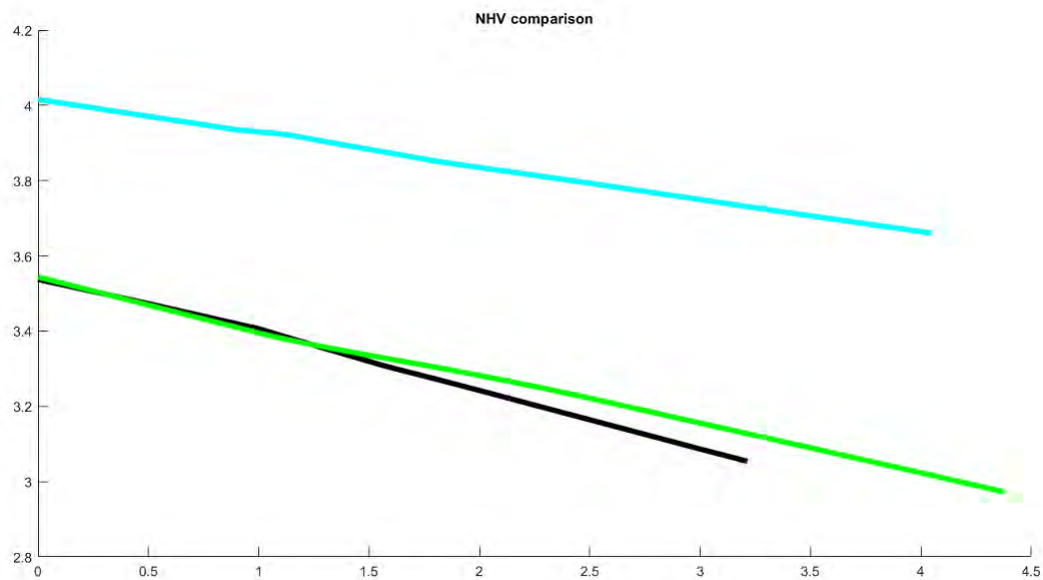


Figura 4.18: Gráfica del control de calidad de la atrofia del volumen hipocampal comparando el estudio de Bernal con el presente. Los datos del estudio de Bernal se muestran en azul (sMCI) y en verde (pMCI). Los datos del estudio actual se muestran en rosa (sMCI, si hubiera) y en negro (pMCI).

Por otro lado, se muestra una gráfica con histogramas sobre los datos en el baseline y la atrofia. Esto se muestra tanto para el volumen hipocampal como para la corteza entorrinal (véanse figuras 4.20 y 4.19).

Este control es algo más exhaustivo y permite comprobar que la dispersión en la población es similar. En este caso se busca que la forma aproximada de los histogramas sea parecida, salvando la diferencia en el número de sujetos de ambos estudios.

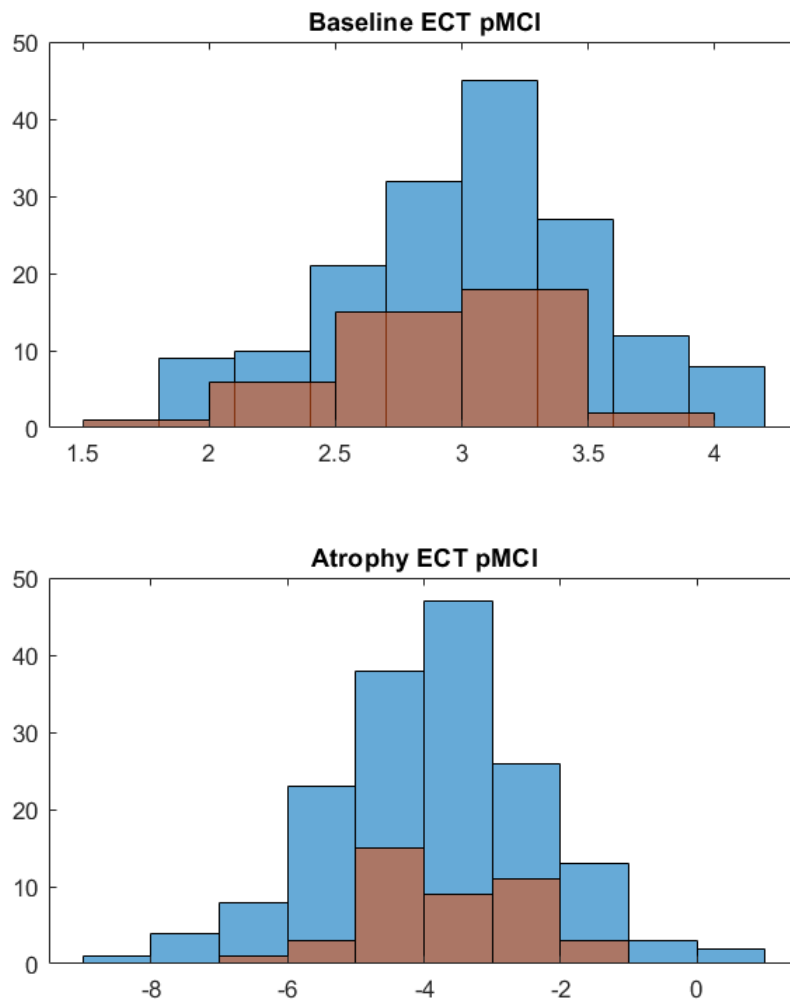


Figura 4.19: Gráfica del control de calidad de la corteza entorrinal. Los valores del estudio de Bernal se muestran en azul. Los valores del presente estudio se muestran en rojo. El objetivo es la verificación de la dispersión de la población. El eje vertical no es un porcentaje, sino una cuenta del número de sujetos.

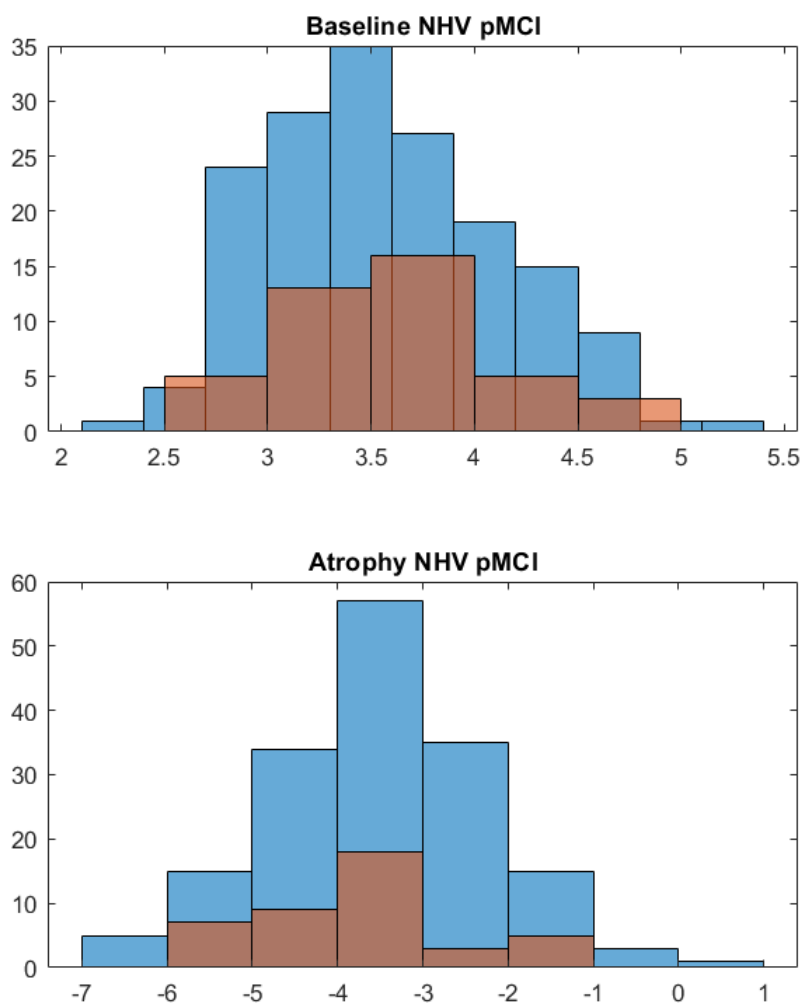


Figura 4.20: Gráfica del control de calidad del volumen hipocampal. Los valores del estudio de Bernal se muestran en azul. Los valores del presente estudio se muestran en rojo. El objetivo es la verificación de la dispersión de la población. El eje vertical no es un porcentaje, sino una cuenta del número de sujetos.

Además, existe otra representación igualmente importante, y es la del mínimo coeficiente de correlación. Ésta se basa en el principio que explica que dos visitas muy próximas en el tiempo deben tener una correlación muy alta.

Este acercamiento ha provocado la división del código en dos partes: una para obtener estas correlaciones y otra para llevar a cabo el control como tal.

Esta parte del control de calidad sólo es relevante para medir el procesamiento longitudinal, ya que se necesita que las imágenes 'Cross' estén alineadas con la plantilla 'Base'. Si el procesamiento longitudinal ha tenido éxito se espera que el mínimo coeficiente de correlación entre imágenes de un mismo sujeto no sea muy alto (véase figura 4.21).

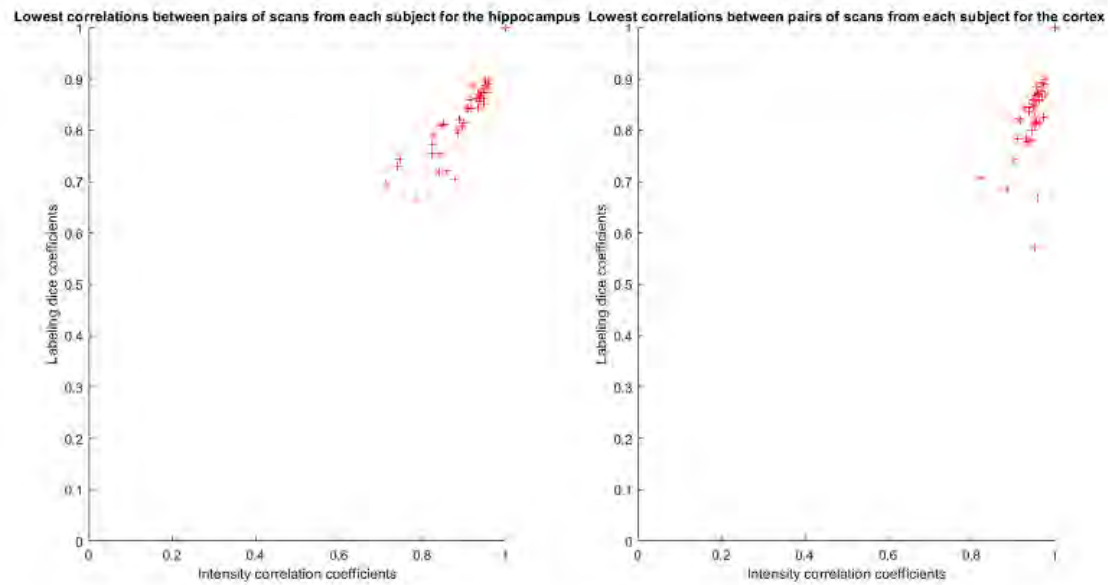


Figura 4.21: Al ofrecer en una gráfica la correlación mínima de parejas de imágenes de un mismo sujeto para todos los sujetos se puede ver cuáles han podido sufrir un fallo en el procesamiento Long.

Este control es importante ya que, una vez añadida la información del análisis masivo, se tendrá esta información como el punto de partida del análisis estadístico (véase sección 4.3).

4.2.10. Análisis masivo

Esta parte de la extracción de datos requiere haber realizado un procesamiento previo de las imágenes de los sujetos. Toma como punto de partida el valor del grosor cortical en cada vértice de la malla. Para entender completamente este procesamiento debe asegurarse el lector de tener claros los conceptos estadísticos explicados en el anexo A.

Esta etapa se divide en cuatro partes principales. Aunque cada una de ellas se describirá con más detalle, se mencionan aquí los aspectos más importantes:

1. Almacenar la información del grosor cortical para cada vértice. En realidad, esta información se genera en el procesamiento de FreeSurfer. Sin embargo, es necesario transformar todas las superficies a un espacio común de vértices para poder comparar el grosor cortical en cada uno de ellos.
2. Realizar un mapa FDR con los q-valores. En esencia es un mapa estadístico del córtex en el que se compara un grupo con otro. Para no caer en una inferencia estadística, este mapa se ha realizado con sujetos del estudio de Bernal [2].

- Hallar los clusters. A partir de este mapa estadístico se definen una serie de clusters y se aplica una máscara a cada uno de ellos.

Estos clusters pueden verse como unas ROIs personalizadas que se ciñen mejor al deterioro local, no guiándose por la estructura anatómica intuitiva sino por las diferencias de atrofia cortical.

- Calcular los estadísticos. Al igual que con las ROIs se debe calcular la media del espesor cortical en cada uno de estos clusters para cada visita de cada sujeto. Tras esto se ponen estos datos en forma de tabla.

Por último, se unirán a la tabla Tresults, que hasta este momento sólo tenía almacenada la información de ROIs y tests neuropsicológicos.

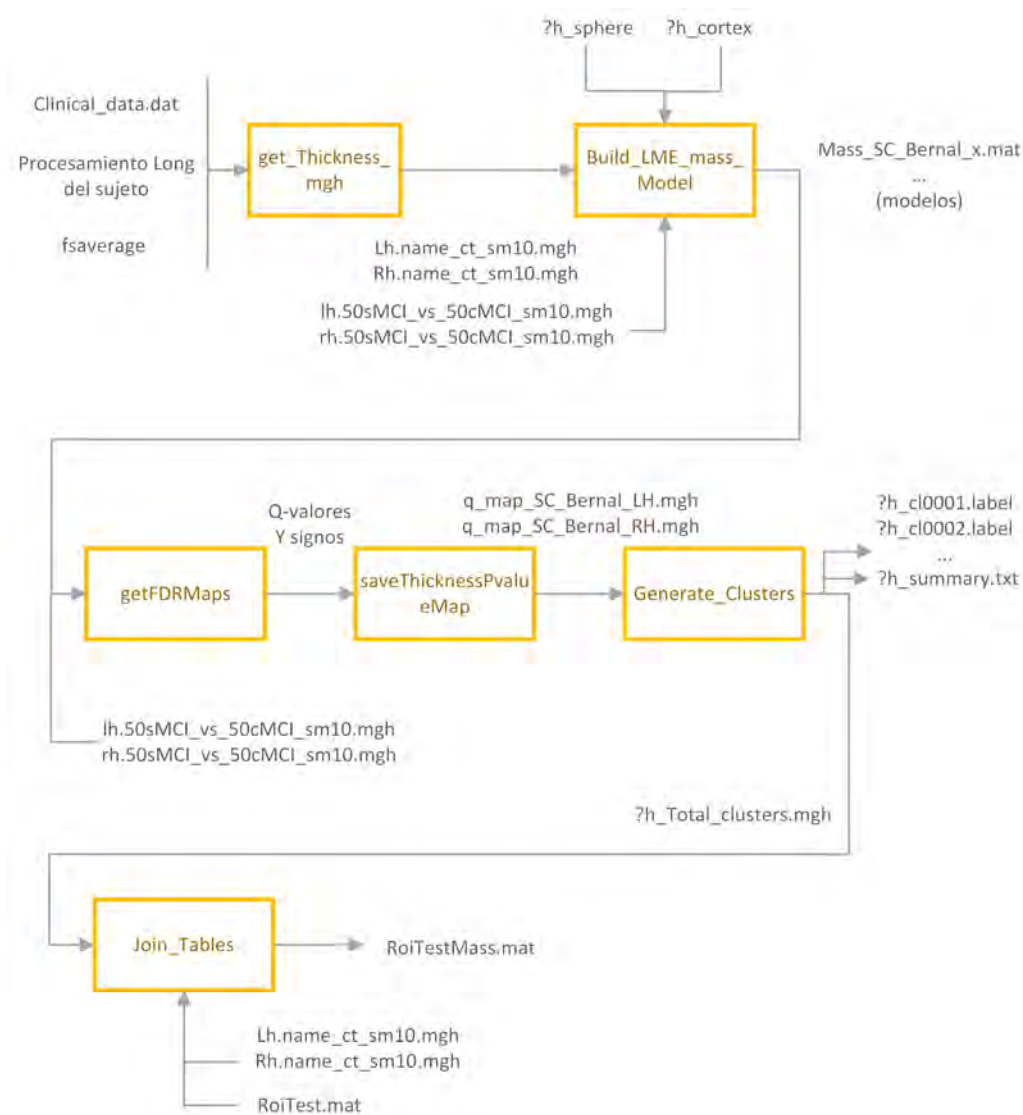


Figura 4.22: Esquema del flujo de trabajo del análisis masivo.

Extracción del grosor cortical

En la implementación de este paso se encarga el script `get_Thickness_mgh` (Véase pseudocódigo 8). Una explicación bastante detallada se puede encontrar en la propia web de FreeSurfer (<https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/GroupAnalysisV6.0>). Sobre todo se puede encontrar información acerca de los dos comandos principales: `mris_preproc` y `mri_surf2surf` (véase figura 4.23).

- **`mris_preproc`** se encarga de reunir los datos en un mismo espacio. En concreto se ha elegido el sujeto `fsaverage` como espacio común. Este sujeto está incluido en las carpetas internas de FreeSurfer y posee alrededor de 160.000 vértices. El comando `mris_preproc` también se encarga de juntar todos estos datos en un mismo archivo `.mgh`. En realidad, este proceso se repite para cada hemisferio, por lo que se tienen dos archivos llamados `lh.thickness.mgh` y `rh.thickness.mgh`.
- **`mri_surf2surf`** se encarga de realizar un suavizado sobre esta malla manteniendo la correspondencia de los vértices con los de `fsaverage`. Como resultado se obtienen los archivos `lh.name_ct_smX.mgh` y `rh.name_ct_smX.mgh`, donde `X` es el número de iteraciones utilizadas para el suavizado.



Figura 4.23: Esquema de los archivos de entrada y salida en el flujo de trabajo de `get_Thickness_mgh` donde se agrupan los valores de grosor cortical.

Este paso es necesario para tener los datos del grosor cortical en una única estructura para todo el procesamiento posterior. Se han procesado dos conjuntos de datos en total:

- El grupo de datos estudiado. De aquí se obtienen los archivos `.mgh` con el grosor cortical de cada sujeto y visita para todos los vértices de la población estudiada.

- Un grupo de 50 sujetos sMCI y 50 sujetos pMCI sacados del estudio de Bernal [2]. Este grupo se utilizará para el estudio estadístico y así no entrar en inferencia con los datos.

En definitiva, los clusters hallados con los datos de estos sujetos simplemente se usarán como máscara para los sujetos de la población estudiada. Esto evita que las diferencias en el grosor cortical estén excesivamente adecuadas a la población estudiada. A los archivos generados se les ha llamado *lh.50sMCI_vs_50cMCI-long_thickness_sm10* y *rh.50sMCI_vs_50cMCI-long_thickness_sm10*.

Algoritmo 8 pseudocódigo de `get.Thickness_mgh`

Entrada:

`clinicaldata.dat`: archivo qdec con la información clínica.
archivos de procesamiento Long

Salida:

`lh.name_ct_sm10.mghTresults`: archivo con los datos del grosor cortical para cada visita y para todos los vértices del sujeto común (`fsaverage`) del hemisferio izquierdo.
`rh.name_ct_sm10.mghTresults`: archivo con los datos del grosor cortical para cada visita y para todos los vértices del sujeto común (`fsaverage`) del hemisferio derecho.

- 1: `qdecfile = clinicaldata.dat`
 - 2: Enlace simbólico a `$FreeSurfer_HOME/subjects/fsaverage`
{Almacenar información del grosor cortical de cada visita.}
 - 3: `mrprep -qdec-long qdecfile -target fsaverage -hemi lh -meas thickness -out lh.thickness.mgh`
 - 4: `mrprep -qdec-long qdecfile -target fsaverage -hemi rh -meas thickness -out rh.thickness.mgh`
{Proyectar de nuevo al espacio `fsaverage`, realizando un suavizado.}
 - 5: `mri_surf2surf -hemi lh -s fsaverage -sval lh.thickness.mgh -tval lh.name_ct_sm10.fs.mgh -fwhm-trg 10 -cortex -noreshape`
 - 6: `mri_surf2surf -hemi rh -s fsaverage -sval rh.thickness.mgh -tval rh.name_ct_sm10.fs.mgh -fwhm-trg 10 -cortex -noreshape`
 - 7: Eliminar el enlace simbólico `fsaverage`
-

Modelos LME masivos para la comparativa entre dos grupos

Como ya se ha dicho, este proceso estadístico va a ser aplicado a los sujetos provenientes del estudio de Bernal [3]. Aunque existen ciertas diferencias, se puede ver una introducción al modelo LME en la sección correspondiente del apartado de análisis estadístico (sección 4.3).

En este caso se parte de los archivos anteriormente generados:

- *50sMCI_vs_50cMCI-long_thickness_sm10.mgh*
- *rh.50sMCI_vs_50cMCI-long_thickness_sm10.mgh*

También se necesita la estructura *ADNI_Long_50sMCI_vs_50cMCI.mat* donde vienen recogidas entre otras cosas el número de visitas y la matriz de efectos fijos X.

A partir de estos datos y de las proyecciones esféricas *lh_sphere_cortex_label_fs* y *rh_sphere_cortex_label_fs* se construirán una serie de modelos. Estos modelos recibirán el nombre de *mass_SC_Bernal_X.mat*, donde X es cada subconjunto estudiado para el contraste. Para más información sobre cómo se obtienen estas salidas véase el pseudocódigo 9.

Cabe mencionar que la obtención de estos modelos con distintos subconjuntos de sujetos permite realizar un mapa de q-valores en lugar de p-valores, tal y como se explica en el anexo A.

Algoritmo 9 pseudocódigo de buildMassLmeModels

Entrada:

lh.50sMCI_vs_50cMCI_long_thickness_sm10.mgh
 rh.50sMCI_vs_50cMCI_long_thickness_sm10.mgh
 ADNI_Long_50sMCI_vs_50cMCI.mat
 lh_sphere_cortex_label_fs
 rh_sphere_cortex_label_fs

Salida:

X archivos *mass_SC_Bernal_X.mat* con la información del modelo.

- 1: Cargar la matriz de efectos fijos desde *ADNI_Long_50sMCI_vs_50pMCI.mat*
 - 2: Cargar las matrices con el grosor en cada vértice por cada sujeto (Y_lh, Y_rh)
 - 3: Cargar los datos de proyección esférica de fsaverage
 - 4: **para** i=1:nboot **hacer**
 - 5: Se divide en un 75 %-25 % aleatorio para el subgrupo de train.
 - 6: Se obtiene el modelo para el hemisferio izquierdo
 - 7: Se obtiene el modelo para el hemisferio derecho
 - 8: Se transforman las estructuras del modelo en Bhat_?H
 - 9: Se guarda el resultado en *mass_SC_Bernal_x.mat*
 - 10: **fin para**
-



Figura 4.24: Esquema de los archivos de entrada y salida en el flujo de trabajo de buildMassLmeModels, donde se generan x modelos distintos basados en distintos subgrupos de sujetos.

Generación de los q-valores

En este script se lleva a cabo una corrección de los p-valores. Si bien este concepto se explica en el anexo A, no está de más recordarlo aquí para garantizar una mejor comprensión.

Al realizar múltiples comparativas (en el caso de este estudio, 8 subconjuntos de la población de Bernal) si se ejecutan suficientes tests también aumentará el número de falsos positivos ($0,05 \times 100 = 5$, $0,05 \times 1000 = 50$). Esto se debe a que en cada test la probabilidad de tener un falso positivo (error de tipo 1) se mantiene constante.

FDR proviene de las siglas en inglés de *False Discovery Rate*. Cambia el paradigma estableciendo que el p-valor se refiera al conjunto de los tests. A este p-valor ajustado se le llama q-valor.

En el caso del análisis masivo, se están realizando tests para cada uno de los 160.000 vértices de la superficie. En este contexto se realizan los ya mencionados 8 subconjuntos de la muestra. Por lo tanto, para cada vértice se obtendrán 8 p-valores.

Existen varios acercamientos. El primero que se intentó fue a manos de Benjamini y Hochberg en 1995 [1]. Acorde a su publicación, si se desea controlar que en un estudio con N comparaciones el FDR no supere un porcentaje D:

1. Deben ordenarse los p-valores de los N tests de menor a mayor. Este resultado se puede entender mejor representando los valores en forma de histograma (véase figura A.4).
2. Se define K como la última posición para la que se cumple que $p_i \leq D \times \frac{i}{N}$, siendo i el índice dentro de la lista de p-valores.
3. Se consideran válidos todos los p-valores hasta la posición K.

