



ugr

Universidad
de Granada

TESIS DOCTORAL

PROGRAMA DE DOCTORADO EN TECNOLOGÍAS DE LA INFORMACIÓN Y
LA COMUNICACIÓN

Caracterización de imágenes cerebrales mediante distribuciones alfa-estables e isosuperficies Combinación de marcadores heterogéneos

Autor

Diego Castillo-Barnes

Tutor

Javier Ramírez Pérez de Inestrosa

Directores

Diego Salas-González

Javier Ramírez Pérez de Inestrosa



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, Mayo de 2021

Editor: Universidad de Granada. Tesis Doctorales
Autor: Diego Castillo Barnes
ISBN: 978-84-1117-015-4
URI: <http://hdl.handle.net/10481/70441>

Caracterización de imágenes cerebrales mediante distribuciones alfa-estables e isosuperficies Combinación de marcadores heterogéneos

Diego Castillo-Barnes

Este documento ha sido redactado en español con excepción del resumen y de las conclusiones de la tesis que han sido presentados también en inglés cumpliendo con lo marcado en la normativa vigente de la Escuela de Posgrado de la Universidad de Granada respecto a la obtención de la mención de Doctorado Internacional.

La normativa completa puede consultarse en el siguiente enlace:

<https://escuelaposgrado.ugr.es/pages/internacional/mencioninternacional/mencion>



Palabras clave: Neuroimagen, Enfermedad de Parkinson, Enfermedad de Alzheimer, Biomarcadores, Aprendizaje máquina, Clasificación, Distribuciones α -Estables, Resonancia Magnética, Tomografía Computarizada por Emisión de Fotón Único, Tomografía por Emisión de Positrones

Keywords: Neuroimaging, Parkinson's Disease, Alzheimer's Disease, Biomarkers, Machine learning, Classification, α -Magnetic Stable distributions, Resonance Imaging, Single Photon Emission Computed Tomography, Positron Emission Tomography

Agradecimientos

Esta tesis doctoral ha sido financiada por el contrato de Formación Personal Investigador (FPI) con referencia BES-2016-076546 asociado al proyecto de excelencia TEC2015-64718-R dentro del Programa Estatal de Investigación, Desarrollo e Innovación Orientada a los Retos de la Sociedad del MINECO/FEDER.

A mis padres, Diego y Encarni, y a mi hermano Pablo. Porque sin vuestros ánimos nunca hubiera llegado a alcanzar esta meta.

A Silvia, porque desde que me conoce nunca se ha separado de mi lado. Por su paciencia, por su inquietud, por su espíritu luchador.

A mis directores de tesis Diego Salas-Gonzalez y Javier Ramírez, por convertiros el mejor ejemplo humano a seguir. Por todo lo que siempre aprendo con vosotros. Por apostar por mí incluso cuando todavía no me conocíais.

A mis compañeros del grupo SiPBA: a Fermín Segovia, a Ignacio Álvarez, a Juan Manuel Górriz, a Andrés Ortiz, a Carmen Jiménez. Por todos los grandes momentos que comparto con vosotros, por todo lo que me enseñáis y por vuestra confianza en mí. Pero especialmente a Francisco Jesús Martínez-Murcia por ser, además, quien más me ha animado a luchar por mi sueño de ser investigador y profesor.

A todos los profesores que a lo largo de mi vida me han transmitido conocimientos, experiencia y valores: (Universidad de Sevilla) Carlos Crespo, Begoña Acha, Rafael Boloix, Mario Bilbao, Daniel Limón, Juan Manuel Virues, Juan José Murillo y Sergio Cruces; (Universidad de Las Palmas de Gran Canaria) Gustavo Callico, Javier García, Victor Manuel Melián y Miguel Ángel Ferrer.

A Ángel Hernández y Cayetano Cabrera, mis canarios.

A Juan Antonio Rivas, por tus consejos y tu apoyo.

A David Muñoz, por tus enseñanzas en Quaip.

A mis amigos del mundo del baile.

A mis alumnos.

Sin vosotros, este camino no hubiera sido posible. Gracias.

Resumen

Introducción

Según la Organización Mundial de la Salud (www.who.int), se estima que en el mundo hay más de 50 millones de personas que padecen algún tipo de demencia. Además, se espera que esta cifra se incremente hasta los 82 millones de personas en 2030 y hasta los 152 millones en 2050. En este contexto, entender cuáles son los mecanismos que desencadenan patologías como la Enfermedad de Alzheimer (EA) o como la Enfermedad de Parkinson (EP), como los dos trastornos neurodegenerativos con mayor incidencia, se riga como un pilar fundamental para intentar contener la inminente necesidad de recursos económicos y de personal, asociados al cuidado y atención de las personas que padecen este tipo de alteraciones.

Tradicionalmente, la evaluación clínica de la demencia solo ha contemplado la evaluación de algunos tests neuropsicológicos y la confirmación de ciertos hallazgos radiológicos en imagen estructural y/o funcional. Pese a que este tipo de pruebas constituyen una enorme fuente de información, su análisis ha sido en gran medida infravalorado limitándose, en muchos casos, a la simple aplicación de tests estadísticos univariantes justificados por el pequeño tamaño muestral de la mayor parte de estudios clínicos. La incorporación de técnicas de análisis de datos y el uso de modelos de inteligencia artificial está permitiendo desarrollar modelos predictivos capaces de determinar con gran nivel de precisión el estado y evolución de pacientes con demencia. Estas propuestas nos permiten comprender mejor cómo funciona el sistema nervioso central y nos ayudan a identificar los mecanismos que desencadenan sus dolencias.

En la última década, el desarrollo progresivo de los sistemas con aprendizaje máquina (*Machine Learning*, ML), ha permitido a la comunidad científica pasar de desarrollar sistemas de análisis de imágenes cerebrales simples, a modelar sistemas complejos con inteligencia artificial capaces de combinar toda la información procedente tanto de varias fuentes de imagen médica como de otros tipos de pruebas clínicas (incluyendo marcadores relacionados con la genética, la proteómica o la metabolómica, entre otros) para la predicción de trastornos neurológicos como la demencia. La mejora de estos sistemas ha sido, por un lado, consecuencia directa de la incorporación de

nuevos algoritmos de preprocesado, extracción de características, selección de características e incluso de bloques de aprendizaje como los árboles de decisión o las redes neuronales. Pero por otro lado también de la incorporación de herramientas como las distribuciones α -estables, capaces de explicar con gran nivel de precisión procesos biológicos más complejos como la distribución de tejidos cerebrales.

Aunque los sistemas con inteligencia artificial están consiguiendo cada vez mejores resultados, su aplicación en la clínica práctica convencional ha recibido también muchas críticas debido, sobre todo, a su falta de interpretabilidad y a su coste computacional. La inclusión de nuevos marcadores de imagen, como por ejemplo el uso de características morfológicas, ayuda a compensar este problema a la vez que permite caracterizar con gran nivel de precisión la evolución de pacientes con enfermedades como la EP.

Gracias a iniciativas como PPMI (*Parkinson's Progression Markers Initiative*) o DIAN (*Dominantly Inherited Alzheimer's Network*), la cantidad de marcadores al servicio de la investigación clínica para el desarrollo de modelos pronósticos de la demencia no deja de crecer. Estas bases de datos no solo sirven como banco de pruebas para el desarrollo de nuevos métodos de análisis sino que, además, nos permiten identificar nuevos marcadores, comparar diferentes modelos matemáticos, e incluso caracterizar sujetos que pueden desarrollar o pueden estar desarrollando algún tipo de demencia incluso en etapas todavía presintomáticas de su enfermedad.

En este contexto, la propuesta y desarrollo de nuevos métodos de análisis, la incorporación de nuevos marcadores y el uso de modelos que permitan combinar toda esta información, constituyen las tres líneas de trabajo más importantes de la neurociencia en la actualidad. Siguiendo esta línea, en esta tesis doctoral se ha propuesto un nuevo modelo de segmentación basado en el uso de distribuciones α -estables (Capítulo 4); un sistema de apoyo al diagnóstico de personas con EP basado en el uso de marcadores de imagen generados a partir de isosuperficies (Capítulo 5); un primer modelo utilizando técnicas de ML para la evaluación y seguimiento de pacientes con Alzheimer hereditario (Capítulo 6); y una nueva metodología de aprendizaje de conjuntos capaz de detectar y eliminar de manera automática aquellas fuentes de información de entrada poco relevantes durante la evaluación del estado y seguimiento de pacientes con algún tipo de demencia (Capítulo 7).

Antecedentes

Al margen del Capítulo 1 en el que se detalla la motivación, objetivos y estructura de este trabajo, en el Capítulo 2 se incluye una profunda revisión sobre los antecedentes de las publicaciones presentadas en esta tesis doctoral. Esta revisión comienza con una introducción sobre las técnicas de imagen médica, sus aplicaciones prácticas y las limitaciones de la neuroimagen clíni-

ca convencional. En este bloque también se desarrolla una revisión sobre las etapas que conforman un sistema de análisis con inteligencia artificial, con especial énfasis en el preprocesado de imágenes cerebrales; se comentan algunas aplicaciones prácticas de las distribuciones α -estables en el campo de la neurociencia; y se comentan detalles sobre el uso de técnicas de análisis multivariante de marcadores heterogéneos en neuroimagen incluyendo una revisión sobre los métodos de aprendizaje de conjuntos (*ensemble*) para el estudio de la demencia.