La función utilizada para validar los p-valores en la demo de Bernal para LME masivo está basada en el criterio FDR de Thomas Nichols, profesor de la universidad de Warwick [13].

Una versión actualizada no sólo de este algoritmo sino de otros trabajos relacionados puede encontrarse en la página de la propia universidad de Warwick: <https://warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/fdr>.

Respecto al presente estudio, esta labor se realiza mediante dos funciones descritas en los pseudocódigos 10 y 11: `getFDRMaps()` y `saveThicknessPvalueMap()`.

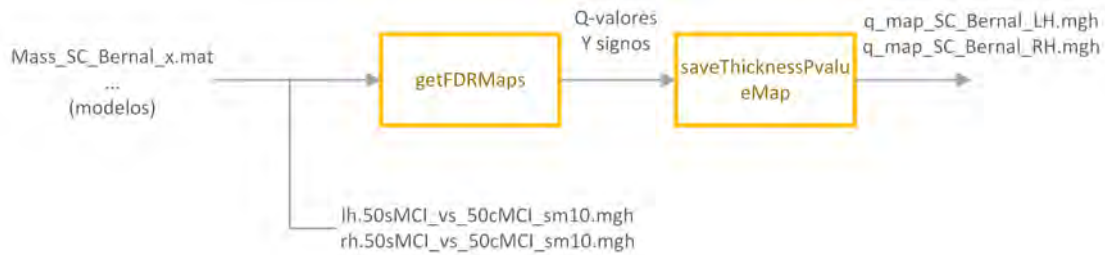


Figura 4.25: Esquema de la función `save_Thickness`, que genera el mapa de significancia con el p-valor corregido a partir de los modelos generados en la etapa anterior. Estos valores se calculan en la función `getFDRMaps`.

En definitiva, se obtienen todos los p-valores sacados de los distintos modelos generados en la etapa anterior. A partir de los mismos, la función `getFDRMaps()` obtiene los q-valores y éstos son almacenados en un archivo llamado `q_map_SC_Bernal_?H.mgh`. a través de la función `saveThicknessPvalueMap()`.

Estos mapas de q-valores pueden observarse en FreeView superponiéndolos como *overlay* en una superficie de `fsaverage` (véase figura 4.26).

Algoritmo 10 pseudocódigo de `saveThicknessPvalueMap`

Entrada:

lh.50sMCI_vs_50cMCI_long_thickness_sm10.mgh
rh.50sMCI_vs_50cMCI_long_thickness_sm10.mgh

Salida:

Un archivo `q_map_SC_Bernal_?H.mgh` por hemisferio con la información de los q-valores.

- 1: [q-valores, signos]= `getFDRMaps()`
 - 2: Se carga la información
 - 3: Cargar la información de la estructura mri de `lh.50sMCI_vs_50cMCI_long_thickness_sm10.mgh`
 - 4: Se almacena, con signo, el logaritmo decimal del q-valor para el hemisferio izquierdo.
 - 5: Se almacena, con signo, el logaritmo decimal del q-valor para el hemisferio derecho.
-

Algoritmo 11 pseudocódigo de `getFDRMaps`

Entrada:

`dir_models`: referencia al directorio donde se encuentran los x archivos `mass_SC_Bernal_x.mat`

Salida:

`qval_lh`: Conjunto de q-valores del hemisferio izquierdo. `qval_rh`: Conjunto de q-valores del hemisferio derecho. `qsign_lh`: signo de los q-valores del hemisferio izquierdo. `qsign_rh`: signo de los q-valores del hemisferio derecho.

- 1: `listMAT` = lista con los x archivos encontrados en `dir_models`.
 - 2: Inicialización de los vectores de p-valores a cero.
 - 3: **para** archivo de `listMAT` **hacer**
 - 4: Se asignan los p-valores y el signo provenientes de un F-test anterior.
 - 5: **fin para**
 - 6: Se aplican los criterios FDR, devolviendo los q-valores.
-

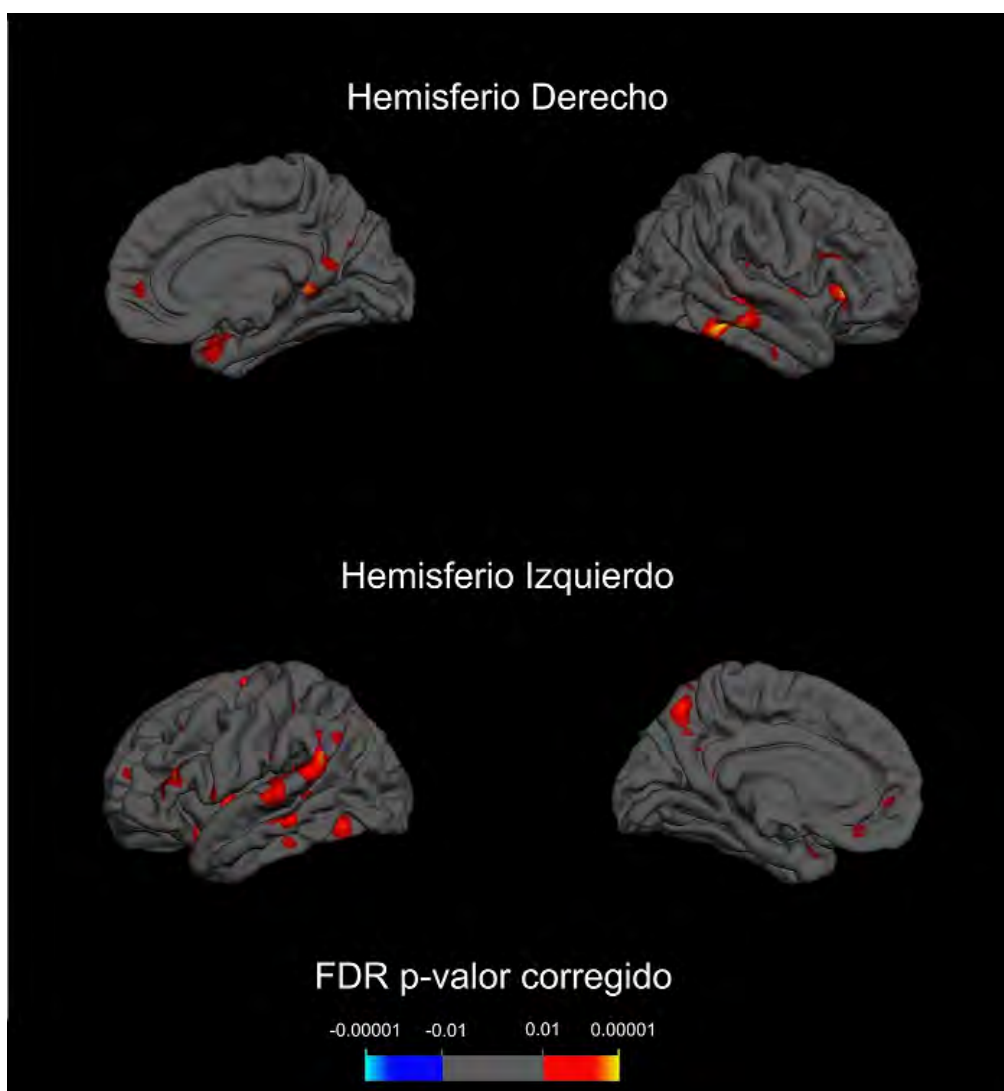


Figura 4.26: Mapa de significancia sMCI-pMCI.

Generación de los clusters

Una vez se tiene el mapa de q-valores es posible formar clusters que engloben las zonas más discriminantes entre los dos grupos. Para ello se ha utilizado la función *mri_surfcluster* de FreeSurfer. Esto simplifica el código considerablemente, consistiendo éste únicamente en dos llamadas a esta función (una por hemisferio). Dentro de los parámetros de la referida función cabe destacar:

- `-minarea`: establece el área mínima que debe tener un cluster para ser considerado válido. Aunque Landin en su estudio [18] defiende que el área mínima debe ser de 50 mm^2 , empíricamente se observa que un área de 100 mm^2 consigue clusters consistentes (véase figura 4.29).

- `-fdr`: posibilita un estudio `fdr`. En este caso, como ya se ha realizado, no debe aplicarse.
- `-thmin`: indica el límite mínimo del `q`-valor. Este valor está expresado como $-\log_{10}(p)$, por lo que para un nivel de significancia del 0.01 se deberá establecer un mínimo de 2 ($-\log_{10}(0,01) = 2$).

Otros *flag* simplemente indican los archivos de entrada y salida de la función. De los utilizados en esta implementación se pueden diferenciar los siguientes archivos:

1. **?h_summary.txt**: consiste en un resumen de la operación. Se trata de un simple archivo de texto donde poder consultar rápidamente el número y tamaño de los clusters encontrados. Se indica con el *flag* `-sum`.
2. **q_map_SC_Bernal.?H.mgh**: es un archivo de entrada con el mapa de los `q`-valores. Se indica con el *flag* `-in`.
3. **?h_label-XXXX.label**: almacena una máscara del cluster. Se generará un archivo por cada cluster encontrado. XXXX se corresponde al índice de cluster. Se indica con el *flag* `-olab`.
4. **?h_Total_clusters.mgh**: se trata de un archivo que agrupa la información de todos los clusters. Se indica con el *flag* `-cwsig`.



Figura 4.27: Esquema de los archivos de entrada y salida en la generación de clusters.

Es importante señalar que la función ofrece una gran versatilidad. Por ejemplo, se puede limitar la búsqueda de clusters a una zona determinada elaborando una máscara previamente.

Por último cabe destacar que, si bien el archivo `?h_Total_clusters.mgh` es el que se va a utilizar en la siguiente etapa, resulta conveniente generar los archivos `.label`. El motivo principal es que su sencilla visualización en `FreeView` permitiría revisar un cluster individualmente si se descubre que resulta importante en el análisis estadístico.

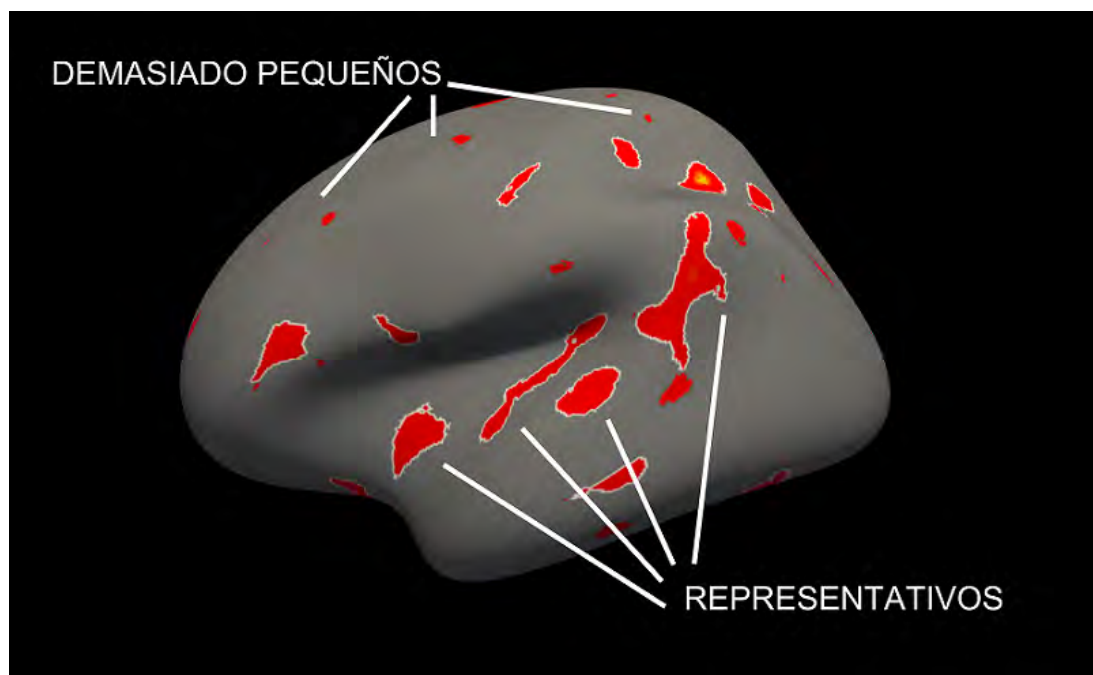


Figura 4.28: Esquema de los clusters obtenidos al realizar análisis masivo. Como se puede apreciar, se han desechado los clusters que no tuviesen un área mínima de 100 mm^2 . La representación se ha realizado en superficie inflada para mayor claridad.

Unión de tablas

Una vez se dispone de los clusters deben tratarse como una ROI, calculando el grosor medio del córtex en cada uno de ellos. Esto debe hacerse tanto para el hemisferio izquierdo como para el derecho.

Para ello se recurre a dos archivos:

- El archivo `?h_Total_clusters.mgh`, que dispone de una máscara para todos los clusters. En concreto, este archivo almacena una matriz de $4 \times nVtx$, donde $nVtx$ es el número de vértices. En este caso, al trabajar con `fsaverage`, $nVtx$ es 160000 aproximadamente.

La cuarta fila de esta matriz contiene por cada vértice de la malla un valor entero que indica a qué cluster pertenece (1,2,3,...) o si no pertenece a ningún cluster (0). Esto permite elaborar las máscaras abriendo un único archivo, en lugar de recorrer cada etiqueta generada en el paso anterior.

- El archivo `?h.name_ct_sm10_fs.mgh` generado en la primera etapa. Este archivo contiene una matriz de $N \times nVtx$, donde N es el número de scans de la población estudiada y $nVtx$ vuelve a ser el número de vértices de la malla. Esta matriz almacena el valor del grosor cortical para cada uno de los vértices y visitas.

Una vez se tiene esta información es fácil aplicar una máscara a la matriz proveniente de `?h.name_ct_sm10_fs.mgh`. Además, es posible que a la hora de generar la

tabla con información de las ROIs y los tests neuropsicológicos se hayan despreciado algunos sujetos y no concuerde el número de visitas. Es por ello que a la hora de unir los resultados es necesario combinar las tablas usando los elementos comunes. En el pseudocódigo 12 se explican los detalles.



Figura 4.29: Esquema del flujo de trabajo de join_Tables.

Algoritmo 12 pseudocódigo de joinTables

Entrada:

?h.Tot_clusters.mgh: archivos donde está almacenada la información de los clusters. En concreto, a qué cluster pertenece cada vértice, en una suerte de máscara.

?h.name_ct_sm10_fs.mgh.: archivos con la información del grosor cortical de la población estudiada para

orig_clinical: archivo con los datos clínicos a partir de los cuales se generaron los archivos ?h.name_ct_sm10_fs.mgh.

RoiTest: tabla con la información de Roi y tests neuropsicológicos.

Salida:

RoiTestMass: tabla con la información de Roi, tests neuropsicológicos y del análisis masivo.

- 1: Pop_thickness = lectura del archivo ?h.name.ct.sm10_fs.mgh
 - 2: index = Se leen los índices para todos los vértices del archivo ?h.Tot_clusters.mgh
 - 3: Se elimina el índice nulo (0) que indica que el vértice no pertenece a ningún cluster.
 - 4: ncl = número de clusters (índices diferentes).
 - 5: **para** i=1:ncl **hacer**
 - 6: nombres(i)= 'cl_i'
 - 7: selection_index = bool(index == i)
 - 8: selection = Pop_thickness (selection_index)
 - 9: statistics = media de los valores para cada fila de selection.
 - 10: **fin para**
 - 11: Se unen las columnas generadas para cada cluster
 - 12: Se repite el proceso para el otro hemisferio
 - 13: Tmass =[Tmassrh, Tmasslh]
 - 14: Tmass.id = id sacado de la tabla orig_clinical
 - 15: Tresult = join(RoiTest, Tmass)
 - 16: Se guarda Tresults como RoiTestMass
-

Una vez se dispone de la tabla RoiTestMass se puede proceder a realizar todos los experimentos del estudio. En la siguiente sección se describen las herramientas matemáticas y estadísticas utilizadas como base, y en el último apartado se describirán los distintos experimentos llevados a cabo.

4.3. Análisis estadístico

4.3.1. Algoritmo de selección mRMR

Explicación

A la hora de abordar un problema de textitmachine learning, el primer reto que aparece es reducir el número de características. Esto es necesario para conseguir que el algoritmo sea capaz de generalizar, y no adecuarse únicamente al conjunto de muestras de entrenamiento. A este problema se le conoce en la literatura como *overfitting* (véase anexo A).

El término *mejores* lo decide el algoritmo de selección, premiando o penalizando ciertas propiedades. Con esta evaluación se puede obtener un cierto orden de las características, de mejor a peor.

En este apartado se va a explicar brevemente el algoritmo de mínima redundancia y máxima relevancia (mRMR) utilizado en este trabajo. Este algoritmo fue propuesto en 2005 por Chris Ding y Hanchuan Peng [8] y con él se pretende optimizar la selección atendiendo a dos criterios:

- **Mínima Redundancia:** si dos características son redundantes, aunque por separado aporten información importante para diferenciar dos grupos, al juntarlas no es esperable mejorar los resultados, o al menos no sustancialmente. El objetivo de esta condición es escoger las características más representativas. Esta condición puede ser expresada como:

$$\min W_C, W_C = \frac{1}{|S|^2} \sum_{i,j} |c(x_i, x_j)|$$

donde S es el subconjunto deseado y $c(x_i, x_j)$ es el coeficiente de correlación de Pearson entre las características x_i y x_j .

- **Máxima Relevancia:** si no hay diferencias significativas entre los valores de los dos grupos para una característica, entonces ésta no servirá para diferenciarlos. Esto resultará en un valor bajo en esta parte de la valoración. Esta condición puede ser expresada como:

$$\max V_F, V_F = \frac{1}{|S|} \sum_{i \in S} F(x_i, h)$$

donde S es el subconjunto deseado y F la prueba F de Fisher de la característica x_i para el conjunto de posibles clases h. Para el caso de dos clases, la prueba F de Fisher puede reducirse a una prueba T de Student ($F = t^2$).

Para optimizar los dos criterios, éstos deben ser combinados en una única fórmula. Se puede hacer recurriendo a distintos métodos, pero los más sencillos son los siguientes:

- Diferencia (d_mRMR): la ecuación que se utiliza para optimizar la selección es la resta de ambos criterios:

$$\max(V_F - W_C)$$

- Coeficiente (q_mRMR): la ecuación que se utiliza para optimizar la selección es el cociente entre ambos criterios:

$$\max(V_F/W_C)$$

Estos dos acercamientos al problema dan lugar a las dos implementaciones introducidas en la herramienta de MATLAB ofrecida por Chris Ding y Hanchuan Peng utilizada en este trabajo (<https://es.mathworks.com/MATLABcentral/fileexchange/14608-mrmr-feature-selection--using-mutual-information-computation-/>).

Implementación

En este trabajo, la selección de características se produce a partir de los datos de las tablas con tests neuropsicológicos y valores medios de estructuras corticales, subcorticales y clusters.

Sea cual sea la implementación del algoritmo de clasificación, el algoritmo de selección de características debe ejecutarse antes de la clasificación. En este estudio en concreto se ha utilizado un algoritmo de regresión lineal de efectos mixtos (véase la sección 4.3.2) y están relacionados por una estructura de K-folds anidados (véase la sección 4.3.3).

En el caso de este estudio, se ha envuelto el algoritmo original mRMR con una selección por frecuencia. Es decir, se ejecuta para un número determinado de iteraciones Nit el algoritmo con subconjuntos de la población. Una vez hecho esto, se obtienen Nit ordenaciones distintas almacenadas en una matriz de $Nit \times Nc$, donde Nc es el número de características evaluadas.

A partir de estas ordenaciones se pueden encontrar, para cada dimensión, las combinaciones más repetidas en los primeros puestos de la clasificación. Éstas se consideran como las más prometedoras al demostrar ser buenas para varios subconjuntos de la población.

De esta manera se puede evitar que la selección se adapte únicamente a la muestra. Este acercamiento se volverá a hacer patente en el algoritmo de K Folds anidados (véase la sección 4.3.3).

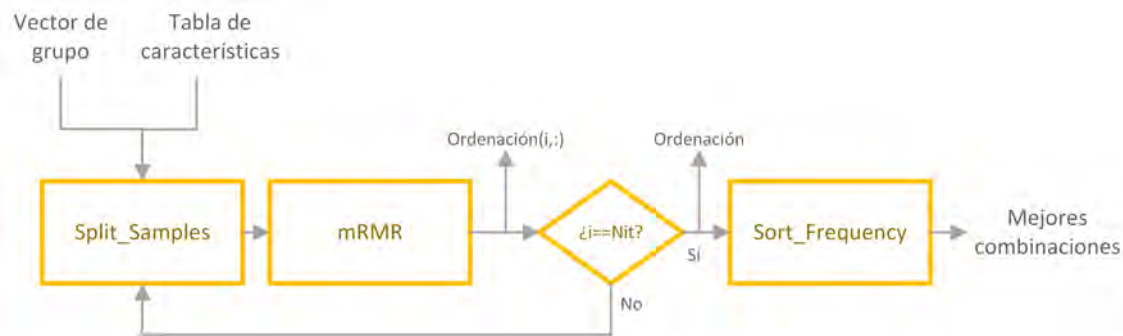


Figura 4.30: Esquema de mRMR donde las características se escogen por la frecuencia de aparición en los primeros puestos.

4.3.2. Análisis longitudinal y modelo lineal de efectos mixtos

Introducción a la regresión lineal

En un estudio transversal se puede intentar predecir el valor y resultante para un elemento con una serie de características X a partir de la llamada regresión lineal. Ésta se puede definir como

$$y = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_n\beta_n + err$$

donde cada elemento x_i es una característica y cada elemento β_i es el peso asociado a la misma. El elemento β_0 corresponde a un valor constante. Puede entenderse como un *offset* que desplaza la función. El error no puede ser calculado, pero representa la diferencia existente entre la salida predicha y el resultado real. Puesto en forma matricial queda:

$$Y = X\beta + err$$

donde a la matriz X se le añade una columna unitaria correspondiente al término $(1 \times \beta_0)$. Con el grupo de entrenamiento el algoritmo utilizado debe determinar los pesos β que minimicen el error en la predicción. Una vez hecho esto se guardan esos valores y se utilizan para predecir el resultado para nuevos elementos conociendo únicamente el vector X de características. Este procedimiento correspondería al proceso de test.

Introducción a los estudios longitudinales

Un análisis longitudinal hace referencia a un estudio sobre la evolución de uno o varios parámetros a lo largo del tiempo (véase figura 4.31). Como ejemplo relacionado con este estudio, no sólo se tiene en cuenta el espesor del córtex, sino que también se mide el ritmo de degradación.

Por lo general, este tipo de estudios permiten obtener mejores resultados que los análisis transversales, especialmente a medida que se aumenta el número de datos a lo largo del tiempo. Sin embargo, tienen como contrapunto el uso de nuevas herramientas y bases matemáticas todavía en desarrollo.

En estudios médicos o anatómicos se espera que medidas cercanas en el tiempo presenten una correlación mayor que aquellas más distantes. Esto es debido a que existe un proceso biológico subyacente a los datos que evoluciona de manera continua y muchas veces con grandes constantes de tiempo.

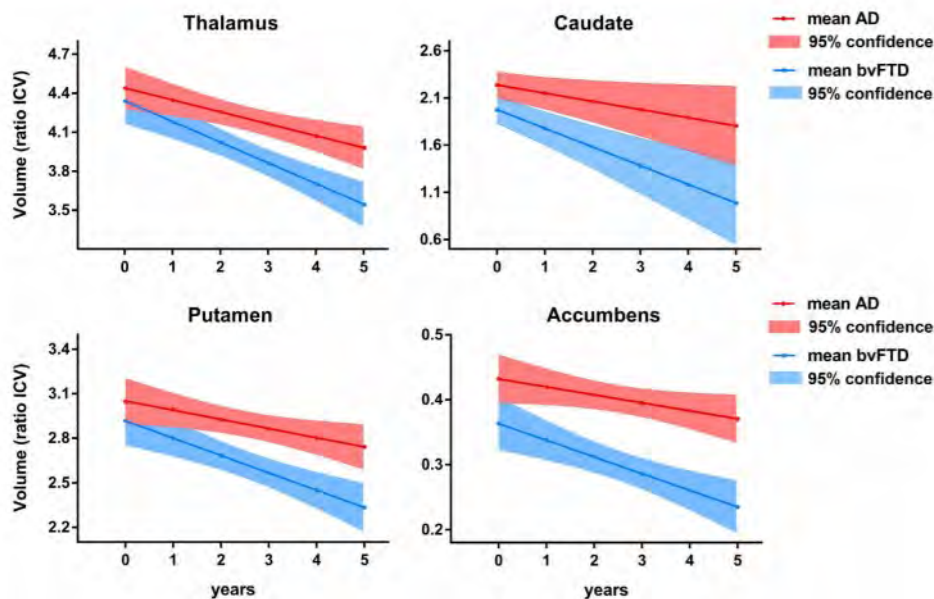


Figura 4.31: Representación de los datos de un estudio longitudinal[18].