Uso de distribuciones α -estables en neuroimagen

Aunque las distribuciones α -estables han demostrado ser una herramienta muy eficaz en campos como la ingeniería eléctrica, la economía o la astrofísica; a la hora de modelar variables de entrada con fuerte impulsividad y/o con colas pesadas, su uso para aplicaciones dentro del campo de la neurociencia es todavía limitada. No obstante, y ya que algunas de estas aplicaciones requieren del uso de modelos de mezclas de distribuciones de probabilidad, una de las contribuciones presentadas para esta tesis doctoral ha estado relacionada con el desarrollo de un modelo de mezclas de distribuciones de probabilidad con α -estables. Esta propuesta, basada en el uso del algoritmo de maximización de la esperanza (*Expectation-Maximization*, EM), se ha utilizado como base para el desarrollo posterior de un sistema para la segmentación automática de imágenes estructurales en combinación con un modelo de campos ocultos de Markov (*Hidden Markov Random Fields*, HMRF) capaz de modelar las relaciones de intensidad entre vóxeles vecinos. Como se demuestra a partir de los resultados de los experimentos propuestos para su evaluación, esta propuesta es mucho más eficiente a nivel computacional que su equivalente bayesiana y consigue un mejor ajuste de la segmentación final. Además, como las distribuciones α -estables consiguen explicar mejor fenómenos como la asimetría de las distribuciones o la impulsividad de los histogramas, se logra reducir el número de componentes necesarias para explicar cualquier distribución de entrada, lo que simplifica la interpretabilidad de los resultados finales. Este Capítulo 4 se cierra con una comparativa que justifica la aplicación de modelos de normalización en intensidad con distribuciones α -estables durante el estudio y evaluación de imágenes funcionales relacionadas con la demencia gracias al aumento de la separación entre clases que añade esta normalización.

Caracterización de imágenes funcionales mediante isosuperficies

No solo con la incorporación de métodos de análisis más precisos se consigue una mejor caracterización de enfermedades como la demencia. A veces resulta necesario incluir nuevos marcadores que logren explicar su evolución.

Siguiendo esta idea, otra de las contribuciones de esta tesis ha sido el desarrollo de un sistema de ayuda al diagnóstico de la EP mediante el análisis morfológico de isosuperficies (el equivalente en 3D de las isolíneas en 2D) generadas a partir de imágenes cerebrales. Según esta propuesta, los niveles de intensidad de una imagen cerebral se pueden estratificar en N niveles diferentes a partir de los cuales se generan las isosuperficies. Luego se pueden extraer algunas de sus características morfológicas como el área, el volumen, la esfericidad o la posición relativa del centro de masas de cada isosuperficie. A partir de estas características, se pueden modelar trastornos como la EP tanto a nivel de diagnóstico como de análisis longitudinal. Para la propuesta presentada en esta tesis, se han analizado un conjunto de imágenes FP-CIT SPECT procedentes tanto de sujetos sanos como de pacientes con EP, y se han comparado los resultados de su clasificación (diagnóstico) utilizando tanto sus valores de intensidad (análisis tradicional) como nuestro modelo con isosuperficies. Como se concluye a partir de los resultados obtenidos, nuestra propuesta con isosuperficies no solo consigue igualar o mejorar los resultados de clasificación, sino que además disminuye el número de variables de entrada necesarias para predecir la EP (lo cual ayuda a limitar el posible sobreajuste a los datos de entrenamiento del sistema), evita la necesidad de aplicar ningún tipo de normalización espacial, y facilita la interpretabilidad. Además, el uso de esta técnica puede extenderse al análisis de otras patologías como la EA (a la hora de medir el grosor del córtex neuronal), la esclerosis múltiple (mediante la detección de focos de intensidad en imagen estructural) o el cáncer (ayudando a parcelar masas tumorales en el sistema nervioso central).

Análisis multivariante de marcadores heterogéneos

Aunque los tests estadísticos univariantes como el t-Test siguen constituyendo el tipo de prueba por defecto en investigación clínica para evaluar la utilidad de un marcador, esta tendencia está cambiando. En los últimos años, el número de estudios que han empezado a incorporar análisis más avanzados, como el uso de modelos predictivos multivariantes, está creciendo exponencialmente. Con el desarrollo de nuevas bases de datos, como es el caso del repositorio DIAN para el estudio de sujetos con Alzheimer hereditario (*Dominantly Inherited Alzheimer's Disease*, DIAD), surge también el interés de proponer un primer modelo con inteligencia artificial que permita estimar el estado y/o evolución de pacientes con esta variante de la EA, a la vez que ayude a validar/rechazar algunas de las hipótesis planteadas en trabajos anteriores sobre la enfermedad. Con este objetivo, en el Capítulo 6 se ha propuesto el desarrollo del primer modelo a nivel mundial para la evaluación de DIAD empleando con técnicas de ML. Este modelo ha tratado de comparar sujetos de control frente a pacientes portadores de

alguna mutación en cualquiera de sus genes PSEN1, PSEN2 y/o APP; con especial atención a las etapas presintomáticas de la enfermedad. Según los resultados obtenidos, se confirma que la EA no debería considerarse como una enfermedad con una etiología única. Además, el sistema ha sido capaz de identificar la existencia de dos tipos de sujetos con el gen PSEN1 mutado, desconocidos hasta la fecha, y con características claramente diferenciadoras incluso a edades muy tempranas. Se espera que a raíz de estas conclusiones se puedan desarrollar modelos más avanzados para predecir la sintomatología de los sujetos con DIAD a la vez que se abre la puerta al desarrollo de nuevas líneas de tratamiento para la EA común.

Por otro lado, y ya que muchos repositorios como DIAN o PPMI han comenzado a incorporar una gran cantidad de pruebas clínicas procedentes de pruebas de imagen, de análisis de sangre o de punciones lumbares, entre otros; también resulta interesante combinar de alguna forma todas estas fuentes de información para modelar de forma más precisa el estado y/o evolución de pacientes con demencia. Para este fin, se ha propuesto una nueva metodología de aprendizaje de conjuntos (*ensemble*) basada en el voto por mayoría tradicional pero que incluye un pesado no-lineal de los votos de cada fuente de información que se calcula en función del rendimiento individual de dicha fuente. Este sistema implica la existencia de un doble bucle de validación cruzada. Por un lado, en el bucle anidado se calcula el rendimiento de cada fuente por separado; y luego, en el bucle exterior, se asigna un peso a cada resultado para luego combinar el aprendizaje de todas las fuentes de entrada. La ventaja de esta propuesta, descrita en el Capítulo 7, es que permite descartar de manera automática fuentes de información de entrada poco fiables.

Conclusiones

La realización de esta tesis doctoral ha incluido una serie de propuestas, evaluadas experimentalmente, que tratan de mejorar el análisis de datos neurológicos para el estudio de la demencia desde tres puntos de vista diferentes:

1. Por un lado, se ha propuesto un método para el cálculo de componentes de un modelo de mezclas de distribuciones α -estables usando el algoritmo EM. Este método se ha empleado también para la segmentación de imágenes estructurales y pone en relieve, a través del análisis de las distribuciones de intensidad de la materia blanca y de la materia gris, la mejora que el uso de este tipo de distribuciones presenta a la hora de modelar datos de entrada con una fuerte impulsividad y/o con colas pesadas.

2. El segundo punto tiene que ver con la necesidad de proponer nuevos marcadores en neuroimagen que ayuden a explicar mejor el estado y la evolución de la demencia. Para ello se ha propuesto un análisis morfológico con isosuperficies que ha permitido cuantificar la pérdida de transportadores dopaminérgicos en la vía nigroestriada relacionadas con la neurodegeneración de la EP. Este modelo cuenta con la ventaja adicional de ser fácilmente interpretable desde el punto de vista físico, y de que su aplicación puede extenderse al análisis de otros tipos de trastornos neurológicos.
3. El último punto trata sobre cómo analizar toda la información disponible en bases de datos en las que se incorporan imágenes médicas y otros tipos de pruebas clínicas. Por un lado, se ha propuesto un análisis de cada una de las fuentes de información disponibles en repositorios como DIAN, y se ha evaluado la importancia que tiene cada uno de estos tipos de datos en diferentes fases de la enfermedad. Este modelo sólo ha tratado de avalar un modelo teórico publicado años atrás sobre la EA. Por otro lado, la segunda contribución en este punto ha estado relacionada con la propuesta de un método de aprendizaje de conjuntos (*ensemble*) basado en el voto por mayoría ponderado pero de forma no-lineal. Como se justifica a partir de los resultados obtenidos analizando el repositorio PPMI, esta propuesta es capaz de descartar de manera automática los tipos de datos de entrada poco fiables y puede emplearse tanto en escenarios de clasificación como de regresión.