Modelo lineal de efectos mixtos

Un ejemplo de estos modelos matemáticos lo encontramos en el modelo longitudinal de efectos mixtos (LME). La principal fuente de información en este trabajo sobre este modelo proviene de un artículo publicado por Bernal en 2013 [2].

Frente a modelos transversales (cross) LME permite corregir errores provenientes del irregular espaciado entre distintas visitas. Además ofrece una respuesta a la alta variabilidad entre distintos sujetos. Esta variabilidad no solo es alta sino que no se mantiene constante durante todo el proceso.

Tal y como expone Bernal en el artículo ya mencionado, existen tres fuentes principales de variabilidad:

1. Variación entre sujetos: distintos sujetos poseen anatomía, condiciones y por lo tanto una evolución diferente.

2. Variación dentro del sujeto: el mismo sujeto en dos momentos diferentes de tiempo poseerá distintas características anatómicas.
3. Errores de medida: incluso en medidas cercanas en el tiempo al mismo sujeto, donde se pueden despreciar las dos variaciones anteriores, existe un error inherente al proceso de medida.

El modelo LME se puede representar como:

$$Y_i = X_i\beta + Z_ib_i + e_i$$

donde Y_i es un vector de dimensiones $(n_i \times 1)$ siendo n_i el número de visitas del sujeto. Este vector Y_i representa el valor de una medida (p.ej.: el espesor medio del córtex) a lo largo de las n visitas del sujeto i . X_i es la llamada matriz de efectos fijos (género, educación, grupo clínico, etc) y tiene una dimensión $(n_i \times p)$ donde p es el número de efectos que se tienen en cuenta. β es un vector $(p \times 1)$ con los coeficientes de regresión, y son compartidos por todos los sujetos. Z_i es la matriz de efectos aleatorios, de dimensiones $(n_i \times q)$, donde $q < p$. Z_i representa el subconjunto de variables de X a las que se pretende añadir una naturaleza aleatoria además de un comportamiento genérico definido por β . Por último, el vector b_i contiene los coeficientes para estos efectos aleatorios.

Para poder aplicar el modelo LME se asume:

- Que b_i sigue una distribución normal centrada en cero. Esto quiere decir que los efectos aleatorios se producen siguiendo una distribución normal. Debe recordarse que el vector b_i es propio de cada sujeto, a diferencia de los coeficientes comunes β . Esto puede expresarse como:

$$b_i \sim N(0, D);$$

- Que el error e_i se distribuye siguiendo una normal centrada en el cero. Esto puede expresarse como:

$$e_i \sim N(0, \sigma^2 I_{n_i});$$

Donde I_{n_i} es la matriz identidad de tamaño n_i (número de visitas del sujeto i)

Si se cumple esto, entonces el valor esperado de Y_i puede darse de dos maneras:

1. El valor condicionado o específico para un sujeto: Se trata del valor esperado teniendo el valor del vector b_i . En este caso, la esperanza corresponde a:

$$E(Y_i|b_i) = X_i\beta + Z_ib_i$$

2. El valor esperado marginal, o valor medio de la población: se calcula teniendo en cuenta el comportamiento general en el grupo. Esto quiere decir que sólo se utiliza la matriz de efectos fijos una vez el vector β (común a todos los sujetos) ya ha sido calculado. Puede expresarse de la siguiente forma:

$$E(Y_i) = X_i\beta$$

A la hora de aplicar el algoritmo, el objetivo es estimar los parámetros β a partir del grupo de entrenamiento. También se deben estimar los parámetros σ y D relacionados con la distribución normal de b_i y e_i .

$$Y_i \sim N(X_i\beta, \Sigma_i)$$

donde $\Sigma_i = (Z_i D Z_i^T + \sigma^2 I_{m_i})$ y el resto de variables han sido ya definidas anteriormente. En este punto, caben dos posibilidades:

- En caso de disponer de estimadores \hat{D} y $\hat{\sigma}$ existe una solución cerrada para el estimador de β :

$$\hat{\beta} = \left(\sum_{i=1}^m X_i^T \hat{\Sigma}_i^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i^T \hat{\Sigma}_i^{-1} y_i$$

donde $\hat{\Sigma}_i = (Z_i \hat{D} Z_i^T + \hat{\sigma}^2 I_{m_i})$ y y_i es la realización del vector aleatorio Y_i .

- En caso contrario se puede recurrir a una fórmula abierta que debe resolverse mediante métodos numéricos:

$$l_{ReML} = \frac{1}{2} \left(\sum_{i=1}^m \log |\Sigma_i^{-1}| - \sum_{i=1}^m (y_i - X_i \hat{\beta})^T \Sigma_i^{-1} (y_i - X_i \hat{\beta}) - \log \left| \sum_{i=1}^m X_i^T \Sigma_i^{-1} X_i \right| \right)$$

Tras haber entrenado el modelo, se usa este estimador de los pesos de efectos fijos $\hat{\beta}$ para predecir el grupo del sujeto.

Implementación en este trabajo

Aunque la implementación en este estudio del algoritmo de clasificación LME está englobada dentro del flujo de trabajo de los K-folds anidados (véase sección 4.3.3), la llamada principal para hacer uso del modelo LME se ha denominado *Test_GroupFeat*.

Esta función toma como entrada una división en un grupo de entrenamiento y test a través de máscaras (`mask_train`, `mask_test`, `mask_bl_test`) y vectores con el número de imágenes (`ni_train`, `ni_test`). Por otro lado se introducen como parámetros de entrada la matriz de efectos fijos X , la matriz de características y los conjuntos

de características con los que se pretende evaluar la predicción (provenientes normalmente de mRMR).

La función realiza esta valoración (Train-Test) una única vez. Las distintas iteraciones para obtener valores estadísticos deben realizarse fuera de la función. La división entre el grupo de Train y Test también queda, como se ha mencionado, a cargo de un script externo.

Lo primero que se hace es reducir la matriz de características a las imprescindibles. Estas son las que aparecen al menos una vez en alguno de los conjuntos de características que se pretenden probar. Tras esto, para cada característica se calcula el error residual a partir del conjunto de entrenamiento.

Una vez se tiene el error residual de todas las características que se van a utilizar, se itera por los distintos grupos y se evalúa la predicción del modelo dados los errores residuales de las características del conjunto en concreto.

De cada grupo se extrae un conjunto de 6 variables estadísticas (sensibilidad, especificidad, PPV, NPV, precisión, área bajo la curva) que se podrán utilizar más adelante para evaluar la eficacia del algoritmo. Esto se realiza tomando todas las imágenes de test y tomando únicamente la primera visita.

Es precisamente esta tabla de estadísticas del error el parámetro de vuelta de la función. Se trata de dos matrices de $ngroups \times 6$ donde cada fila es un conjunto distinto de características y cada columna es un valor estadístico del error. Una matriz corresponde a la evaluación utilizando todas las imágenes del test, y la otra utilizando el *baseline* únicamente. Los detalles quedan mejor explicados en el pseudocódigo 13 y la figura 4.32.

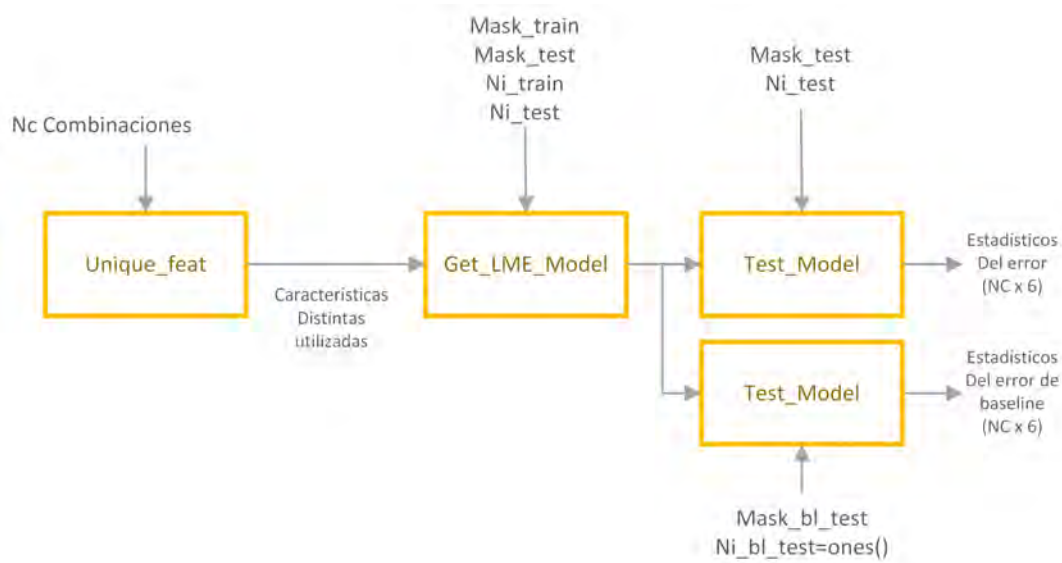


Figura 4.32: Esquema de la implementación LME, donde se obtiene tanto el estadístico en baseline como el estadístico con todas las imágenes del subconjunto de test.

Algoritmo 13 pseudocódigo de LME

Entrada:

- Grupos de características que se quieren probar.
- Tabla con los datos de todos los sujetos.
- Matriz de efectos fijos X .
- Máscara de entrenamiento-test-baseline_test.
- Número de imágenes de entrenamiento-test.

Salida:

- Datos del error de clasificación con test.
 - Datos del error de clasificación con baseline_test.
 - 1: Utilizar subconjunto de la tabla de datos con las características que se quieren probar.
 - 2: Calcular los residuos (LME) para cada característica que vaya a aparecer.
 - 3: **para** $i=1$ hasta número de grupos que se quierñ probar **hacer**
 - 4: Entrenar el modelo con el grupo de entrenamiento.
 - 5: Probar el modelo con el grupo de test.
 - 6: Probar el modelo con el grupo de baseline_test.
 - 7: **fin para**
 - 8: Devolver los resultados del error
-

4.3.3. Modelo de K-Folds anidados

Modelo original

Como ya se ha mencionado en el capítulo *Estado de la técnica*, la idea del uso de K-Folds anidados proviene del estudio de Korolev16 [17]. El modelo original funciona de la siguiente manera:

1. Primero se realiza una división de grupos Entrenamiento-Test del 90%-10%. En el caso de Sorensen esta división se hace 10 veces y se trata de 10 conjuntos disjuntos. Sin embargo, una selección aleatoria tiene el mismo efecto. Esto se debe a que los sujetos que en una iteración son entrenamiento, en otra iteración formarán parte del grupo de test.
2. Una vez hecho esto, se pasa a ejecutar en cada iteración externa (por cada *external fold*) una serie de divisiones internas (*internal fold*): de ahí el adjetivo *anidados*.

De nuevo se realizan, por tanto, 10 divisiones distintas, pero esta vez únicamente del conjunto de entrenamiento. Esta división se corresponderá a un grupo de selección de características y un segundo grupo de validación.

Para la selección de características se usa un algoritmo de selección. En el caso de este estudio se trata de mRMR, aplicado con una selección interna tal y como se ha comentado en su correspondiente sección. Esto devuelve una

selección de las características más prometedoras. En el caso de este estudio se ha establecido que se devuelvan las 10 mejores combinaciones según mRMR para cada dimensión.

3. De cada par de subconjuntos se prueban las 10 mejores combinaciones mRMR en el subconjunto de validación (10 % del 90 %) y con los resultados de área bajo la curva del baseline obtenidos se escogen las 3 mejores para cada dimensión.

De esta manera, por cada *fold* interno se obtienen los siguientes resultados:

- *10 combinaciones elegidas por mRMR \times D dimensiones* como candidatas. Debe recordarse que esto estará multiplicado por los 10 *fold*s internos por cada *fold*s externos. Todas estas selecciones de características pasarán por el subconjunto de validación.
 - *3 mejores combinaciones validadas \times D dimensiones*. De nuevo debe recordarse que esto se repetirá para todos los *fold*s internos por cada *fold* externo.
4. Se recuerda que esta última selección se obtiene de un subconjunto del 90 % del total. Por cada *fold* externo, esta selección se prueba en el 10 % restante para obtener los resultados definitivos (test y baseline_test).
 5. Estas divisiones evitan una contaminación estadística: el grupo de selección no se cruza con el de validación. Por otro lado, el grupo entero de entrenamiento (90 %) no se cruza con el grupo de test (10 %).

El resultado de este análisis, así como la selección, se pueden guardar en archivos .mat para poder procesarlo más adelante y ofrecer resultados.

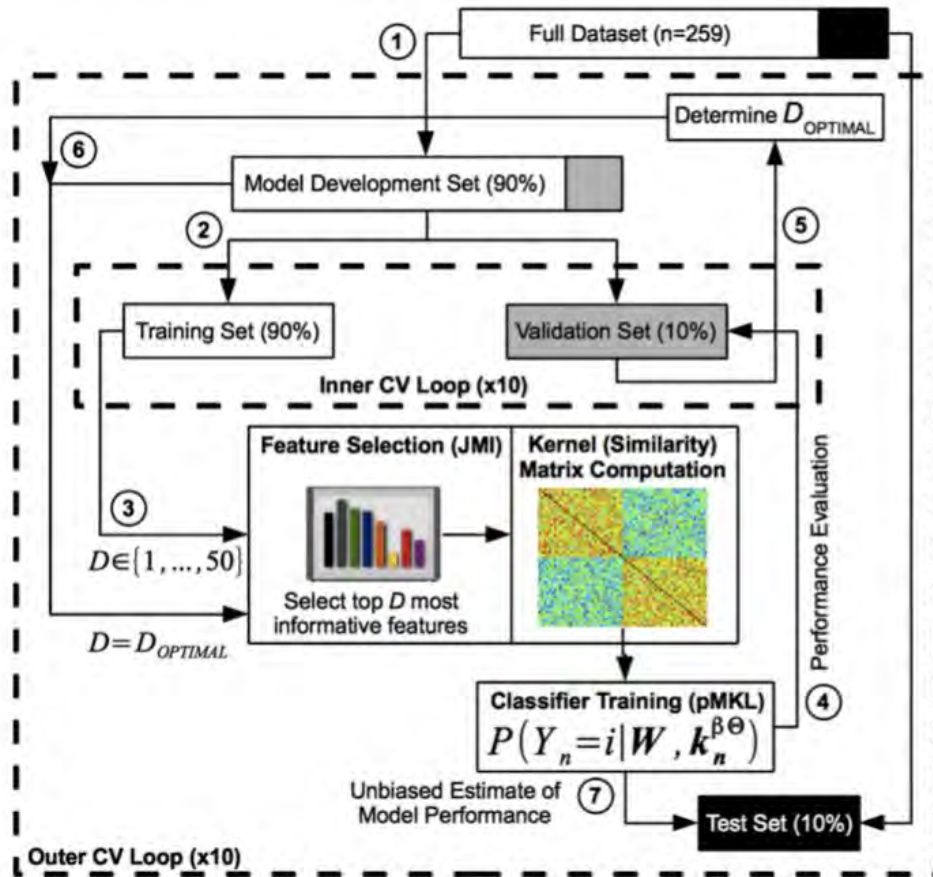


Figura 4.33: Esquema de la validación mediante k-folds llevada a cabo en el estudio de Korolev[17]. Imagen del artículo original.

Variante implementada en este proyecto

El modelo K-Fold utilizado en este estudio es una variante basada en elecciones por la frecuencia de selección de características. Este modelo K-Fold basado en una Selección por Frecuencia (KFSF, de ahora en adelante) presenta ciertas diferencias con el modelo original:

1. Aunque no se trate de una diferencia como tal, el algoritmo mRMR utilizado para la selección inicial de características está igualmente basado en la frecuencia de selección, tal y como se describe en el apartado correspondiente.
2. La siguiente diferencia fundamental es que para poder escoger características por su frecuencia de aparición se debe repetir la división externa de los K-Folds más veces.

Por lo tanto, aunque se mantiene una división interna en 10 folds *train-validation*, se realiza un número mayor de divisiones *train-test*. Esto permite asegurar que

en dimensiones altas (8,9,10,etc.) se repite alguna combinación lo suficiente como para considerarlo representativo.

3. Surgida de este paradigma, la siguiente diferencia fundamental es el uso de una selección aleatoria (90 %-10 %) en lugar de conjuntos disjuntos (folds) puesto que, como ya se ha explicado, tienen el mismo efecto.

4. La búsqueda de la dimensión óptima, así como de los conjuntos más prometedores para cada dimensión, se realiza de una manera diferente. De todos los conjuntos de cada dimensión no se consideran más representativos aquellos con mejor puntuación de error, sino aquellos que más se han repetido al ser seleccionados y validados.

El motivo de este cambio es que es más probable que estos conjuntos de características generalicen mejor su predicción para cualquier subconjunto de Train-test. Esto significa robustez y que se espera que al procesar varias veces estos resultados con distintos conjuntos el valor del error no varíe.

El guiarse por el error desde el principio puede llevar a seleccionar conjuntos de características que son especialmente buenos para una división determinada pero pésimos para otras.

De esta manera, para cada dimensión se seleccionarán los cinco conjuntos más repetidos en los K-Folds externos y se probarán en un número igual de iteraciones (p.ej.: 500). Una vez hecho esto se compara el valor estadístico medio de cada una para determinar los mejores conjuntos de cada dimensión y la dimensión óptima. De todo esto se hablará con más de talle en la sección 5.2.

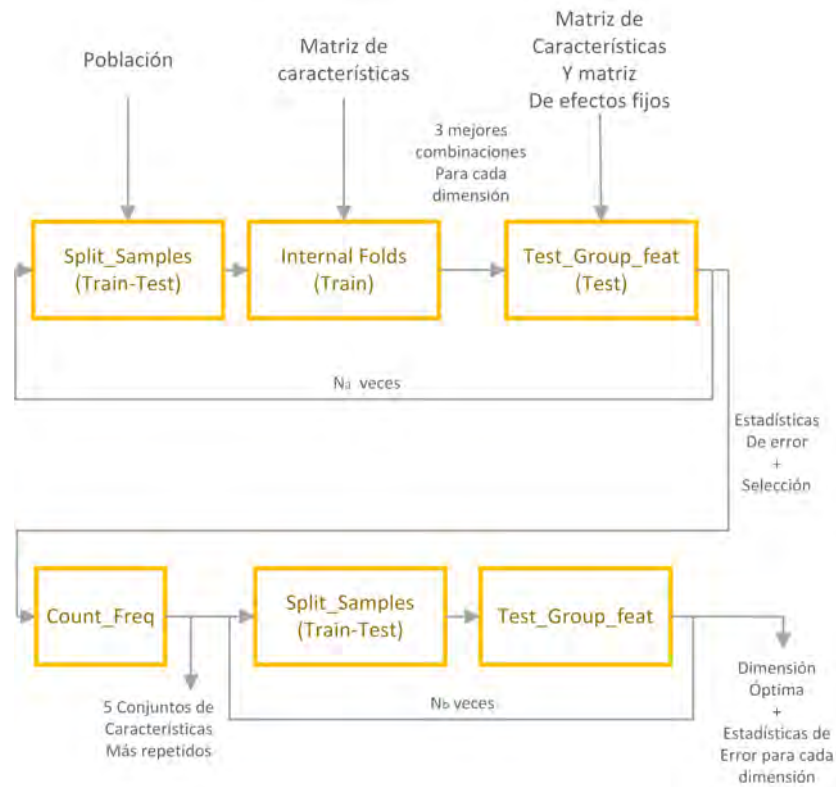


Figura 4.34: Esquema de la variante de K-Folds con selección por frecuencia (KFSF), que es la base de todos los experimentos llevados a cabo en este estudio.

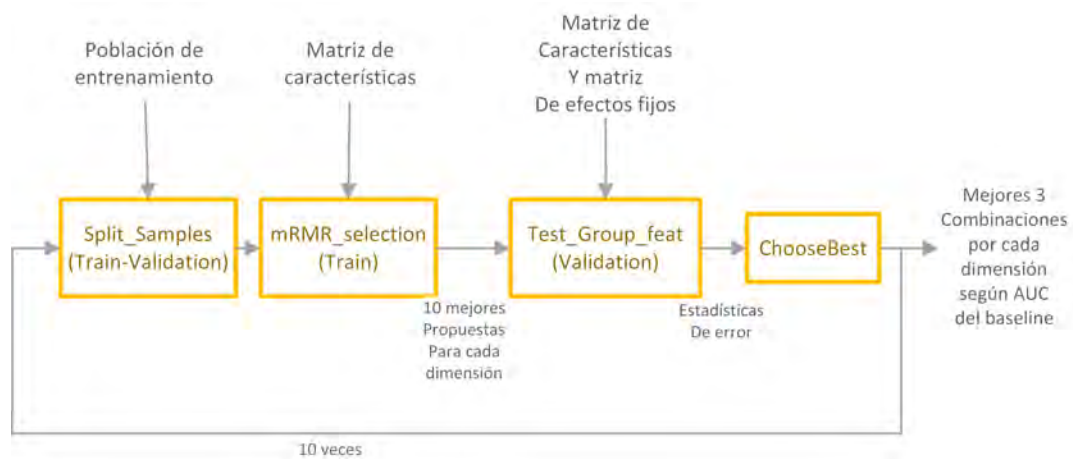


Figura 4.35: Esquema de los folds internos, ejecutados 10 veces por cada fold externo. En el interior de cada fold interno se seleccionan los conjuntos de características más prometedores mediante mRMR y se evalúan con el subconjunto de validación mediante LME.

Respecto a la implementación en MATLAB, se ha establecido un tercer bucle que

permite un procesamiento en paralelo. Tal y como se expone en la figura 4.36 antes de ejecutar el algoritmo de K-Folds se realiza una división inicial en la que se estipula cada cuántas iteraciones se desea guardar el resultado en un archivo .mat numerado.

Con esta información se procede a ejecutar en paralelo cada Fold externo. Esta decisión se ha tomado teniendo en cuenta distintas propiedades de la división del trabajo entre varios núcleos:

- Cuanto más global sea la división en núcleos, mayor será el ratio de tiempo de procesamiento frente al tiempo de comunicación entre los núcleos. Este último permanece más o menos constante, sea la tarea que sea. Sin embargo, si se aprovecha esta división del trabajo para tareas extensas, el tiempo de comunicación es despreciable frente al tiempo de procesamiento.

Este es el motivo por el que no se realiza la paralelización con los bucles internos, sino con los externos.

- Para guardar los datos es necesario garantizar una cierta secuenciación en el bucle. Es por ello que se ha combinado un bucle externo con una paralelización interna.

Este es el motivo por el que no se realiza una paralelización directamente con todos los bucles externos que se desean ejecutar, sino que se realiza una división previa.

Puesto en un ejemplo práctico: imagínese el caso en el que se desea realizar un total de 750 folds externos, y se desea guardar el resultado cada 100 iteraciones. Entonces se harán 8 llamadas a la función `KFolds()` con un parámetro de entrada `external_kfolds = 100` excepto en la última iteración, donde `external_kfolds = 50`.

Dentro de esta función `KFolds()` se ejecutan en paralelo los folds externos para esas 100 ó 50 iteraciones. Al acabar estas iteraciones se guarda el resultado en un archivo y se sale de la función, volviendo al bucle inicial de 8 iteraciones. Esto puede entenderse mejor acudiendo al pseudocódigo 14 y al esquema de la figura 4.36.

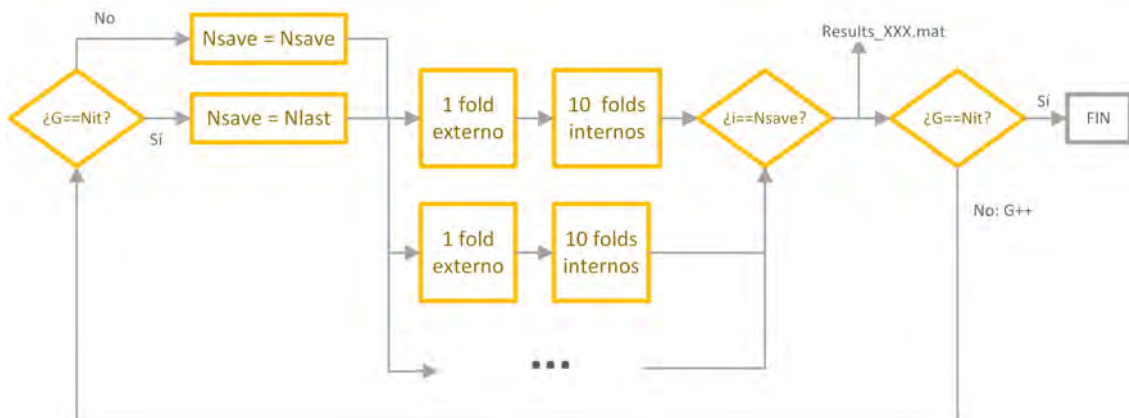


Figura 4.36: Esquema de la implementación del cálculo paralelo en KFSF.

Algoritmo 14 pseudocódigo del paralelismo del procesamiento de K-Folds

Entrada:

Nextern = Número de external_KFolds.
Nsave = Cada qué número de iteraciones se desean guardar los resultados en un archivo.

- 1: Nit = floor(Nnextern/Nsave)
- 2: Nlast = módulo(Nnextern,Nit)
- 3: **si** Nlast == 0 **entonces**
- 4: Nlast = Nit
- 5: **si no**
- 6: Nit = Nit + 1
- 7: **fin si**
- 8: **para** g=1 hasta Nit **hacer**
- 9: **si** g==Nit **entonces**
- 10: Nsave = Nlast
- 11: **fin si**
- 12: **para** paralelo i=1 hasta Nsave **hacer**
- 13: Dividir en Train-Test
- 14: Realizar los 10 Folds internos
- 15: Probar la selección de características en el test.
- 16: **fin para**
- 17: Guardar los resultados de las Nsave iteraciones en un archivo
- 18: **fin para**

Capítulo 5

Resultados

5.1. Introducción del capítulo

En esta sección se procede a describir los experimentos realizados, basados en las técnicas explicadas en el capítulo anterior, así como los resultados de los mismos.

Respecto a la discusión sobre estos resultados, los detalles se plantean en esta sección, mientras que en el capítulo 6 se plantea la discusión general junto a las conclusiones más relevantes.

Un aspecto importante de las comparativas entre estudios es que los resultados comparados serán los calculados con la información de la primera visita, para poder establecer una correlación entre la predicción de estudios transversales y el presente estudio.

En las distintas secciones de este capítulo se pueden observar los resultados de los mismos experimentos modificando la matriz de efectos fijos.

5.2. Experimentos

En esta sección se describen todos los experimentos que han dado lugar a los resultados de este estudio, utilizando los métodos anteriormente explicados. Gran parte de los experimentos se basan en una ejecución previa del K-Folds, como es el caso de la búsqueda de la dimensión óptima o el cálculo de las características más repetidas.

5.2.1. Obtención de los datos de la población

Este experimento es en verdad sencillo, pero no por ello menos importante. Se trata de una exploración de los datos recogidos en la tabla *Tresults*, atendiendo especialmente a los datos clínicos y demográficos, así como al número de visitas. Este experimento se ha recogido en un único script, `getPopulationData()`.

El experimento tiene dos objetivos principales. Por un lado, permite conocer una serie de características de la población imprimiéndolas por pantalla. Esto se ha utilizado para calcular todos los datos mostrados en la sección 3, *Materiales y Herramientas*.

El otro objetivo de este experimento es cerciorarse de que los grupos están equilibrados. El equilibrio más importante es que haya una cantidad similar de sMCI y pMCI. No obstante, también es importante que ambos grupos tengan una edad parecida y que haya igualmente un equilibrio de género en cada población.

Si hubiera un desequilibrio demasiado grande de uno de los dos grupos clínicos (sMCI o pMCI) las medidas de los resultados podrían quedar falseadas.

Tómese como ejemplo que si una de las poblaciones es muy pequeña respecto a la otra, el algoritmo puede dar una precisión elevada prediciendo prácticamente por defecto el diagnóstico mayoritario.

Estos fallos se reflejarían en la sensibilidad y especificidad, pero ante grupos equilibrados la precisión representa una buena medida del error junto al área bajo la curva.

5.2.2. Cálculo de la dimensión óptima

El objetivo de este experimento es calcular la dimensión óptima dada la matriz de efectos fijos estudiada. Para ello, de las combinaciones más repetidas por el algoritmo K-Folds se escogen las N_{best} más repetidas para cada dimensión (en concreto se ha escogido $N_{best}=5$).

Una vez hecho esto se prueban todas ellas un número determinado de veces con subconjuntos distintos y se realiza una media. Se puede decir que K-Folds se encarga de elaborar la selección y esta parte del experimento se encarga de dar un valor estadístico al error.

No se acude al error proporcionado por K-Folds porque en varias ocasiones puede que sólo se haya valorado una única vez, especialmente en dimensiones altas.

De entre todos estos valores se calcula el óptimo. En este punto la interpretación se realiza manualmente, ya que este óptimo queda definido por un equilibrio entre la precisión, la sensibilidad, especificidad y el área bajo la curva.

5.2.3. Cálculo de las características más repetidas

En este experimento se evalúa por separado las características más repetidas para cada dimensión. De esta manera se puede entrever qué características resultan útiles por separado y qué características se complementan bien a la hora de ofrecer una predicción.

Como ya se ha explicado, contar el número de veces que aparece una característica frente al total supone sopesar el poder generalizador de esta característica.

En concreto, la manera de proceder ha sido la siguiente:

1. A partir de los resultados de K-Fold, se colapsan en un archivo donde constan todas las repeticiones de cada combinación.
2. A partir de este número de repeticiones, se calcula el porcentaje de veces que la combinación en concreto ha resultado elegida frente al total de cada dimensión.
3. Se seleccionan las tres combinaciones más repetidas para cada combinación. También se expone el número de veces que se ha seleccionado frente al resto de su misma dimensión.
4. Se recogen los nombres y la frecuencia de las combinaciones más repetidas para cada dimensión en tablas.

5.2.4. Cálculo de otros efectos fijos

La primera iteración de los experimentos solamente tiene en cuenta una matriz de efectos fijos que estudia la edad como único efecto fijo. Esto quiere decir que se tiene en cuenta el deterioro natural, consecuencia del envejecimiento, para todos los sujetos. Sin embargo, todos estos procedimientos pueden repetirse modificando la matriz de efectos fijos.

En el experimento original se ha considerado un vector de dimensiones que va desde $D=1$ hasta $D=8$. Sin embargo, en estas repeticiones las dimensiones estudiadas pueden ser reducidas a las cercanías de la dimensión óptima ($D_{opt} \pm 1$) y la dimensión única ($D = 1$) para poder realizar los mismos experimentos. Esto es así, ya que aunque la dimensión óptima puede cambiar no se espera que lo haga en más de una dimensión.

Por ejemplo: si en el experimento original la dimensión óptima era $D=4$, en la repetición con una matriz de efectos fijos (X) distinta es posible que la dimensión óptima cambie a $D=5$ o $D=3$, pero no es probable que cambie a $D=6$.

5.3. Resultados con la edad como efecto fijo

En esta sección se muestran los distintos resultados que derivan de considerar una matriz de efectos fijos con la información referente a la edad únicamente.

5.3.1. Características más repetidas

Tras ejecutar el algoritmo de K-Folds con selección por frecuencia se han obtenido unos ratios de aparición de las combinaciones. A partir de la dimensión D el porcentaje de aparición es tan pequeño que no resulta concluyente. Es importante

recordar que las frecuencias están tomadas respecto al número de combinaciones de cada dimensión, y no respecto al total de dimensiones. Cada una de las siguientes tablas recoge los resultados para un tipo de características distinto:

únicamente ROIs		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MTT	1426/10500	13.58 %
MeT	1050/10500	10.00 %
IPT	969/10500	9.23 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MTT + LHV	1304/10500	12.42 %
IPT.lh + LHV	678/10500	6.46 %
IPT.lh + NAV	627/10500	5.97 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
IPT.lh + MTT + LHV	396/10500	3.77 %
PRT.lh + MTT + LHV	340/10500	3.24 %
MTT + LPaV + LHV	239/10500	2.28 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ECT.rh +IPT.lh +MTT +LHV	163/10500	1.55 %
TPT.lh +IPT.lh +MTT +LHV	161/10500	1.53 %
IPT.lh +MTT +LPaV +LHV	134/10500	1.28 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
TPT.lh +IPT.lh +MTT +LPaV +LHV	73/10500	0.70 %
ECT.rh +IPT.lh +MTT +LHV +NAV	55/10500	0.52 %
ECT.rh +TPT.lh +IPT.lh +MTT +LHV	49/10500	0.47 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ITT.rh +ECT.rh +TPT.lh +IPT.lh +MTT +LHV	21/10500	0.20 %
ITT.rh +TPT.lh +IPT.lh +MTT +LPaV +LAV	18/10500	0.17 %
TPT.lh +IPT.lh +MTT +LPaV +LHV +NAV	17/10500	0.16 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ITT.rh +ECT.rh +TPT.lh +IPT.lh +MTT +RAV +LHV	10/10500	0.10 %
ITT.rh +TPT.lh +MTT.lh +IPT.lh +LPaV +LAV +LHV	10/10500	0.10 %
MTT.rh +ITT.rh +ECT.rh +TPT.lh +IPT.lh +LHV +NAV	8/10500	0.08 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ITT.rh +ECT.rh +TPT.lh +IPT.lh +MTT +LPaV +LHV +NAV	6/10500	0.06 %
MTT.rh +ITT.rh +ECT.rh +TPT.lh +MTT.lh +IPT.lh +LHV +NAV	5/10500	0.05 %
MTT.rh +IPT.rh +ECT.rh +MTT.lh +IPT.lh +MeT +LHV +NAV	5/10500	0.05 %

Tabla 5.1: Frecuencia de las combinaciones más comunes contando con las características provenientes de la segmentación de FreeSurfer.

A partir de estos resultados se puede observar lo siguiente:

1. Que en combinaciones de una dimensión mayor a 4 el ratio de aparición desciende, empezando a ser despreciable.
2. Que, por separado, las medidas más representativas son el espesor del lóbulo temporal medio (MTT), el espesor medio (MeT) y el espesor parietal inferior (IPT).

Sin embargo, se observa también que al poder complementar la información con otras características, el valor medio del espesor cortical pierde relevancia. De la misma manera, el espesor del lóbulo temporal medio (MTT), así como el espesor del parietal inferior del hemisferio izquierdo (IPT_LH) empiezan a ganarla. Es decir, que estos espesores se muestran como características más relevantes.

3. Que la mayoría de características escogidas están relacionadas con el espesor cortical, y un reducido número de volúmenes complementan esta información.
4. Que al permitir la inclusión de más características, el volumen hipocampal izquierdo (LHV) complementa en la mayoría de casos la información cortical. Obsérvese que este volumen aparece en la mayoría de combinaciones a partir de dimensión 2.
5. Que las características del hemisferio izquierdo son escogidas con mayor frecuencia que las del derecho.

Al utilizar únicamente tests neuropsicológicos, las características más repetidas para cada dimensión han sido las siguientes:

únicamente tests		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ADAS13	2624/10500	24.99 %
ADASQ4	2137/10500	20.35 %
FAQ	1711/10500	16.30 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADASQ4	3075/10500	29.29 %
FAQ + ADAS13	2349/10500	22.37 %
ADASQ4 + RAVLT_immediate	1189/10500	11.32 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADASQ4 + RAVLT_immediate	1622/10500	15.45 %
FAQ + ADAS13 + RAVLT_immediate	1359/10500	12.94 %
FAQ + ADASQ4 + ADAS13	1290/10500	12.29 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MMSE + FAQ + ADASQ4 + ADAS13	1407/10500	13.40 %
CDR + FAQ + ADASQ4 + RAVLT_immediate	1136/10500	10.82 %
FAQ + ADASQ4 + ADAS13 + RAVLT_forget	1026/10500	9.77 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
CDR + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate	1917/10500	18.26 %
MMSE + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate	1303/10500	12.41 %
MMSE + FAQ + ADASQ4 + ADAS13 + RAVLT_forget	1296/10500	12.34 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
CDR + MMSE + FAQ + ADASQ4 + ADAS13 + + RAVLT_immediate	3957/10500	37.69 %
CDR + MMSE + FAQ + ADASQ4 + ADAS13 + + RAVLT_forget	1212/10500	11.54 %
CDR + FAQ + ADASQ4 + ADAS13 + RAVLT_forget + + RAVLT_immediate	952/10500	9.07 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
CDR +MMSE +FAQ +ADASQ4 +ADAS13 +RAVLT_forget + + RAVLT_immediate	3554/10500	33.85 %
CDR +MMSE +FAQ +ADASQ4 +ADAS13 +RAVLT_learn + + RAVLT_immediate	2270/10500	21.62 %
CDR + MMSE + FAQ + ADASQ4 + ADAS13 + ADAS11 + + RAVLT_immediate	2067/10500	19.69 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
CDR +MMSE +FAQ +ADASQ4 +ADAS13 +ADAS11 +RAVLT_forget + + RAVLT_immediate	4439/10500	42.28 %
CDR +MMSE +FAQ +ADASQ4 +ADAS13 +ADAS11 +RAVLT_learn + + RAVLT_immediate	3543/10500	33.74 %
CDR +MMSE +FAQ +ADASQ4 +ADAS13 +RAVLT_forget +RAVLT_learn + + RAVLT_immediate	1656/10500	15.77 %

Tabla 5.2: Frecuencia de las combinaciones más comunes contando con las características provenientes de los tests neuropsicológicos.

Respecto al significado de las siglas, se recomienda acudir al diccionario de siglas mostrado en el anexo C. Partiendo de este análisis de frecuencia de aparición se pueden resaltar los siguientes puntos:

1. Al tratar con pocas características (11 en total) las combinaciones se repiten con mayor frecuencia que en el caso de las ROIs. De hecho, en dimensiones altas ($D=8$, p.ej.), al no haber tantas combinaciones posibles, la frecuencia es inusualmente elevada.

Respecto hasta qué punto esta frecuencia está influenciada por una mejor predicción o por la poca variedad de combinaciones, se deberá comprobar al ejecutar una serie de pruebas *Train-Test* con los más repetidos. Los resultados de esta prueba se exponen en la siguiente sección.

2. Por separado, las características más representativas son ADAS13, ADASQ4 y FAQ.
3. Al aumentar la dimensión, estas tres puntuaciones se complementan especialmente bien. De hecho, se encuentran presentes en la mayoría de combinaciones en todas las dimensiones.
4. Además, en dimensiones mayores se incluyen como características complementarias el test RAVLT_immediate y las pruebas CDR y MMSE.

únicamente clusters		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
rh_cl1	1843/10500	17.55 %
lh_cl1	1569/10500	14.94 %
lh_cl8	1323/10500	12.60 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl10 +rh_cl1	912/10500	8.69 %
lh_cl9 +rh_cl1	873/10500	8.31 %
lh_cl13 +rh_cl1	783/10500	7.46 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl10 +lh_cl9 +rh_cl1	590/10500	5.62 %
lh_cl10 +lh_cl1 +rh_cl1	586/10500	5.58 %
lh_cl13 +lh_cl10 +rh_cl1	381/10500	3.63 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl10 +lh_cl9 +lh_cl1 +rh_cl1	339/10500	3.23 %
lh_cl10 +lh_cl1 +rh_cl2 +rh_cl1	286/10500	2.72 %
lh_cl14 +lh_cl9 +rh_cl2 +rh_cl1	279/10500	2.66 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl10 +lh_cl9 +lh_cl1 +rh_cl2 +rh_cl1	315/10500	3.00 %
lh_cl14 +lh_cl13 +lh_cl9 +rh_cl2 +rh_cl1	277/10500	2.64 %
lh_cl14 +lh_cl10 +lh_cl9 +rh_cl2 +rh_cl1	165/10500	1.57 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl14 +lh_cl10 +lh_cl9 +lh_cl1 +rh_cl2 +rh_cl1	130/10500	1.24 %
lh_cl14 +lh_cl10 +lh_cl9 +lh_cl8 +rh_cl2 +rh_cl1	118/10500	1.12 %
lh_cl14 +lh_cl13 +lh_cl10 +lh_cl9 +rh_cl2 +rh_cl1	105/10500	1.00 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl14 +lh_cl13 +lh_cl10 +lh_cl9 +lh_cl1 +rh_cl2 +rh_cl1	76/10500	0.72 %
lh_cl14 +lh_cl10 +lh_cl9 +lh_cl8 +lh_cl1 +rh_cl2 +rh_cl1	72/10500	0.69 %
lh_cl10 +lh_cl9 +lh_cl8 +lh_cl1 +rh_cl5 +rh_cl2 +rh_cl1	71/10500	0.68 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl14 +lh_cl13 +lh_cl10 +lh_cl9 +lh_cl1 +rh_cl5 +rh_cl2 +rh_cl1	75/10500	0.71 %
lh_cl14 +lh_cl10 +lh_cl9 +lh_cl8 +lh_cl1 +rh_cl5 +rh_cl2 +rh_cl1	72/10500	0.69 %
lh_cl14 +lh_cl13 +lh_cl9 +lh_cl8 +lh_cl1 +rh_cl5 +rh_cl2 +rh_cl1	57/10500	0.54 %

Tabla 5.3: Frecuencia de las combinaciones más comunes contando con las características provenientes del análisis masivo.

Para la localización de los clusters más relevantes se pueden consultar las imágenes 5.1 y 5.2. A partir de estos resultados se puede llegar a varias conclusiones:

1. Que el porcentaje de repetición a partir de dimensiones mayores a 6 comienza a descender ($< 1\%$) debido a la gran variedad de combinaciones posibles.
2. Que los clusters más relevantes por sí solos son el clúster 1 y 8 del hemisferio izquierdo y el clúster 1 del hemisferio derecho.
3. Que al aumentar la dimensión sólo el clúster 1 del hemisferio derecho se complementa bien con otras características.

4. Que en dimensiones superiores surgen nuevas características que combinadas sustituyen a las mejores características individuales. En concreto destacan los clusters 9 y 10 del hemisferio izquierdo.
5. Que, por lo general, las características del hemisferio izquierdo aparecen con mayor frecuencia que las del hemisferio derecho.

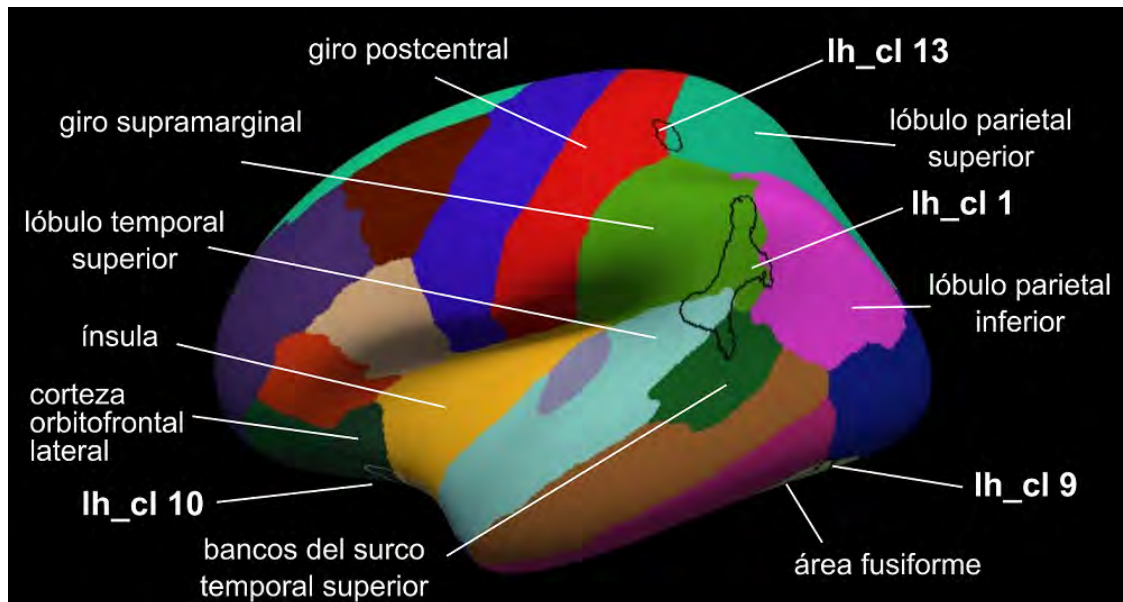


Figura 5.1: Esquema de la distribución de los clusters 1 y 9 del hemisferio izquierdo. Estos dos clusters aparecen repetidamente en las combinaciones de características.

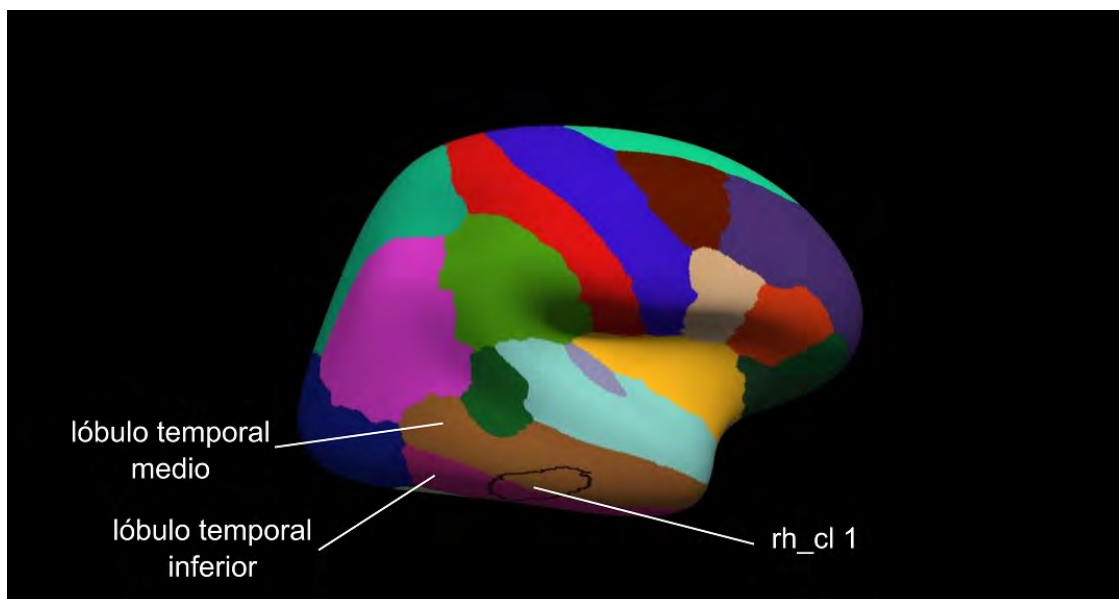


Figura 5.2: Esquema del cluster 1 del hemisferio derecho. Este cluster aparece repetidamente en las combinaciones de características.

Al evaluar apoyándose en las características provenientes de las ROIs y los tests neuropsicológicos, las combinaciones más repetidas han sido las siguientes:

ROIs y Tests neuropsicológicos		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ADAS13	2341/10500	22.30 %
ADASQ4	2076/10500	19.77 %
FAQ	1882/10500	17.92 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADAS13	1952/10500	18.59 %
FAQ + ADASQ4	1804/10500	17.18 %
FAQ + IPT_lh	494/10500	4.70 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADASQ4 + IPT_lh	1146/10500	10.91 %
FAQ + ADAS13 + IPT_lh	1058/10500	10.08 %
FAQ + ADASQ4 + PRT_lh	779/10500	7.42 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADASQ4 + ADAS13 + IPT_lh	338/10500	3.22 %
FAQ + ADASQ4 + IPT_lh + NAV	307/10500	2.92 %
FAQ + ADAS13 + IPT_lh + NAV	240/10500	2.29 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADASQ4 + ADAS13 + TPT_lh + IPT_lh	256/10500	2.44 %
FAQ + ADASQ4 + ADAS13 + IPT_lh + NAV	251/10500	2.39 %
FAQ + ADASQ4 + ADAS13 + IPT_lh + LHV	219/10500	2.09 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MMSE + FAQ + ADASQ4 + ADAS13 + ITT_rh + IPT_lh	97/10500	1.10 %
MMSE + FAQ + ADASQ4 + ADAS13 + IPT_lh + NAV	86/10500	1.04 %
FAQ + ADASQ4 + ADAS13 + ITT_rh + IPT_lh + LHV	79/10500	0.94 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MMSE + FAQ + ADASQ4 + ADAS13 + ITT_rh + IPT_lh + NAV	37/10500	0.35 %
CDR + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate + IPT_lh + NAV	32/10500	0.30 %
MMSE + FAQ + ADASQ4 + ADAS13 + ITT_rh + IPT_lh + LHV	29/10500	0.28 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
CDR + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate + IPT_lh + MTT + LAV	16/10500	0.15 %
CDR + MMSE + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate + IPT_lh + NAV	16/10500	0.15 %
CDR + MMSE + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate + ITT_rh + IPT_lh	16/10500	0.15 %

Tabla 5.4: Frecuencia de las combinaciones más comunes contando con las características provenientes de la segmentación de FreeSurfer y los tests neuropsicológicos.

A partir de estos resultados se puede destacar lo siguiente:

1. Que la información fundamental es aportada por los tests neuropsicológicos incluso para dimensiones bajas.
2. Que los tests neuropsicológicos más relevantes son el FAQ, el ADAS13 y el ADASQ4. En dimensiones mayores que 6 MMSE hace su aparición como una característica complementaria.

3. Que dentro de las ROIs, aquellas que mayor importancia cobran son sin duda el volumen de la amígdala (NAV) y el espesor del lóbulo parietal inferior izquierdo (IPT_lh).
4. Que de entre las ROIs seleccionadas, existe una abundancia de características provenientes del hemisferio izquierdo.

Al evaluar apoyándose en las características provenientes de las ROIs y los clusters, las combinaciones más repetidas han sido las siguientes:

ROIs y Clusters		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MTT	1403/10500	13.36 %
rh.cl1	1081/10500	10.30 %
IPT	814/10500	7.75 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
MTT +LHV	958/10500	9.12 %
rh.cl1 +LHV	484/10500	4.61 %
lh.cl1 +LHV	372/10500	3.54 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl13 +MTT +LHV	449/10500	4.28 %
lh.cl9 +MTT +LHV	396/10500	3.77 %
lh.cl13 +rh.cl1 +LHV	237/10500	2.26 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl13 +lh.cl9 +MTT +LHV	210/10500	2.00 %
lh.cl9 +lh.cl8 +MTT +LHV	68/10500	0.65 %
lh.cl9 +rh.cl5 +MTT +LHV	64/10500	0.61 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl13 +lh.cl9 +rh.cl5 +MTT +LHV	34/10500	0.32 %
lh.cl13 +lh.cl10 +lh.cl1 +rh.cl1 +LHV	26/10500	0.25 %
lh.cl14 +lh.cl13 +lh.cl9 +MTT +LHV	23/10500	0.22 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl14 +lh.cl13 +lh.cl9 +rh.cl5 +MTT +LHV	8/0.08	1.10 %
lh.cl14 +lh.cl13 +lh.cl9 +rh.cl5 +rh.cl1 +NAV	7/0.07	1.04 %
lh.cl10 +lh.cl9 +lh.cl1 +rh.cl5 +rh.cl1 +LHV	7/0.07	0.94 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl13 +lh.cl9 +rh.cl5 +rh.cl1 +TPT_lh +IPT_lh +LHV	5/10500	0.05 %
lh.cl13 +lh.cl9 +lh.cl1 +rh.cl1 +MTT +RAV +LHV	4/10500	0.04 %
lh.cl14 +lh.cl13 +lh.cl9 +rh.cl1 +TPT_lh +MTT +LHV	4/10500	0.04 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl9 +lh.cl8 +rh.cl5 +rh.cl1 +TPT_LH +MTT +LHV +NAV	3/10500	0.03 %
lh.cl13 +lh.cl9 +lh.cl1 +rh.cl5 +rh.cl3 +rh.cl1 +IPT_lh +LHV	3/10500	0.03 %
lh.cl10 +lh.cl9 +lh.cl8 +lh.cl1 +rh.cl5 +rh.cl1 +LHV +NAV	3/10500	0.03 %

Tabla 5.5: Frecuencia de las combinaciones más comunes contando con las características provenientes de la segmentación de FreeSurfer y los clusters provenientes del análisis masivo.

A partir de estos resultados se pueden extraer las siguientes conclusiones:

1. A partir de la dimensión $D=4$ la frecuencia de repetición deja de ser representativa.
2. Los clusters provenientes del análisis masivo adquieren un papel predominante respecto a las ROIs.
3. Dentro de los clusters, los más repetidos son los ya mencionados en el experimento en el que se consideraban únicamente los clusters. Esto es, los clusters 1, 9, 10 y 13 del hemisferio izquierdo y el clúster 1 del hemisferio derecho.
4. De entre todas las ROIs destacan dos en concreto: el volumen hipocampal izquierdo (LHV) y el espesor del lóbulo temporal medio (MTT). Estas características se encuentran presentes desde las dimensiones más bajas.
5. Por lo general, al igual que en otros casos, hay una predilección por las características del hemisferio izquierdo.

Al evaluar apoyándose en las características provenientes de los clusters y los tests neuropsicológicos, las combinaciones más repetidas han sido las siguientes:

Clusters y Tests neuropsicológicos		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ADAS13	2422/10500	23.07 %
ADASQ4	2125/10500	20.24 %
FAQ	1886/10500	17.96 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADAS13	1453/10500	13.84 %
FAQ + ADASQ4	1155/10500	11.00 %
rh_cl1 + ADAS13	1046/10500	9.96 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
rh_cl1 + FAQ + ADAS13	1396/10500	13.30 %
lh_cl9 + FAQ + ADASQ4	1192/10500	11.35 %
rh_cl1 + FAQ + ADASQ4	1186/10500	11.30 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl9 + rh_cl1 + FAQ + ADASQ4	606/10500	5.77 %
lh_cl9 + rh_cl1 + FAQ + ADAS13	548/10500	5.22 %
lh_cl1 + rh_cl1 + FAQ + ADAS13	433/10500	4.12 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl9 + rh_cl1 + FAQ + ADASQ4 + ADAS13	648/10500	6.17 %
lh_cl1 + rh_cl1 + FAQ + ADASQ4 + ADAS13	533/10500	5.08 %
lh_cl9 + lh_cl1 + FAQ + ADASQ4 + ADAS13	373/10500	3.55 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl9 + lh_cl1 + rh_cl1 + FAQ + ADASQ4 + ADAS13	293/10500	2.79 %
lh_cl9 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13	226/10500	2.15 %
lh_cl10 + lh_cl1 + rh_cl1 + FAQ + ADASQ4 + ADAS13	180/10500	1.71 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl9 + lh_cl1 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13	190/10500	1.81 %
lh_cl9 + lh_cl8 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13	116/10500	1.10 %
lh_cl14 + lh_cl9 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13	91/10500	0.87 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh_cl10 + lh_cl9 + lh_cl1 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13	49/10500	0.47 %
lh_cl9 + lh_cl1 + rh_cl1 + MMSE + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate	39/10500	0.37 %
lh_cl9 + lh_cl1 + rh_cl1 + CDR + FAQ + ADASQ4 + ADAS13 + RAVLT_immediate	38/10500	0.36 %

Tabla 5.6: Frecuencia de las combinaciones más comunes contando con las características provenientes de la segmentación de FreeSurfer y los tests neuropsicológicos.

A partir de estos resultados se obtiene lo siguiente:

1. Se mantiene que los tests aportan la mayor parte de la información y que los más relevantes son FAQ, ADAS13 Y ADASQ4.
2. Se mantiene que los clusters más relevantes están relacionados en su mayoría con el hemisferio izquierdo (lh_cl1,9 y 10) y el clúster 1 del hemisferio derecho (rh_cl1). La situación de estos clusters se puede consultar en las figuras 5.1 y 5.2.

Al evaluar apoyándose en todas las características, las combinaciones más repetidas han sido las siguientes:

todas las características		
D = 1. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
ADAS13	2395/10500	22.81 %
ADASQ4	2154/10500	20.51 %
FAQ	1860/10500	17.71 %
D = 2. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
FAQ + ADAS13	1385/10500	13.19 %
FAQ + ADASQ4	1124/10500	10.70 %
rh.cl1 + ADAS13	869/10500	8.28 %
D = 3. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
rh.cl1 + FAQ + ADAS13	1069/10500	10.18 %
lh.cl9 + FAQ + ADASQ4	1021/10500	9.72 %
lh.cl1 + FAQ + ADAS13	889/10500	8.47 %
D = 4. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl9 + rh.cl1 + FAQ + ADASQ4	335/10500	3.19 %
lh.cl9 + rh.cl1 + FAQ + ADAS13	300/10500	2.86 %
lh.cl1 + FAQ + ADASQ4 + ADAS13	216/10500	2.06 %
D = 5. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl1 + rh.cl1 + FAQ + ADASQ4 + ADAS13	308/10500	2.93 %
lh.cl9 + rh.cl1 + FAQ + ADASQ4 + ADAS13	273/10500	2.60 %
lh.cl9 + lh.cl1 + FAQ + ADASQ4 + ADAS13	147/10500	1.40 %
D = 6. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl9 + rh.cl1 + FAQ + ADASQ4 + ADAS13 + NAV	116/10500	1.10 %
lh.cl9 + lh.cl1 + rh.cl1 + FAQ + ADASQ4 + ADAS13	109/10500	1.04 %
lh.cl10 + lh.cl1 + rh.cl1 + FAQ + ADASQ4 + ADAS13	99/10500	0.94 %
D = 7. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl9 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13 + NAV	39/10500	0.37 %
lh.cl9 + lh.cl1 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13	38/10500	0.36 %
lh.cl10 + lh.cl1 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13	32/10500	0.30 %
D = 8. Combinaciones totales: 10500		
	Frec. absoluta	Frec. relativa
lh.cl10 + lh.cl9 + lh.cl1 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13	12/10500	0.11 %
lh.cl9 + lh.cl1 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13 + NAV	9/10500	0.09 %
lh.cl9 + lh.cl1 + rh.cl1 + MMSE + FAQ + ADASQ4 + ADAS13 + LAV	8/10500	0.08 %

Tabla 5.7: Frecuencia de las combinaciones más comunes contando con las características provenientes de la segmentación de FreeSurfer.

Partiendo de este análisis de frecuencia de aparición se pueden resaltar los siguientes puntos:

1. El porcentaje de repetición a partir de dimensiones mayores a 5 comienza a descender ($< 1\%$) debido a la gran variedad de combinaciones posibles.
2. Los tests neuropsicológicos aportan la mayor parte del poder discriminatorio dentro de las combinaciones, al igual que sucedía al realizar otro tipo de com-

binaciones entre familias de características.

3. Los clusters obtenidos del análisis masivo desplazan a la información de las ROIs, ya que están creados atendiendo al desarrollo de la enfermedad en lugar de a estructuras anatómicas.
4. Los clusters relevantes han sido rh_cl1, lh_cl1 y lh_cl9. Todos ellos guardan relación con las áreas del lóbulo temporal, tal y como se puede apreciar en las figuras 5.1 y 5.2.
5. Las ROIs más destacadas están relacionadas con el volumen de la amígdala (NAV, LAV). Es importante mencionar que la información recogida por el espesor del lóbulo parietal inferior complementa pobremente a los clusters del análisis masivo, mientras que el volumen de la amígdala sigue siendo una característica relevante.

5.3.2. Dimensión óptima

Una vez seleccionadas las combinaciones de características más repetidas, se ha ejecutado una serie de procesos de entrenamiento-test con distintos subconjuntos y se ha calculado el error medio. De este error medio para cada combinación de cada dimensión se extraen los siguientes gráficos, cada uno correspondiente a una familia de características distinta:

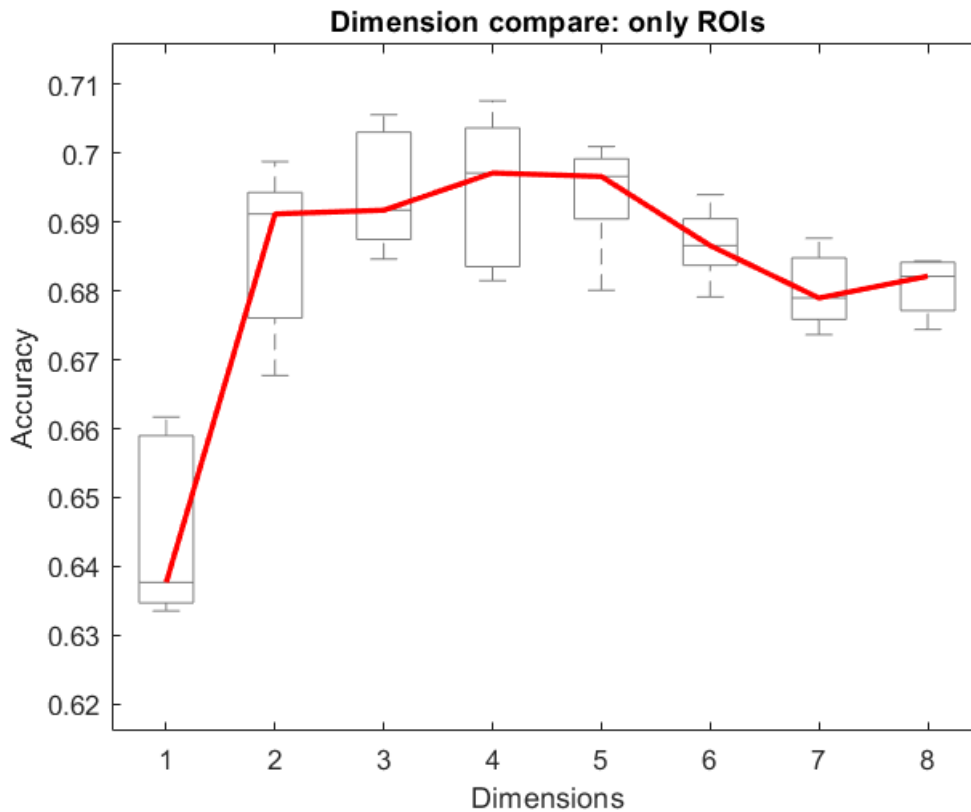


Figura 5.3: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta únicamente las ROIs.

Se puede observar que la dimensión óptima al considerar únicamente las características provenientes de la segmentación de FreeSurfer se puede situar en torno a $D=4$ ó $D=5$. Atendiendo a la distribución se puede observar que combinaciones de cuatro características otorgan un resultado algo más elevado.

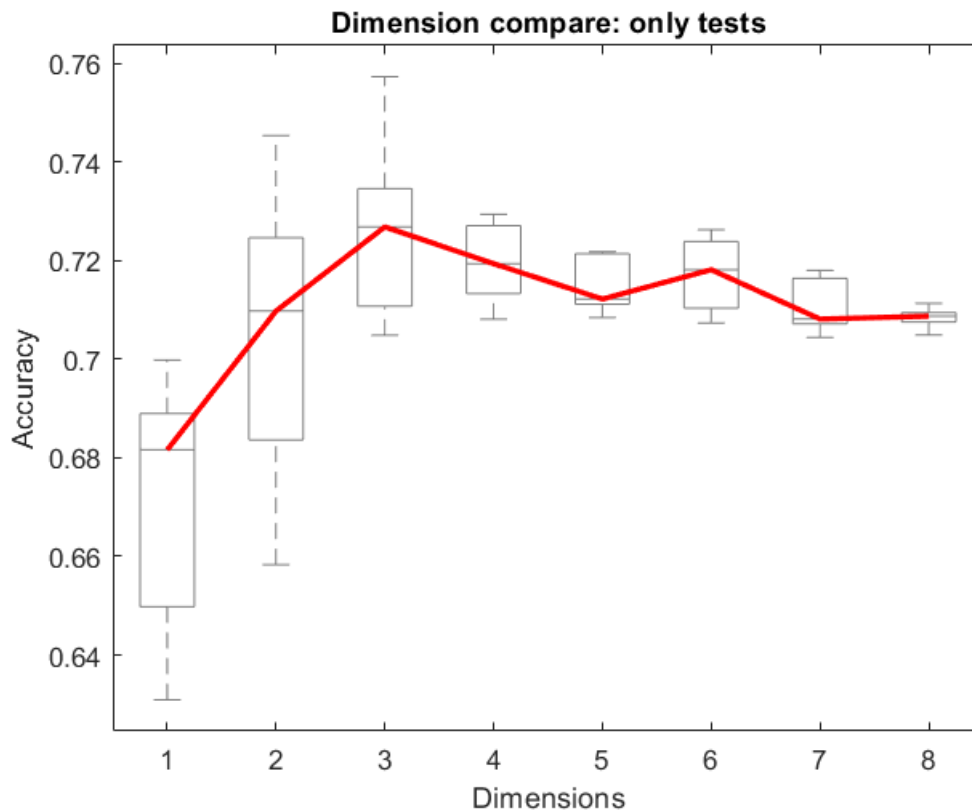


Figura 5.4: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta únicamente los tests neuropsicológicos.

Se puede observar que la dimensión óptima al considerar únicamente los tests neuropsicológicos se sitúa en $D=3$. En este caso no hay duda al respecto, ya que la diferencia es evidente.

Además se puede apreciar que la medida de error, por lo general, es superior a la vista en el caso de las ROIs. Esto confirma, como ya se dejaba intuir por la selección de características, que los tests neuropsicológicos tienen un mayor poder de clasificación que las ROIs.

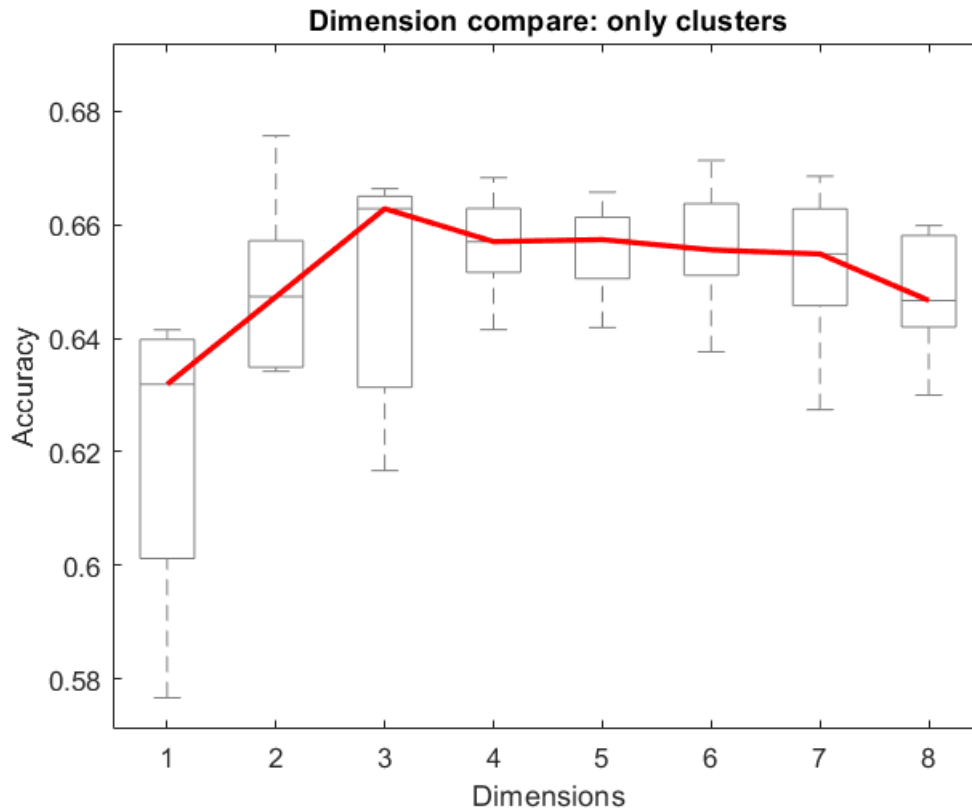


Figura 5.5: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta únicamente los clusters provenientes del análisis masivo.

Se puede observar cómo la dimensión óptima se encuentra en $D=3$. Sin embargo, se puede comprobar que el error es mayor que en cualquiera de los casos anteriores.

Por lo tanto, el auténtico poder de clasificación de estas características se manifiesta al ser usadas como complemento de los tests neuropsicológicos.

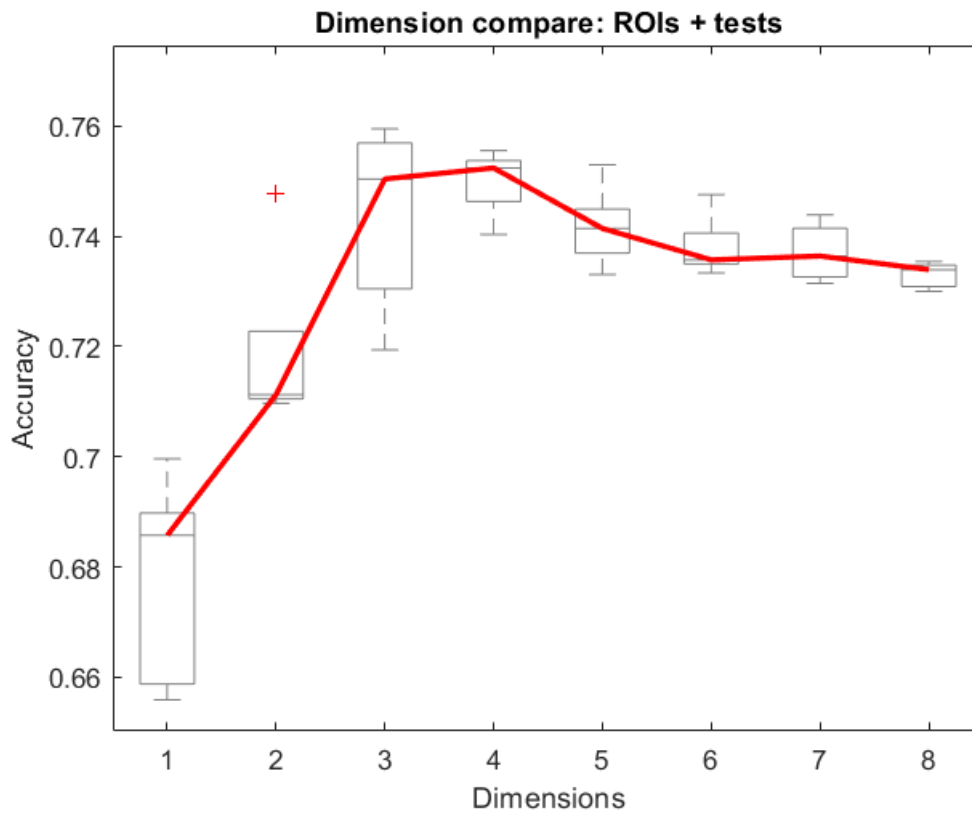


Figura 5.6: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta las ROIs y los tests neuropsicológicos.

Se puede observar cómo la dimensión óptima se encuentra entre $D=3$ y $D=4$. No obstante, la distribución de errores indica que es preferible $D=4$.

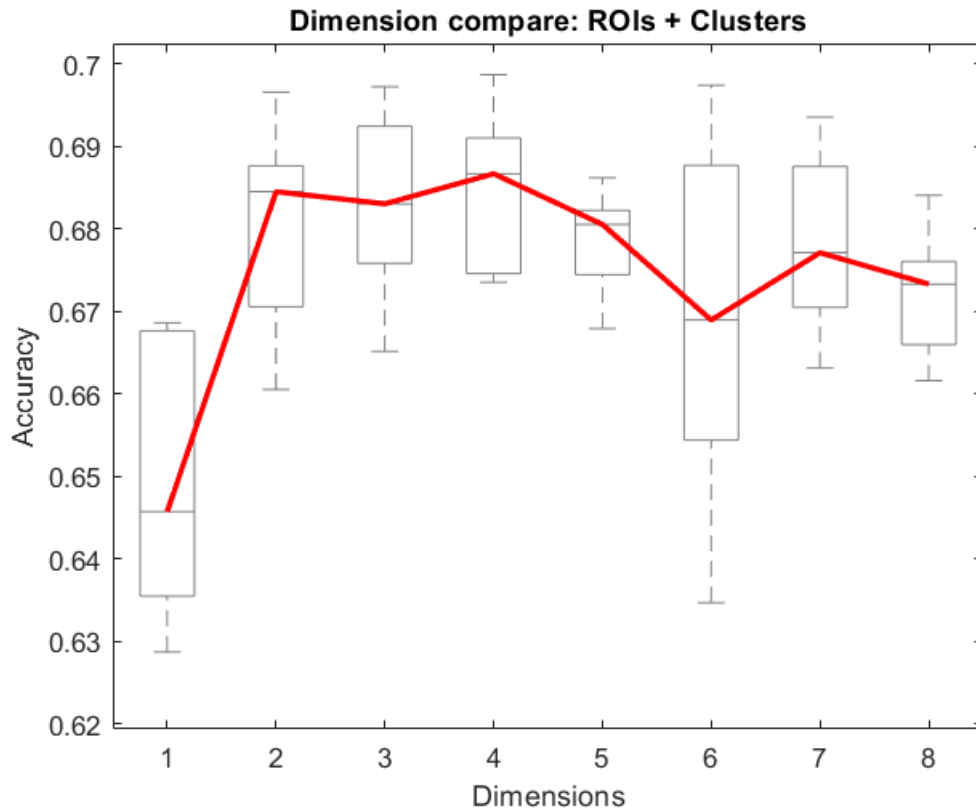


Figura 5.7: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta las ROIs y los clusters provenientes del análisis masivo.

Se puede observar cómo la dimensión óptima se encuentra en $D=4$. Sin embargo, el error es mayor que cualquiera de las combinaciones resultantes de combinar ROIs o clusters con los tests neuropsicológicos.

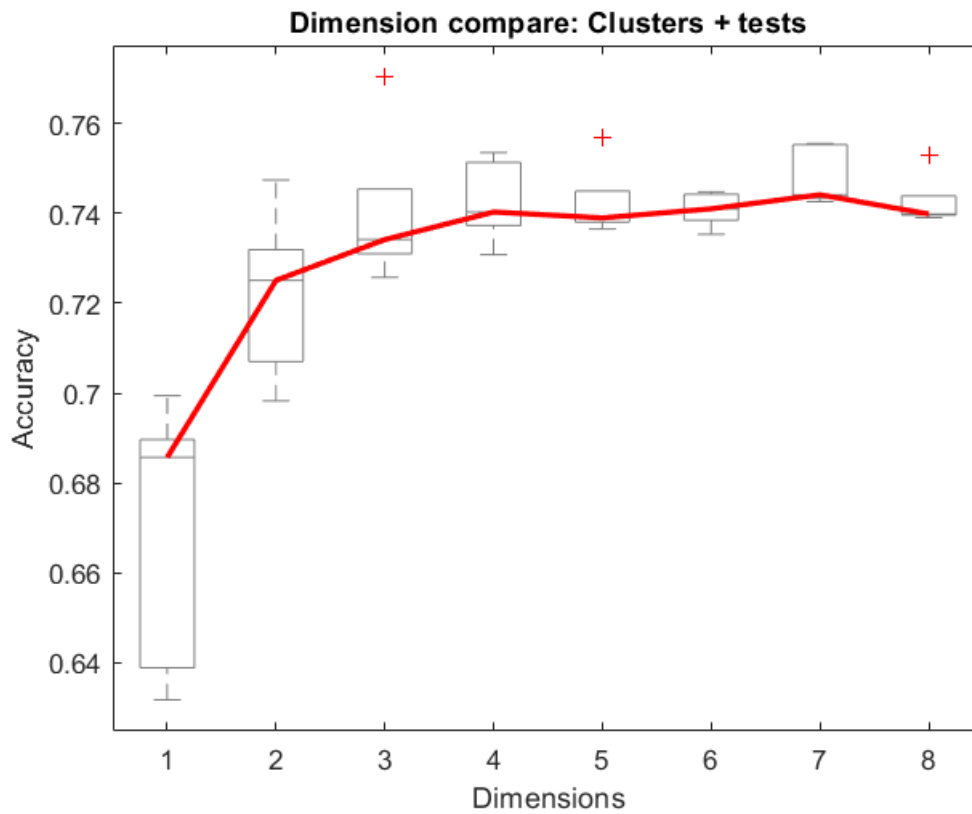


Figura 5.8: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta los clusters y los tests neuropsicológicos.

Se puede observar cómo la dimensión óptima se encuentra entre $D=4$ y $D=7$. De hecho, la variación del error, incluso al fijarse en la distribución, no resuelve el conflicto. Teniendo en cuenta las recomendaciones en la literatura de escoger una combinación lo más reducida posible con una explicación intuitiva, se ha decidido designar $D=4$ como la dimensión óptima.

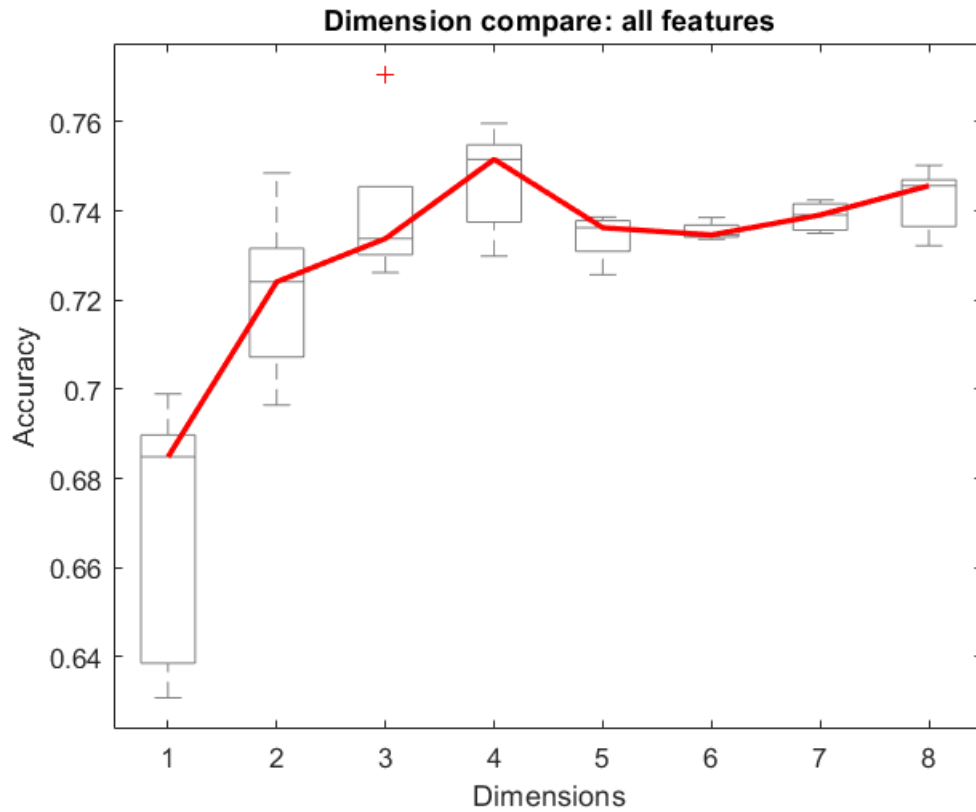


Figura 5.9: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas teniendo en cuenta todas las características.

Se puede observar que el máximo al considerar todas las características se encuentra para la dimensión $D=4$. Es de esta dimensión de donde se ha escogido la exactitud (*accuracy*) máxima a partir de la información del baseline como valor representativo del estudio.

5.3.3. Mejora de la predicción con todas las visitas

En la siguiente tabla se muestra la mejora que se observa al contar con todas las imágenes posibles de cada paciente. Aunque no sirva para comparar con otros estudios, es importante para entender el poder de los estudios longitudinales. El resultado está dado como los errores medios de la combinación óptima hallada.

únicamente ROIs				
	Sen %	Spec %	Acc %	AUC %
baseline	76.57 (8.58)	64.19 (9.83)	70.24 (5.86)	75.88 (6.14)
todas las visitas	74.45 (9.20)	72.90 (8.90)	73.65 (5.67)	79.20 (5.68)
únicamente tests				
	Sen %	Spec %	Acc %	AUC %
baseline	84.35 (7.83)	67.47 (9.14)	75.73 (5.42)	81.73 (5.17)
todas las visitas	73.22 (8.95)	83.65 (6.83)	78.54 (5.02)	88.06 (4.02)
únicamente clusters				
	Sen %	Spec %	Acc %	AUC %
baseline	68.26 (9.21)	65.10 (9.54)	66.64 (6.06)	70.49 (6.71)
todas las visitas	68.86 (9.71)	73.94 (9.68)	71.45 (6.00)	74.72 (6.37)
ROIs + tests				
	Sen %	Spec %	Acc %	AUC %
baseline	83.00 (8.12)	67.98 (9.33)	75.31 (5.48)	82.80 (5.21)
todas las visitas	79.15 (8.52)	84.31 (6.56)	81.79 (4.91)	88.56 (4.08)
ROIs + Clusters				
	Sen %	Spec %	Acc %	AUC %
baseline	72.22 (9.09)	67.68 (9.62)	69.87 (6.06)	74.63 (6.15)
todas las visitas	71.30 (8.74)	73.46 (8.48)	72.39 (5.66)	77.74 (5.73)
Tests + Clusters				
	Sen %	Spec %	Acc %	AUC %
baseline	82.68 (8.11)	68.32 (8.98)	75.35 (5.40)	83.27 (5.04)
todas las visitas	72.56 (9.01)	82.55 (6.53)	77.67 (4.98)	88.77 (3.84)
todas las características				
	Sen %	Spec %	Acc %	AUC %
baseline	83.81 (7.41)	68.44 (9.31)	75.97 (5.26)	83.38 (5.02)
todas las visitas	74.83 (8.53)	82.93 (6.58)	78.98 (4.96)	88.54 (3.84)

Tabla 5.8: Mejora de los resultados para el óptimo con las distintas combinaciones de características, contando únicamente con la primera visita (baseline) o con todas las visitas del mismo sujeto.

A partir de estos resultados se pueden sacar las siguientes conclusiones:

1. Que al tener información de más visitas a lo largo del tiempo la precisión de la predicción aumenta. Este efecto es independiente de la familia de características. He aquí el auténtico poder de los estudios longitudinales.
2. Que al aumentar el número de imágenes, dentro del error general, se sacrifica algo de sensibilidad a cambio de una mejora sustancial en la especificidad.
3. Que la combinación de todas las características posee un error menor de predicción, seguido de la información de las combinaciones con tests (Clusters+Tests, ROIs+Tests).
4. Que aunque en la combinación general los clusters desplazan a las ROIs, la combinación de ambas con la información de tests ofrece resultados bastante similares.

- Que al comparar los resultados al contar únicamente con clusters o ROIs se puede ver que las ROIs ofrecen un error considerablemente menor, probablemente debido a que no sólo contienen información cortical, sino también volumétrica. Debe recordarse que, de hecho, el volumen de la amígdala es una de las características más repetidas.

5.4. Resultados con la edad y APOE4 como efectos fijos

A continuación se reflejan los resultados al utilizar como efectos fijos entre la población únicamente la edad y la presencia del alelo APOE4. Tanto en este experimento como en los siguientes debe recordarse que las combinaciones de características escogidas por frecuencia sólo varían sensiblemente respecto al experimento principal (con la matriz de efectos fijos que sólo contempla la edad). Esto es debido a que el algoritmo mRMR previo a la selección por medida del error en los K-Folds internos está reaccionando al mismo input. Es decir, a las mismas características.

5.4.1. Dimensión óptima

Al ejecutar la consecución de entrenamientos-tests sobre las combinaciones más repetidas de cada dimensión se ha obtenido el siguiente resultado:

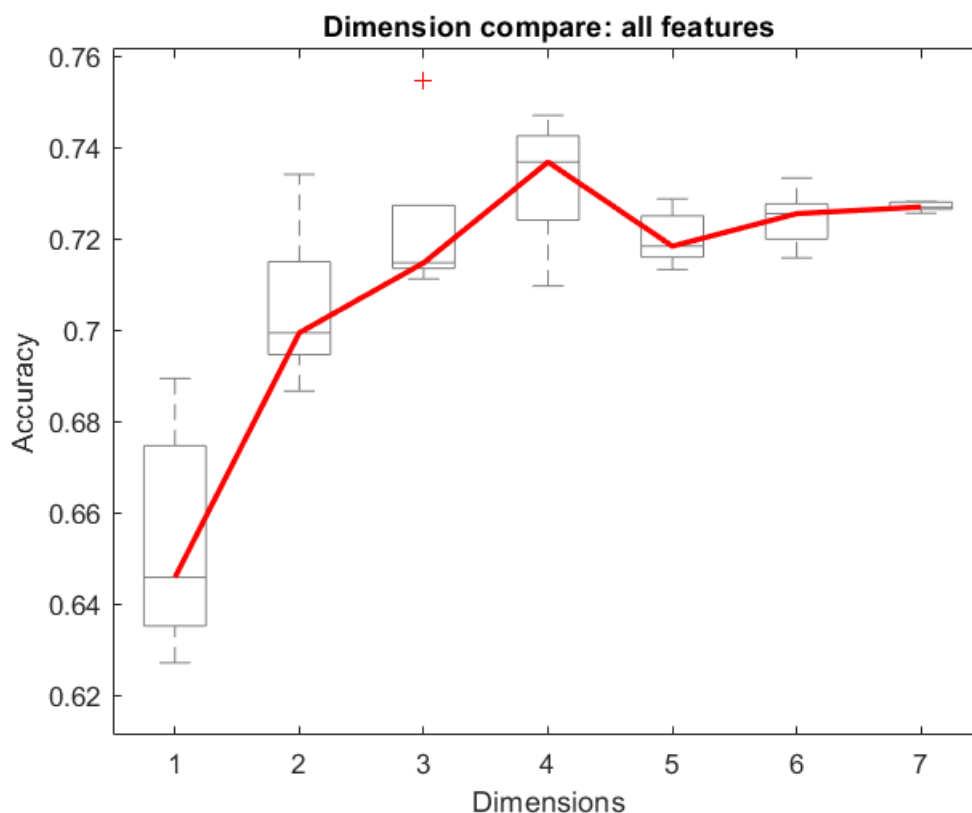


Figura 5.10: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones.

En el gráfico se puede observar a simple vista que la dimensión óptima es de combinaciones de 4 características. De esta dimensión se obtendrá la mejor puntuación de error medio. A esta puntuación (contando únicamente con imágenes de baseline) se hará referencia al hablar de los resultados con esta matriz de efectos fijos:

Matriz de efectos fijos	Sensibilidad	Especificidad	Exactitud	AUC
KFSF_all (X=edad+APOE4)	82.23 (8.46)	67.55 (9.37)	74.71 (5.76)	82.52 (5.07)

Tabla 5.9: Mejor medida de error cuando la matriz de efectos fijos tiene en cuenta tanto la edad como el alelo APOE4.

5.5. Resultados con la edad y la educación como efectos fijos

A continuación se reflejan los resultados al utilizar como efectos fijos entre la población únicamente la edad y los años de educación.

5.5.1. Dimensión óptima

Al ejecutar la consecución de entrenamientos-tests sobre las combinaciones más repetidas de cada dimensión se ha obtenido el siguiente resultado:

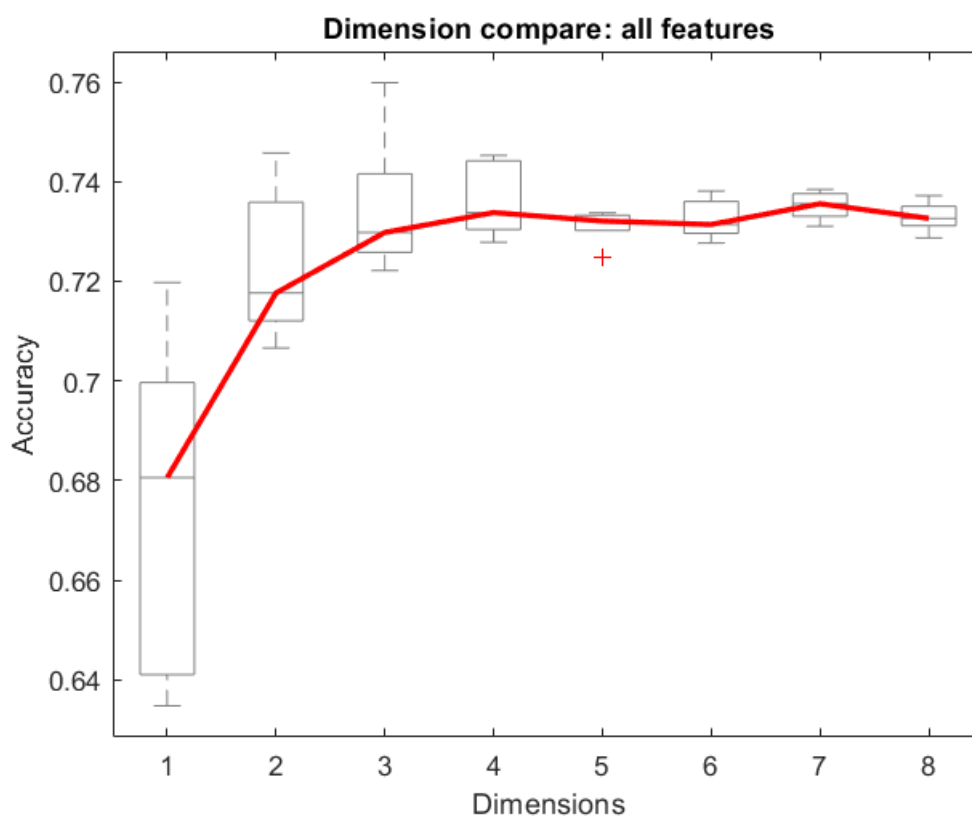


Figura 5.11: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas.

En el gráfico se puede observar que en principio no existe una dimensión predominante, a partir de $D=4$. Sin embargo, atendiendo a las distribuciones y medianas, se puede considerar precisamente $D=4$ como la dimensión óptima. Es de esta dimensión de donde se obtendrá la mejor puntuación de error medio. Esta puntuación (contando únicamente con imágenes de baseline) es a la que se hará referencia al hablar de los resultados con esta matriz de efectos fijos:

Matriz de efectos fijos	Sensibilidad	Especificidad	Exactitud	AUC
KFSF_all (X=edad+educación)	82.65 (8.27)	66.54 (9.61)	74.38 (5.60)	82.71 (5.14)

Tabla 5.10: Mejor medida de error cuando la matriz de efectos fijos tiene en cuenta tanto la edad como los años de educación.

5.6. Resultados con la edad, la educación y APOE4 como efectos fijos

A continuación se reflejan los resultados al utilizar como efectos fijos entre la población únicamente la edad y los años de educación.

5.6.1. Dimensión óptima

Al ejecutar la consecución de entrenamientos-tests sobre las combinaciones más repetidas de cada dimensión se ha obtenido el siguiente resultado:

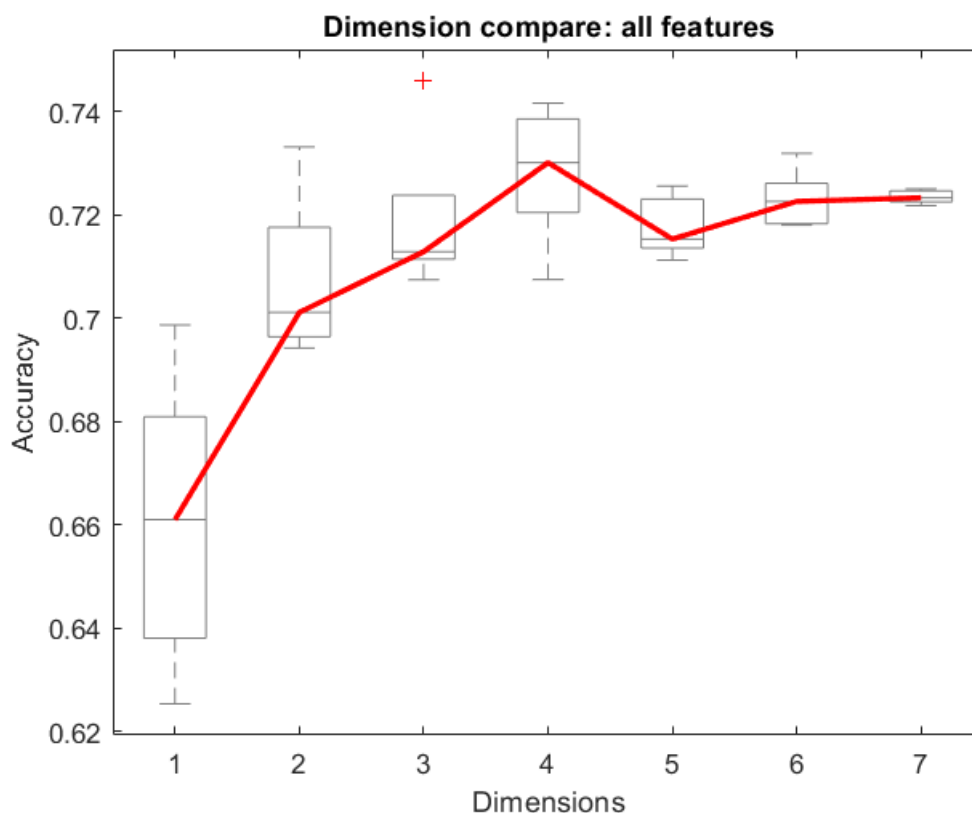


Figura 5.12: Gráfica donde se recogen los resultados medios (medida de exactitud (*accuracy*)) para todas las dimensiones estudiadas.

En el gráfico se puede observar que en la dimensión $D=4$ ofrece mejores resultados. De esta dimensión se obtendrá la mejor puntuación de error medio. A esta puntuación (contando únicamente con imágenes de baseline) se hará referencia al hablar de los resultados con esta matriz de efectos fijos:

Matriz de efectos fijos	Sensibilidad	Especificidad	Exactitud	AUC
KFSF_all (X=edad+educación+APOE4)	82.20 (8.49)	66.52 (9.42)	74.16 (5.71)	82.28 (5.09)

Tabla 5.11: Mejor medida de error cuando la matriz de efectos fijos tiene en cuenta tanto la edad como los años de educación.

5.7. Comparativa entre distintas matrices de efectos fijos

Se procede en este apartado a realizar una comparativa entre los resultados con distintas matrices de efectos fijos. La comparativa se realizará mediante la información del baseline y apoyándose en la evolución del error según se dispone de más imágenes:

contando con baseline				
Matriz de efectos fijos	Sensibilidad	Especificidad	Exactitud	AUC
KFSF_all (X=edad)	83.81 (7.41)	68.44 (9.31)	75.97 (5.26)	83.38 (5.02)
KFSF_all (X=edad+APOE4)	82.23 (8.46)	67.55 (9.37)	74.71 (5.76)	82.52 (5.07)
KFSF_all (X=edad+educación)	82.65 (8.27)	66.54 (9.61)	74.38 (5.60)	82.71 (5.14)
KFSF_all (X=edad+educación+APOE4)	82.20 (8.49)	66.52 (9.42)	74.16 (5.71)	82.28 (5.09)
contando con todas las visitas				
Matriz de efectos fijos	Sensibilidad	Especificidad	Exactitud	AUC
KFSF_all (X=edad)	74.83 (8.53)	82.93 (6.58)	78.98 (4.96)	88.54 (3.84)
KFSF_all (X=edad+APOE4)	71.58 (9.49)	82.66 (6.47)	77.24 (5.11)	88.21 (3.92)
KFSF_all (X=edad+educación)	72.89 (9.08)	82.13 (7.19)	77.61 (5.27)	89.01 (4.10)
KFSF_all (X=edad+educación+APOE4)	71.65 (9.46)	82.57 (6.59)	77.23 (5.06)	87.97 (3.91)

Tabla 5.12: Comparativa de errores según las matrices de efectos fijos.

Como se puede observar, ninguna de las nuevas matrices de efectos fijos aporta información útil, por lo que realmente la mejor predicción se encuentra al considerar únicamente el deterioro natural causado por la edad. Como se ha podido observar, para todas estas matrices de efectos fijos la dimensión óptima se encuentra en $D=4$, y las combinaciones utilizadas se basan principalmente en tests neuropsicológicos apoyados con información de los clusters.

5.8. Comparativa con otros estudios

Para poder comparar con estudios transversales, se utiliza aquí la predicción a partir de la información del baseline, contando con la edad como único efecto fijo.

Estudio	Sensibilidad	Especificidad	Exactitud	AUC
Eskilsen 2015 (a)	60.9 %	66.1 %	63.9 %	66.4 %
Eskilsen 2015 (b)	67.7 %	69.6 %	68.8 %	71.5 %
Eskilsen 2015 (c)	64.6 %	69.2 %	67.3 %	70.1 %
Eskilsen 2015 (d)	65.8 %	69.2 %	67.8 %	72.8 %
Eskilsen 2015 (e)	67.7 %	70.9 %	69.6 %	73.3 %
Eskilsen 2015 (f)	69.6 %	73.6 %	71.9 %	76.6 %
Korolev 2016	83.4 %	76.4 %	79.9 %	87.0 %
KFSF_all (X=edad)	83.81 (7.41)	68.44 (9.31)	75.97 (5.26)	83.38 (5.02)

Tabla 5.13: Comparación de los resultados obtenidos en este estudio mediante KFSF aplicado a todas las características frente a otros estudios. En el caso de Eskilsen: a)63 ROI b)4 gradientes + 4 volúmenes c)71 características combinadas d)3 ROI e)L/R gradiente hipocampal f)L/R gradiente hipocampal + 3 ROI.

5.9. Discusión

Sobre los resultados obtenidos, cabe destacar que la combinación $mRMR+LME$ plantea una mejora respecto a estudios transversales, ofreciendo mejores resultados incluso contando con la misma información (baseline).

También cabe destacar que no se han podido igualar las puntuaciones obtenidas por Korolev[17]. Esto hace pensar que el acercamiento mediante biomarcadores de plasma y la decisión de habilitar una tercera clase, *ambiguo*, han marcado la diferencia entre ambos estudios.

Respecto a las distintas características, tanto los tests más frecuentes (FAQ, ADAS13, ADASQ4), las ROIs (Amígdala y espesor del lóbulo parietal inferior) y los clusters (zonas temporal y fusiforme) se corresponden a lo aparecido en la literatura.

Existe una cierta tendencia a la elección de características del hemisferio izquierdo. Esta tendencia también está recogida en la literatura. El mismo Korolev hace notar este hecho al analizar sus resultados [17].

Capítulo 6

Conclusiones

6.1. Conclusión

En este estudio se ha demostrado que un acercamiento longitudinal posee mayor poder de diagnóstico que un acercamiento transversal. Además, la combinación de los dos métodos mRMR y LME envueltos en una estructura de K-Folds ofrece una gran capacidad de clasificación y generalización.

Por otra parte, también ha quedado patente la mejora de los resultados según se cuenta con más visitas del mismo paciente.

Dentro de las características utilizadas, han surgido varios conceptos interesantes:

- Que la mayor parte del poder de predicción lo aportan los tests neuropsicológicos. En la mayoría de casos se trata de FAQ, ADAS13 y ADASQ4.
- Que existe una tendencia a escoger características asociadas con el hemisferio izquierdo del cerebro.
- Que los clusters provenientes del análisis masivo sustituyen a las ROIs si se consideran ambas familias de características.
- Que una de las ROIs con mayor índice de aparición es el volumen normalizado de la amígdala.
- Que la dimensión óptima se encuentra en torno a $D=4$. Esto quiere decir que por lo general los mejores resultados se obtendrán con grupos de 3, 4 ó 5 características.

Por otro lado, se ha profundizado en el funcionamiento de FreeSurfer y se ha desarrollado un pequeño flujo de trabajo donde se solucionan problemas comunes que se han experimentado a lo largo del proyecto.

Se han proporcionado todos los códigos utilizados para la realización de este estudio: tanto la parte de extracción de datos como la parte de análisis y experimentos.

Además, de manera indirecta son diversos los conocimientos que se han adquirido a propósito de este trabajo. Estos conocimientos incluyen un manejo más fluido de MATLAB, un acercamiento a distintos conceptos del aprendizaje automático, conocimientos de anatomía y neurociencia y, finalmente, la adquisición de soltura manejando el lenguaje LaTeX, indispensable a la hora de escribir documentos con un estilo cerrado.

6.2. Desarrollos futuros

De entre todas las actividades y modificaciones que pueden llevarse a cabo para mejorar el presente estudio cabe destacar las siguientes:

- Aumentar el tamaño de la muestra para poder proporcionar nuevos acercamientos al problema. Por ejemplo, con una muestra suficientemente grande se puede plantear una medida del error más fina con un modelo K-Folds, o bien estudiar modelos de entrenamiento y test distintos.
- Ampliar, mejorar y seguir documentando el código desarrollado. Como trabajo complementario pueden estudiarse acercamientos dedicados únicamente a la guía sobre el uso y manejo de las herramientas utilizadas, especialmente en el caso de ADNI y FreeSurfer. Esta propuesta se debe a la sensación de torpeza y lentitud que se experimenta la primera vez que se entra en contacto con estas herramientas.

Aunque en este estudio se ha intentado esclarecer ciertos puntos al respecto, todo material complementario puede ayudar a reducir drásticamente el tiempo de adaptación, permitiendo así a estudios limitados en el tiempo llegar más lejos en la investigación.

- La introducción del cálculo de la rugosidad del hipocampo y otras medidas adicionales puede ser una futura vía para tratar de mejorar los resultados. Tal y como se ha visto en varios estudios, existe la posibilidad de añadir a las características alguna medida más específica de ciertas zonas de interés, como podría ser el hipocampo o la corteza entorrinal. En cualquier caso, después habría que verificar si mejoran la predicción sustancialmente o no.
- Tratar de buscar un lenguaje de programación más adecuado para el análisis estadístico, como podría ser el caso de R. Si bien es cierto que MATLAB se trata de un lenguaje muy extendido por su versatilidad, no debe menospreciarse el uso de otros lenguajes especializados.

R es un lenguaje de alto nivel con una estructura parecida a MATLAB, especialmente diseñado para el análisis de datos y que ofrece mejores herramientas de representación de resultados. En este contexto puede plantearse trasladar las funciones del modelo LME de Bernal a este otro lenguaje. Sólo esta migración podría considerarse un trabajo de final de grado por sí misma.

Bibliografía

- [1] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.
- [2] Jorge L Bernal-Rusiel, Douglas N Greve, Martin Reuter, Bruce Fischl, Mert R Sabuncu, Alzheimer’s Disease Neuroimaging Initiative, et al. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*, 66:249–260, 2013.
- [3] Jorge L Bernal-Rusiel, Martin Reuter, Douglas N Greve, Bruce Fischl, Mert R Sabuncu, Alzheimer’s Disease Neuroimaging Initiative, et al. Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage*, 81:358–370, 2013.
- [4] Kyle Brauer Boone, Po Lu, and Johnny Wen. Comparison of various ravlt scores in the detection of noncredible memory performance. *Archives of Clinical Neuropsychology*, 20(3):301–319, 2005.
- [5] Terry L Brink, JA Yesavage, O Lum, et al. Geriatric depression scale. *Evidence-Based Diagnosis A Handb. Clin. Predict. Rules*, page 297, 2013.
- [6] OPAS Demencia. Una prioridad de salud pública. *Washington, DC: Organización Panamericana de la Salud, World Health Organization*, 112, 2013.
- [7] David M Diez, Christopher D Barr, and Mine Cetinkaya-Rundel. *OpenIntro statistics*, volume 12. CreateSpace, 2012.
- [8] Chris Ding and Hanchuan Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02):185–205, 2005.
- [9] John Duncan and Adrian M Owen. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in neurosciences*, 23(10):475–483, 2000.
- [10] Simon F Eskildsen, Pierrick Coupé, Vladimir S Fonov, Jens C Pruessner, and D Louis Collins. Structural imaging biomarkers of alzheimer’s disease: predicting disease progression. *Neurobiology of aging*, 36:S23–S31, 2015.
- [11] Bruce Fischl and Anders M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055, 2000.

- [12] Consensus Recommendations for the Post-Mortem Diagnosis of Alzheimer's Disease. The national institute on aging, and reagan institute working group on diagnostic criteria for the neuropathological assessment of alzheimer's disease. *Neurobiol Aging*, 18:1–S2, 1997.
- [13] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, 2002.
- [14] Bradley T Hyman and John Q Trojanowski. Editorial on consensus recommendations for the postmortem diagnosis of alzheimer disease from the national institute on aging and the reagan institute working group on diagnostic criteria for the neuropathological assessment of alzheimer disease. *Journal of neuropathology and experimental neurology*, 56(10):1095, 1997.
- [15] Ryo Ishibashi, Matthew A Lambon Ralph, Satoru Saito, and Gorana Pobric. Different roles of lateral anterior temporal lobe and inferior parietal lobule in coding function and manipulation tool knowledge: evidence from an rtms study. *Neuropsychologia*, 49(5):1128–1135, 2011.
- [16] Zaven S Khachaturian. Diagnosis of alzheimer's disease. *Archives of neurology*, 42(11):1097–1105, 1985.
- [17] Igor O Korolev, Laura L Symonds, Andrea C Bozoki, Alzheimer's Disease Neuroimaging Initiative, et al. Predicting progression from mild cognitive impairment to alzheimer's dementia using clinical, mri, and plasma biomarkers via probabilistic pattern classification. *PloS one*, 11(2):e0138866, 2016.
- [18] Ramon Landin-Romero, Fiona Kumfor, Cristian E Leyton, Muireann Irish, John R Hodges, and Olivier Piguet. Disease-specific patterns of cortical and subcortical degeneration in a longitudinal study of alzheimer's disease and behavioural-variant frontotemporal dementia. *Neuroimage*, 151:72–80, 2017.
- [19] Yang Li, Yaping Wang, Guorong Wu, Feng Shi, Luping Zhou, Weili Lin, and Dinggang Shen. Discriminant analysis of longitudinal cortical thickness changes in alzheimer's disease using dynamic and network features. *Neurobiology of aging*, 33(2):427–e15, 2012.
- [20] Lin Lin. Análisis longitudinal desde imágenes de resonancia magnética para la detección temprana de la enfermedad del alzheimer. 2017.
- [21] John H McDonald. *Handbook of biological statistics*, volume 2. Sparky House Publishing Baltimore, MD, 2009.
- [22] Ian McDowell. *Measuring health: a guide to rating scales and questionnaires*. Oxford University Press, USA, 2006.
- [23] Guy McKhann, David Drachman, Marshall Folstein, Robert Katzman, Donald Price, and Emanuel M Stadlan. Clinical diagnosis of alzheimer's disease report of the nincds-adrda work group* under the auspices of department of health and human services task force on alzheimer's disease. *Neurology*, 34(7):939–939, 1984.

- [24] Samuel Barreto Melo. Transformaciones geométricas sobre imágenes digitales. *Facultad de Ciencias-Carrera de Matemáticas. Universidad Distrital Francisco José de Caldas*, 2009.
- [25] Alex J Mitchell and Mojtaba Shiri-Feshki. Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies. *Acta Psychiatrica Scandinavica*, 119(4):252–265, 2009.
- [26] Daniel Peña. *Fundamentos de estadística*. Alianza Editorial, 2014.
- [27] Ronald C Petersen. Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine*, 256(3):183–194, 2004.
- [28] Ronald C Petersen, Rosebud O Roberts, David S Knopman, Bradley F Boeve, Yonas E Geda, Robert J Ivnik, Glenn E Smith, and Clifford R Jack. Mild cognitive impairment: ten years later. *Archives of neurology*, 66(12):1447–1455, 2009.
- [29] Martin James Prince. *World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer’s Disease International, 2015.
- [30] Olivier Querbes, Florent Aubry, Jérémie Pariente, Jean-Albert Lotterie, Jean-François Démonet, Véronique Duret, Michèle Puel, Isabelle Berry, Jean-Claude Fort, Pierre Celsis, et al. Early diagnosis of alzheimer’s disease using cortical thickness: impact of cognitive reserve. *Brain*, 132(8):2036–2047, 2009.
- [31] Saima Rathore, Mohamad Habes, Muhammad Aksam Iftikhar, Amanda Shacklett, and Christos Davatzikos. A review on neuroimaging-based classification studies and associated feature extraction methods for alzheimer’s disease and its prodromal stages. *NeuroImage*, 155:530–548, 2017.
- [32] E García Río. Una introducción a la curvatura. *Universidade de Santiago de Compostela*, 2002.
- [33] Nelson Rivera, Vivian Andrea Perdomo, Guillermo Barreto, et al. Frecuencias polimórficas del gen de apolipoproteína e en el valle del cauca, colombia. *Revista de Ciencias*, 19(1):10–10, 2018.
- [34] Katharina Schmidt. Clinical dementia rating scale. In *Encyclopedia of Quality of Life and Well-Being Research*, pages 957–960. Springer, 2014.
- [35] Joel Andrés Silva Sánchez. Determinación de la frecuencia de los polimorfismos de la apolipoproteína e (apoe) en la población de la sierra ecuatoriana. B.S. thesis, Quito: Universidad de las Américas, 2017., 2017.
- [36] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.
- [37] Kyle Smith, Paul Kang, and Marwan N Sabbagh. The effect of statins on rate of cognitive decline in mild cognitive impairment. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association*, 12(7):P619, 2016.

- [38] Lauge Sørensen, Christian Igel, Akshay Pai, Ioana Balas, Cecilie Anker, Martin Lillholm, Mads Nielsen, Alzheimer's Disease Neuroimaging Initiative, et al. Differential diagnosis of mild cognitive impairment and alzheimer's disease using structural mri cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical*, 13:470–482, 2017.
- [39] Mary C Tierney, Rory H Fisher, Anthony J Lewis, Maria L Zorzitto, W Gary Snow, David W Reid, and Paula Nieuwstraten. The nincds-adrda work group criteria for the clinical diagnosis of probable alzheimer's disease a clinicopathologic study of 57 cases. *Neurology*, 38(3):359–359, 1988.

Anexos

Anexo A

Conceptos básicos de estadística

Con el fin de asegurarse de que el lector entienda de manera inequívoca los distintos conceptos estadísticos que aparecerán en este trabajo (especialmente en la parte de conclusiones) se procede en este apartado a definir algunos de los términos más importantes. De todo lo aquí expuesto se pueden encontrar más referencias en los libros *Handbook of biological statistics* [21], *OpenIntro Statistics*[7] o *Fundamentos de Estadística*, de Daniel Peña [26].

A.1. Introducción al contraste de hipótesis

A.1.1. Hipótesis nula e hipótesis alternativa

El contraste de hipótesis es un método estadístico que permite aceptar o rechazar una afirmación o hipótesis con un cierto grado de incertidumbre. Esta hipótesis cuya veracidad se desea demostrar recibe el nombre de **hipótesis nula**, y se suele denominar con H_0 . A la hipótesis contraria se denomina **hipótesis alternativa** H_1 .

Por lo general, la hipótesis nula suele asociarse a la falta de cambio. Son hipótesis del estilo: las dos poblaciones son iguales, la media de las dos medidas es la misma, etc. La hipótesis alternativa recoge, precisamente, la posibilidad de cambio: las poblaciones son distintas, la media de las medidas es distinta, etc.

A.1.2. Test de hipótesis

Una vez se han planteado las dos hipótesis debe decidirse si se hará un test unilateral o bilateral. Un test unilateral mide únicamente una diferencia positiva o negativa entre las medidas estudiadas. Se trata de hipótesis caracterizadas por los signos $<$ y $>$. Por otro lado existen los tests bilaterales, donde se estudia la diferencia en cualquiera de los dos sentidos. Estas hipótesis suelen estar caracterizadas por los signos $=$ (H_0) y \neq (H_1).

En este contexto entra en juego el p-valor. El p-valor representa la probabilidad de obtener los resultados resultantes si la hipótesis nula fuese cierta. Por ejemplo, si se espera que la media de dos medidas fuera la misma (H_0) y se ha medido una diferencia D entre las dos medias: ¿Qué probabilidad hay de que esto ocurra siendo H_0 cierta?

A este valor debe unírsele otra definición. El nivel de significación α representa el valor mínimo de seguridad más allá del cual se rechazará la hipótesis nula, y es definido por el usuario antes del estudio (véase figura A.1). En estudios de biología se suele usar $\alpha = 0,05$. Sin embargo, en el análisis masivo se verá que se utilizará un $\alpha = 0,01$, ya que esto permite obtener las zonas más significativas. Es decir, aquellas en las que se puede decir con mayor confianza (99 %) que existe una diferencia entre los dos grupos.

Se tiene por lo tanto que α puede entenderse como una suerte de frontera. Si el p-valor es demasiado bajo entonces la hipótesis nula no se sostiene (para ese nivel de significancia) y debe rechazarse. Esto, tal y como se ha visto antes, suele significar que las medidas se entienden como distintas, o que los dos grupos son diferentes.

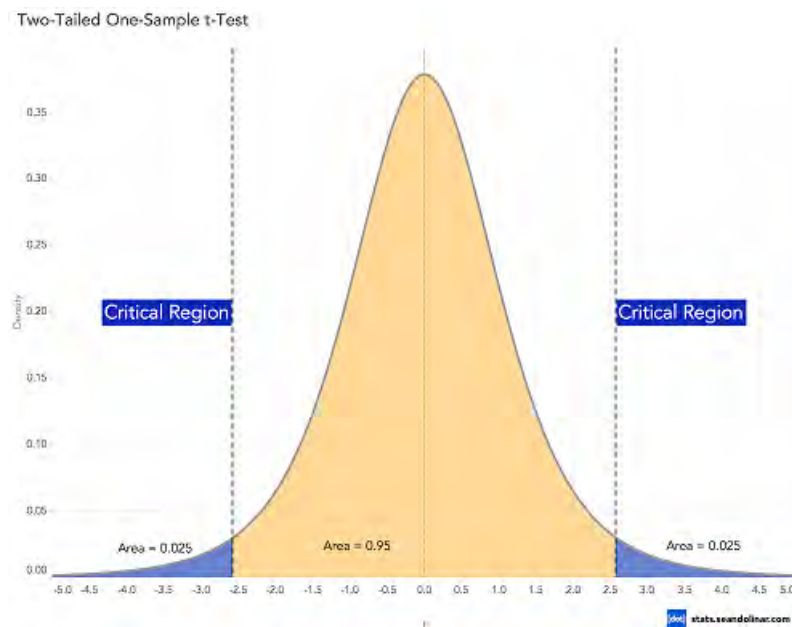


Figura A.1: Esquema de la distribución de la hipótesis nula así como los valores límite para un nivel de significación de $\alpha = 0,05$, por Sean Dolinar.

Una vez se ha llegado a este punto cabe preguntarse: ¿Cómo se calcula entonces el p-valor? Lo primero que se plantea es una distribución para la hipótesis nula. Una vez asumida una distribución se utiliza la función de probabilidad para calcular la probabilidad (p-valor) en el punto dado (medido). En el caso de la figura A.1 se trata de una distribución T de Student. Al realizar un T-test sobre una serie de medidas se obtendrá la probabilidad según esta distribución. Las distribuciones más comunes y utilizadas para modelar estas distribuciones son la T de Student (T-test), la F de Fisher (F-test), la distribución normal (Z-test) y

A.1.3. Tipos de errores

Una vez se ha introducido la manera de aceptar o rechazar la hipótesis nula es importante comprender los tipos de errores. Existen dos tipos principales dependiendo de la incongruencia :

1. **Errores de tipo I:** se habla de errores de tipo 1 cuando se rechaza la hipótesis nula cuando en realidad era válida.
2. **Errores de tipo II:** los errores de tipo 2 recogen los casos en los que se ha aceptado la hipótesis nula cuando en verdad era falsa.

Todos los escenarios posibles, tanto aciertos como errores, se muestran en la llamada matriz de confusión:

	H_0 aceptada	H_0 rechazada
H_0 verdadera	Verdadero positivo	Error tipo I
H_0 falsa	Error tipo II	Verdadero negativo

Tabla A.1: Matriz de confusión. Las filas representan el verdadero valor de la hipótesis, mientras que cada columna representa la predicción a partir del test.

Aparte de estos errores, también disponemos de los verdaderos positivos (VP) y de verdaderos negativos (VN). Estos dos casos se dan cuando tanto la predicción como el resultado real coinciden. El objetivo, por supuesto, es maximizar estos valores. No obstante, para medir los errores se suelen utilizar una serie de valores que ofrecen relaciones con un significado práctico entre ellos:

- Exactitud (accuracy, ACC): define la proporción de elementos que han sido correctamente identificados respecto al total. Puede expresarse como:

$$\frac{VP + VN}{P + N}$$

- Sensibilidad (Sensitivity): define la proporción de positivos que han sido correctamente identificados. Se puede expresar como:

$$\frac{VP}{P}$$

- Especificidad (Specificity): define la proporción de negativos que han sido correctamente identificados. se puede expresar como:

$$\frac{VN}{N}$$

- Área bajo la curva (AUC): se construye a partir de, precisamente, la sensibilidad y la especificidad. Aporta una medida de cómo de bueno es el resultado. La curva referida se trata de una curva en el llamado espacio ROC. Cada punto de la curva se define modificando el punto de corte para la diferenciación entre dos clases.

Esto cambia la proporción de elementos correctamente predichos, construyendo de la siguiente manera la curva: En el eje x encontramos $(1 - \text{sensibilidad})$ y en el eje y encontramos la *especificidad* (véase figura A.2).

Los distintos casos corresponden a distintos puntos de corte. El caso ideal se encontraría en un valor de $(1 - \text{sensibilidad}) = 0$ (o lo que es lo mismo, una sensibilidad máxima) y *especificidad* = 1. Si la curva se asemeja a la diagonal eso significa que el método no ofrece ventaja real ante lanzar una moneda al aire.

Por último es necesario tener en cuenta que la precisión de estas curvas depende del tamaño del grupo de test. Si la N de este grupo es grande, se asemejarán a una curva continua. Sin embargo, si la población de test no cuenta con muchos sujetos entonces el resultado será más parecido a la figura A.3.

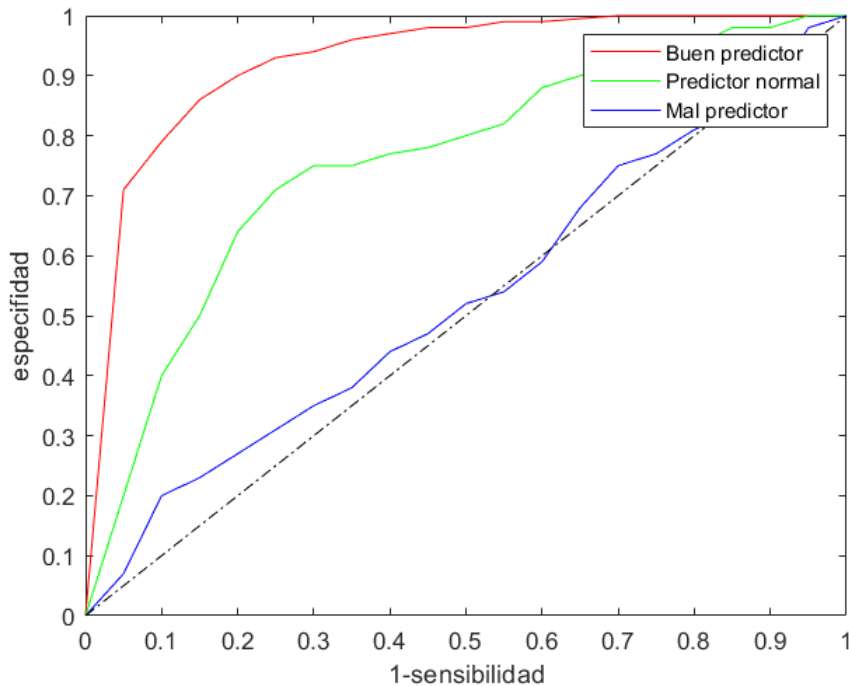


Figura A.2: Ejemplo conceptual de tres curvas ROC y su significado: En rojo, un buen predictor. En verde, un predictor normal. En azul, un mal predictor que apenas mejoraría el resultado de lanzar una moneda al aire.

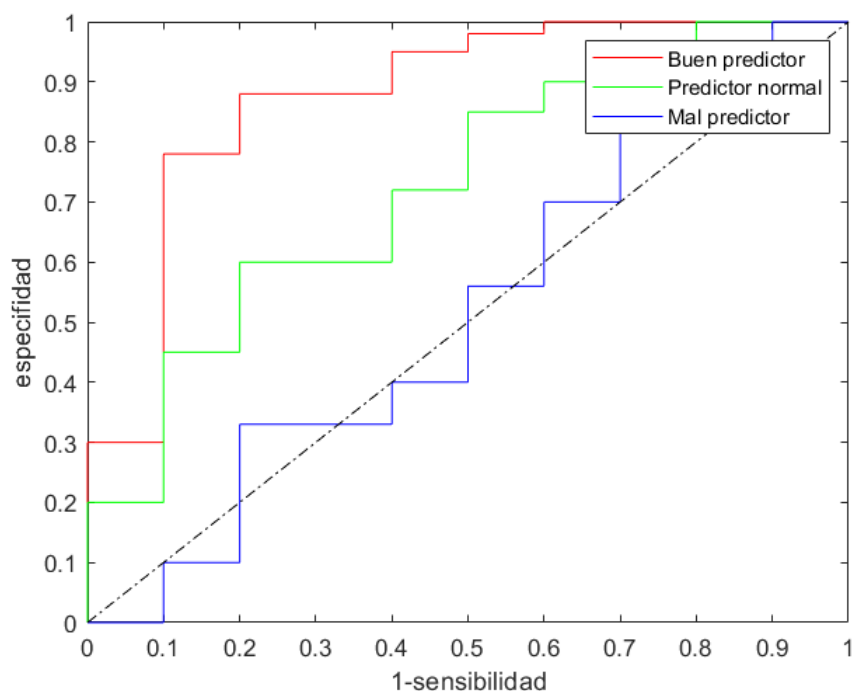


Figura A.3: Con una población de test más pequeña las curvas ROC se vuelven más poligonales.

A.2. La comparativa múltiple: FDR y q-valores

En este apartado se explica cómo pueden ampliarse los conceptos relacionados con el p-valor y el contraste de hipótesis para llevar a cabo un análisis de múltiples comparativas. En este estudio estos conceptos han sido aplicados al análisis masivo.

Para entender el problema de múltiples comparativas se debe recordar el sentido del nivel de significancia α . Un α de 0.05 significa que hay un 5% de probabilidad de caer en un error de tipo 1. Mientras que esta medida puede estar bien para un solo test, puede resultar inaceptable si se repite varias veces.

Como en cada test esta probabilidad se mantiene, si se ejecutan suficientes tests también aumentará el número de falsos positivos ($0,05 \times 100 = 5$, $0,05 \times 1000 = 50$).

FDR proviene de las siglas en inglés de *False Discovery Rate*. Cambia el paradigma estableciendo que el p-valor se refiera al conjunto de los tests. A este p-valor ajustado se le llama q-valor.

En el caso del análisis masivo, se están realizando tests para cada uno de los 160.000 vértices de la superficie. En este contexto se realizan varios subconjuntos

de la muestra (póngase 8 subconjuntos distintos). Por lo tanto, para cada vértice se obtendrán 8 p-valores.

La primera aproximación para trabajar con el FDR fue descrita por Benjamini y Hochberg en 1995 [1]. Acorde a su publicación, si se desea controlar que en un estudio con N comparaciones el FDR no supere un porcentaje D :

1. Deben ordenarse los p-valores de los N tests de menor a mayor. Este resultado se puede entender mejor representando los valores en forma de histograma (véase figura A.4
2. Se define K como la última posición para la que se cumple que $p_i \leq D \times \frac{i}{N}$, siendo i el índice dentro de la lista de p-valores.
3. Se consideran válidos todos los p-valores hasta la posición K .

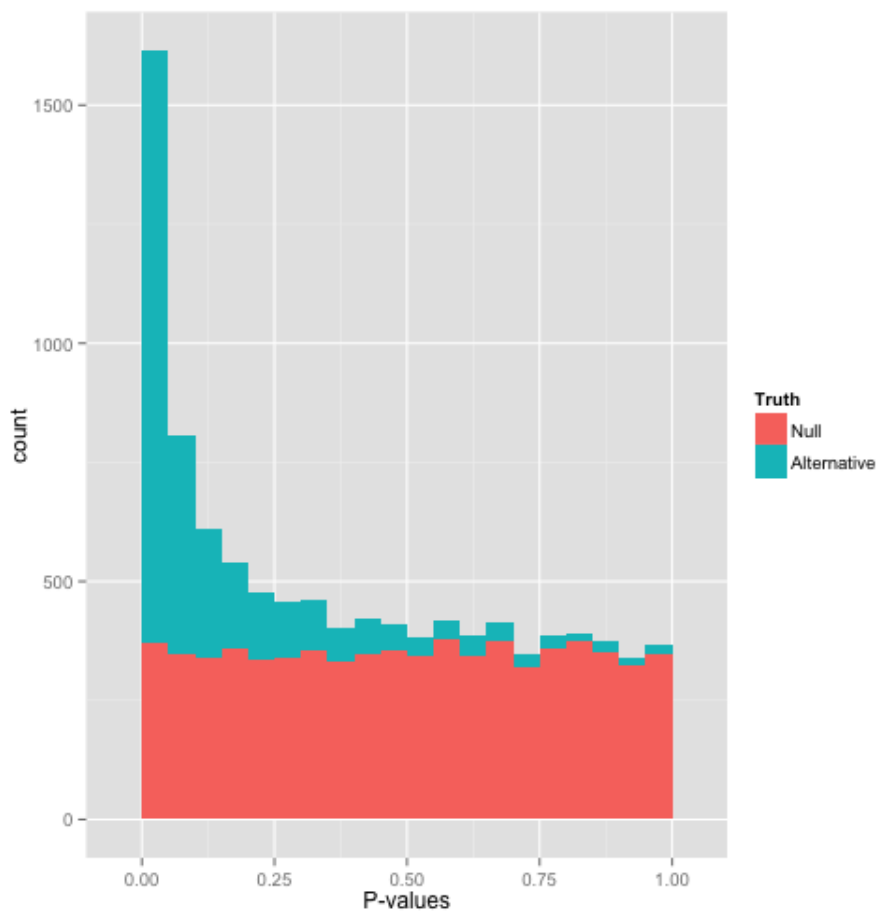


Figura A.4: Esquema del histograma de p-valores para comparativas múltiples. En un caso donde realmente la hipótesis nula sea falsa, la mayoría de casos darán un p-valor por debajo del nivel de significancia. Si la hipótesis nula es cierta, la distribución será más o menos homogénea a lo largo de todo el espectro.

Una ventaja del FDR es su fácil interpretación. Se trata de la proporción esperada de falsos positivos entre todos los tests considerados como significativos. Este concepto cobrará especial importancia en la introducción de elementos masivos.

Ante la hipótesis nula de que los dos conjuntos de córtex son iguales, escogeremos las zonas donde el p-valor indique que debemos descartar la hipótesis nula, ya que esto quiere decir que son zonas que presentan diferencias significativas entre los dos grupos.

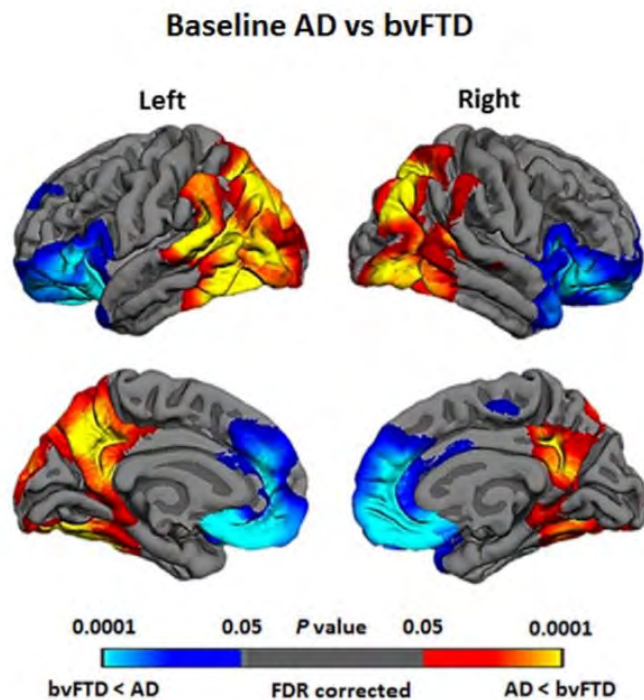


Figura A.5: Ejemplo de representación de p-valores corregidos procedente del estudio de Landin2017 [18] para el análisis masivo del córtex.

A.3. Tipos de grupos en algoritmos de aprendizaje

A la hora de abordar un problema de *machine learning* es necesario contar con tres poblaciones distintas:

- El grupo de entrenamiento (o training set): es un conjunto de elementos con un vector de características X y una salida Y . El objetivo es entrenar al algoritmo con ejemplos sobre qué parámetros X dan lugar a la salida Y .

En el caso de este estudio, la salida es el diagnóstico. El uso de un sistema longitudinal de efectos mixtos (LME) hace que no se introduzca exactamente un vector de características X en este caso. Para más información, acúdase al correspondiente apartado de Métodos.

- El grupo de validación (o validation set): es un conjunto de elementos de la misma naturaleza que el grupo de entrenamiento. En este caso, se utilizan para comprobar el poder de generalización del algoritmo. Como norma general, la precisión en este grupo será menor que en el grupo de entrenamiento. Sin embargo, si la diferencia es muy acusada, podemos estar ante un problema común en este tipo de algoritmos: el llamado *overfitting*. Consiste esencialmente en un bajo poder de generalización, normalmente debido a una muestra pequeña de entrenamiento o un exceso de características en el vector X . En esos casos el algoritmo se ciñe al grupo de entrenamiento, haciendo muy buenas predicciones. Sin embargo, al introducir un nuevo caso no es capaz de realizar una predicción correcta.

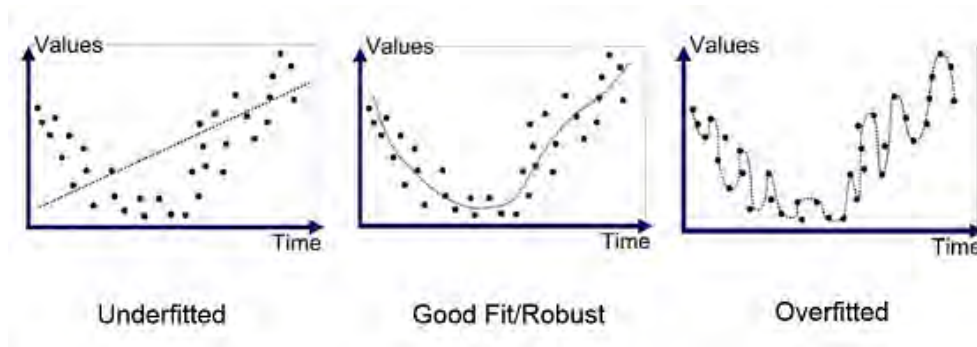


Figura A.6: Esquema de la problemática del overfitting. Si la población es muy baja o el vector de características es muy extenso la curva no se adapta bien a nuevos casos.

- El grupo de test (o test set): es un conjunto de elementos sobre los que se va a medir el error. Es importante no confundirlo con el grupo de validación. La función de este último es la de comprobar el poder de generalización, pero siempre se estará ajustando esa generalización al grupo de validación.

La función del grupo de test es la de, una vez hallados los parámetros que obtienen la salida Y a partir del vector X , ejecutarlo sobre sujetos con características X y mostrar el error producido. Es este resultado el que aparece publicado en el documento final, como prueba de la eficacia del algoritmo.

Respecto al porcentaje que debe constituir cada grupo respecto a la muestra total de la que disponemos, una decisión sensata suele ser 60 % de sujetos de entrenamiento, 20 % de sujetos de validación y 20 % de sujetos de test. Sin embargo, hay otras combinaciones de porcentaje dependiendo de la naturaleza del estudio.

Por último, es importante que estas tres poblaciones no intersecten. Es cierto que en el caso de la validación se puede recurrir a métodos como el Leave-One-Out cross

validation (LOO-cv). Este método es utilizado sobre todo en poblaciones pequeñas. Consiste en dejar un elemento del grupo de entrenamiento fuera y después usarlo como validación. Tras esto se repite el proceso con todos los elementos del conjunto de entrenamiento para obtener un error estadístico. Sin embargo, siempre que se disponga de suficientes sujetos es recomendable dividir los grupos.

Con este fin surge el modelo de K-Folds y su variante de K-Folds anidados utilizada en este estudio. El concepto es parecido al de LOO-cv, solo que en este caso se realizan K divisiones de la población. Con K-1 muestras se entrena y valida el algoritmo de aprendizaje, mientras que la última se utiliza como población de test. Más información sobre este método y su variante se puede encontrar en la sección correspondiente del capítulo de Métodos.

Anexo B

Adquisición y procesamiento de imágenes MRI

B.1. Imágenes de resonancia magnética

Aunque este estudio parte de las imágenes de resonancia magnética (MRI), es importante conocer el proceso que hay detrás, para entender de esta manera la información procurada.

Para adquirir estas imágenes se somete al cuerpo humano a un campo magnético fuerte. En este estado se envían ondas de radio y se reciben las señales emitidas por los núcleos atómicos (véase figura B.1). Un ordenador codifica esta información en forma de intensidades.

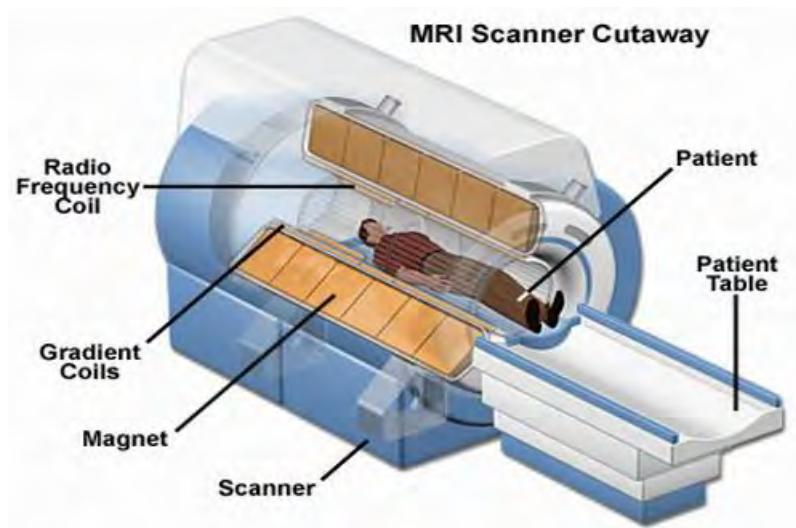


Figura B.1: Esquema del equipo utilizado en la adquisición de imágenes de resonancia magnética.

Entre las ventajas de esta técnica se encuentra que:

1. No utiliza radiación ionizante, por lo que no causa daño alguno al paciente.
2. Se pueden adquirir imágenes en varios planos sin mover al paciente de posición (véase figura B.2).

- Ofrece un buen contraste de tejidos. Además, dependiendo del estudio, esta característica se puede mejorar añadiendo agentes de contraste.

Una desventaja la encontramos en el tiempo que necesita. Normalmente la prueba dura entre 30 y 60 minutos, mientras que otras técnicas toman entre 5 y 15 minutos.

Por último, cabe destacar que al aumentar el campo magnético se puede aumentar la resolución obtenida en la imagen. Sin embargo, el equipo se encarece. Es por este motivo que los equipos más utilizados son los de 1.5T, razón por la cual este estudio se basa en este tipo de imágenes MRI-T1.

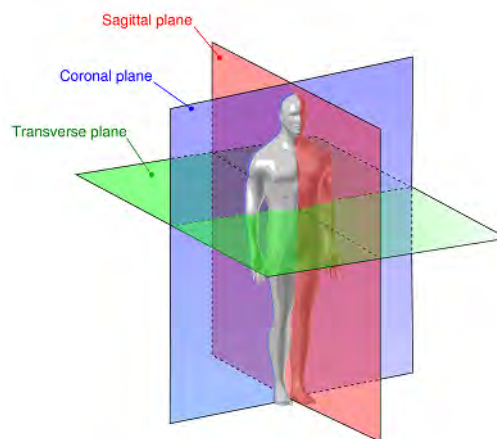


Figura B.2: Los tres planos principales utilizados en la adquisición de imágenes MRI, expresados en relación al cuerpo del paciente.

B.2. Procesamiento de imágenes

Sobre el procesamiento de imágenes, la principal fuente de información ha sido la propia página web de Freesurfer. Sin embargo, un primer acercamiento se puede encontrar en el estudio de Fischl [11]. En este artículo se introduce brevemente la importancia que tiene este biomarcador, tal y como hemos visto ya en la sección dedicada a la evolución de la enfermedad.

Sin embargo, lo más importante del artículo de Fischl (y el tema central del mismo) se encuentra en su descripción del método de procesamiento y la introducción a problemas que debe solventar cualquier método utilizado. Esto permitirá entender mejor los pasos seguidos en el procedimiento de ADNI.

Se muestra que la escasa resolución de MRI hace complicado el cálculo directo de la superficie pial. Por ello se calcula a partir de las frontera de materia gris y de materia blanca. El primer paso es clasificar los vóxeles indicando si forman parte de la materia blanca.

Seguidamente se nos muestra la restricción que se hace en varios algoritmos de que la superficie cerebral sea suave (sin picos) en la escala de milímetros en la que se trabaja. La manera habitual, tal y como se nos describe, implica minimizar la

curvatura Gaussiana o la curvatura media.

La curvatura gaussiana de una superficie es un número real k que se describe a partir de dos números llamados curvaturas principales [32]. Estas curvaturas principales se obtienen de la siguiente manera:

1. Se tiene que S es una superficie contenida en \mathbb{R}^3 , que p es un punto de esta superficie, y que N es un vector unitario y normal a S en p .
2. Se elige un plano Π que pase por p y que contenga a N . La intersección entre Π y S es una curva plana $\gamma \subset \Pi$ que pasa por p .
3. Se calcula la curvatura con signo (k_N) de γ y se repite el proceso para todos los planos normales Π . Las curvaturas principales, K_1 y K_2 , se definen como el mínimo y el máximo de todas estas curvaturas.

Una vez obtenidas K_1 y K_2 , se define la curvatura gaussiana como:

$$K = K_1 K_2$$

Por otra parte, se define la curvatura media gaussiana como:

$$H = \frac{1}{2}(K_1 + K_2)$$

Sin embargo, aunque la superficie general es suave, existen zonas del cerebro con un alto índice de rugosidad. Se observa este fenómeno en el fondo de los surcos y la cima de los giros cerebrales (Véase figura B.3).

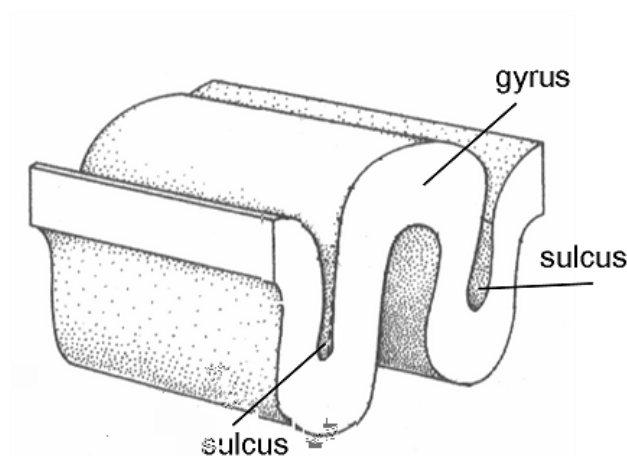


Figura B.3: Representación de los giros cerebrales (colinas) y los surcos (valles) donde el suavizado provoca más error.

El recurrir a la restricción de curvatura supone introducir un error considerable en estas zonas. Para corregirlo se calcula la curvatura de parcelas de superficie. Tras esto se altera la representación de la superficie para que la superficie local tenga también la escala más apropiada posible (i.e.: milímetros).

Tras esto el algoritmo se asegura de que la malla que representa la superficie cortical no está intersectando consigo misma. Para ello se recurre a un algoritmo de intersección de triángulos que aumenta su coste computacional linealmente respecto al número de vértices. Este algoritmo, según el propio Fischl, será el causante de gran parte del coste computacional del proceso completo de suavizado y corrección de errores.

Respecto al procesamiento orientado a la comparación para el análisis masivo, no se incluye en esta versión del trabajo ya que no se ha podido llevar a cabo todavía la parte correspondiente a la introducción de características provenientes del análisis masivo.

Anexo C

Diccionario de siglas

Durante todo el trabajo se han sustituido algunos conceptos por su acrónimo. Si bien cada término aparece explicado en su primera aparición, se incluye en este anexo una correspondencia entre cada sigla y su significado. Para conceptos relacionados con secciones del cerebro se recuerda que variantes con R o L se refieren al mismo concepto aplicado a uno de los hemisferios del cerebro (derecho o izquierdo, respectivamente).

C.1. Ordenación alfabética

- **ACC:** Precisión (*Accuracy*).
- **AD:** Enfermedad de Alzheimer (*Alzheimer Disease*).
- **ADAS-Cog:** *Alzheimer's Disease Assessment Scale-Cognitive sub-scale*.
- **ADNI:** Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (*Alzheimer's Disease Neuroimaging Initiative*).
- **APARC:** Parcelación Automática (superficie-cortical) (*Automatic Parcelation*).
- **APOE4:** Apolipoproteína E, alelo 4.
- **ASEG:** Segmentación Automática (Volumen-subcortical) (*Automatic Segmentation*).
- **AUC:** Área Bajo la Curva (*Area Under the Curve*).
- **CDR:** *Clinical Dementia Rating scale*.
- **ECT:** Espesor de la Corteza Entorrinal (*Entorhinal Cortical Thickness*).
- **FAQ:** *Functional Activities Questionnaire*.
- **FDR:** (*False Discovery Rate*).
- **GDS:** *Geriatric Depression Scale*.
- **IPT:** Espesor del lóbulo Parietal Inferior (*Inferior Parietal Thickness*).
- **ITT:** Espesor del lóbulo Temporal Inferior (*Inferior Temporal Thickness*).
- **KFSF:** modelo K-Fold con Selección por Frecuencia.

- **LME**: Modelo Longitudinal de Efectos Mixtos(*Longitudinal Mixed Effects model*).
- **MCI**: Deterioro Cognitivo Leve (*Mild Cognitive Impairment*).
- **MeT**: Espesor Medio(*Mean Thickness*).
- **MMSE**: *Mini-Mental State Examination*.
- **MRI**: Imagen de Resonancia Magnética (*Magnetic Resonance Image*).
- **MRI-1.5T**: Imagen de Resonancia Magnética de 1.5 Teslas.
- **mRMR**: Algoritmo de Mínima Redundancia y Máxima Relevancia(*Minimum Redundancy and Maximum Relevance*).
- **MTT**: Espesor del lóbulo Temporal Medio(*Middle Temporal Thickness*).
- **NAV**: Volumen de la Amígdala Normalizado (*Normalized Amygdala Volume*).
- **NCV**: Volumen del núcleo Caudado Normalizado (*Normalized Caudate Volume*).
- **NHV**: Volumen Hipocampal Normalizado (*Normalized Hippocampal Volume*).
- **NPaV**: Volumen del Pallidum Normalizado (*Normalized Pallidum Volume*).
- **NPV**: Volumen del Putamen Normalizado (*Normalized Putamen Volume*).
- **NPVal**: Valor Profético Negativo(*Negative Predictive Value*).
- **PCT**: Espesor del giro Cingulado Posterior(*Posterior Cingulate Thickness*).
- **pMCI**: Deterioro Cognitivo Leve progresivo (*progressive Mild Cognitive Impairment*).
- **PPVal**: Valor Profético Positivo(*Positive Predictive Value*).
- **PRT**: Espesor del Precúneo(*Precuneus Thickness*).
- **RAVLT**: *Ray Auditory Verbal Learning Test*.
- **ROIs**: Regiones de Interés (*Regions of Interest*).
- **sMCI**: Deterioro Cognitivo Leve estable (*stable Mild Cognitive Impairment*).
- **SPT**: Espesor del Parietal Superior(*Superior Parietal Thickness*).
- **TPT**: Espesor del Polo Temporal(*Temporal Pole Thickness*).

C.2. Ordenación temática

C.2.1. Diagnóstico

- **AD:** Enfermedad de Alzheimer (*Alzheimer Disease*).
- **MCI:** Deterioro Cognitivo Leve (*Mild Cognitive Impairment*).
- **pMCI:** Deterioro Cognitivo Leve progresivo (*progressive Mild Cognitive Impairment*).
- **sMCI:** Deterioro Cognitivo Leve estable (*stable Mild Cognitive Impairment*).

C.2.2. Tests neuropsicológicos

- **ADAS-Cog:** *Alzheimer's Disease Assessment Scale-Cognitive sub-scale*.
- **CDR:** *Clinical Dementia Rating scale*.
- **FAQ:** *Functional Activities Questionnaire*.
- **GDS:** *Geriatric Depression Scale*.
- **MMSE:** *Mini-Mental State Examination*.
- **RAVLT:** *Rey Auditory Verbal Learning Test*.

C.2.3. Segmentación

- **APARC:** Parcelación Automática (superficie-cortical) (*Automatic Parcelation*).
- **ASEG:** Segmentación Automática (Volumen-subcortical) (*Automatic Segmentation*).
- **ROIs:** Regiones de Interés (*Regions of Interest*).

- **NHV:** Volumen Hipocampal Normalizado (*Normalized Hipocampal Volume*).
- **NAV:** Volumen de la Amígdala Normalizado (*Normalized Amygdala Volume*).
- **NPV:** Volumen del Putamen Normalizado (*Normalized Putamen Volume*).
- **NCV:** Volumen del núcleo Caudado Normalizado (*Normalized Caudete Volume*).
- **NPaV:** Volumen del Pallidum Normalizado (*Normalized Pallidum Volume*).

- **ECT:** Espesor de la Corteza Entorrinal (*Entorhinal Cortical Thickness*).
- **IPT:** Espesor del lóbulo Parietal Inferior (*Inferior Parietal Thickness*).
- **ITT:** Espesor del lóbulo Temporal Inferior (*Inferior Temporal Thickness*).
- **MeT:** Espesor Medio (*Mean Thickness*).
- **MTT:** Espesor del lóbulo Temporal Medio (*Middle Temporal Thickness*).

- **PCT**: Espesor del giro Cingulado Posterior(*Posterior Cingulate Thickness*).
- **PRT**: Espesor del Precúneo(*Precuneus Thickness*).
- **SPT**: Espesor del Parietal Superior(*Superior Parietal Thickness*).
- **TPT**: Espesor del Polo Temporal(*Temporal Pole Thickness*).

C.2.4. Medidas de error

- **ACC**: Precisión (*Accuracy*).
- **AUC**: Área Bajo la Curva (*Area Under the Curve*).
- **FDR**: (*False Discovery Rate*).
- **NPVal**: Valor Profético Negativo(*Negative Predictive Value*).
- **PPVal**: Valor Profético Positivo(*Positive Predictive Value*).

C.2.5. Miscelánea

- **ADNI**: Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (*Alzheimer's Disease Neuroimaging Initiative*).
- **APOE4**: Apolipoproteína E, alelo 4.
- **KFSF**: modelo K-Fold con Selección por Frecuencia.
- **LME**: Modelo Longitudinal de Efectos Mixtos(*Longitudinal Mixed Effects model*).
- **MRI**: Imagen de Resonancia Magnética (*Magnetic Resonance Image*).
- **MRI-1.5T**: Imagen de Resonancia Magnética de 1.5 Teslas.
- **mRMR**: Algoritmo de Mínima Redundancia y Máxima Relevancia(*Minimum Redundancy and Maximum Relevance*).