En definitiva, la evolución del aprendizaje máquina en neurociencia requiere del desarrollo de nuevos modelos matemáticos más precisos; de una mejor caracterización de los procesos biológicos del sistema nervioso central, a partir de nuevos marcadores; y de la combinación, en términos de aprendizaje, de todas las fuentes de información relevantes que se les pueda proporcionar al sistema de aprendizaje.

Summary

Introduction

According to the World Health Organization (www.who.int), it is estimated that more than 50 million people worldwide suffer from some form of dementia. Moreover, this number is expected to increase to 82 million by 2030 and to 152 million by 2050. Dementia is caused by a number of pathologies, being the two most prevalent Alzheimer's and Parkinson's disease. Understanding the underlying mechanisms of these pathologies is a fundamental pillar to contain the imminent need for economic and human resources associated with the care and attention of people affected by them.

Historically, the clinical evaluation of dementia has only considered the evaluation of neuropsychological and biological tests and the confirmation of certain radiological findings in structural and/or functional imaging. Though such tests represent a vast amount of information, their analysis has been largely undervalued and limited, in many cases, to the simplistic application of univariate statistical tests justified by the small sample size of most clinical studies. The incorporation of advanced data analysis techniques and the use of artificial intelligence models is leading to the development of predictive models that can accurately determine the status and evolution of patients with dementia. These proposals improve our understanding of how our central nervous system works and can help to identify the mechanisms that trigger these disorders.

In the last decade, the progressive development of machine learning (ML) has allowed scientists to go from developing simple brain image analysis systems to modeling complex systems with artificial intelligence. These systems can combine all the information from different medical imaging sources and other types of clinical tests (including data from genetics, proteomics or metabolomics, among others) for the prediction of neurological disorders such as dementia. The improvement of these systems has been, on the one hand, a consequence of incorporating new algorithms for preprocessing, feature extraction, feature selection and using learning tools such as decision trees or neural networks. But on the other hand, it has also been the result of the use of probabilistic modelling with α -stable distributions, which are capable of explaining complex biological processes such as the distribution of brain

tissues with a high level of precision.

Although artificial intelligence systems are increasingly achieving outstanding results, their application in clinical practice has received much criticism, mainly due to their lack of interpretability and their computational cost. The inclusion of new imaging markers, such as morphological features, helps to compensate this problem at the same time that they explain the evolution of patients with diseases such as Parkinson's disease with a high level of accuracy.

Thanks to initiatives such as PPMI (Parkinson's Progression Markers Initiative) or DIAN (Dominantly Inherited Alzheimer's Network), the availability of markers at the service of clinical research for the development of prognostic models of dementia continues to grow. These databases not only serve as a benchmark for the evaluation of new methodologies, but also allow us to identify new biomarkers, compare different mathematical procedures, and even characterize subjects who may or not be developing some form of dementia, even at the pre-symptomatic stages of their disease.

In this context, the proposal of new analysis methods, the incorporation of new biomarkers and the development of predictive models combining all this information constitute the three main lines of work in the neuroscience field. Following this idea, this thesis proposes a new segmentation model based on the use of α -stable distributions (Chapter 4); a computer-aided-system for diagnosis of patients with Parkinson's disease based on the use of geometric image markers derived from 3D isosurfaces (Chapter 5); the first predictive model using ML techniques that attempts to characterize patients with dominantly inherited Alzheimer's disease (Chapter 6); and a novel methodology for ensemble learning based on majority voting that can automatically detect and eliminate irrelevant sources of input information (Chapter 7).

Background

In addition to Chapter 1, which details the motivation, objectives and structure of this work, Chapter 2 includes a thorough review of the background of the publications presented in this dissertation. This review begins with an introduction on medical imaging techniques, their practical applications and the limitations of conventional clinical neuroimaging. This chapter also includes a review of the stages involved in an artificial intelligence system with special emphasis on brain image preprocessing; the practical use of α -stable distributions in the field of neuroscience; and a discussion about the potential use of multivariate analysis for heterogeneous markers in neuroimaging. Finally, a review on ensemble learning methods for the study of dementia has also been included.

Use of α -stable distributions in neuroimaging

Although α -stable distributions have proven their ability in fields such as electrical engineering, economics or astrophysics when modeling input variables with strong impulsivity and/or heavy tails, their use in neuroscience applications is limited. Nevertheless, since these applications often require the use of mixture models of probability distributions, one of the contributions presented for this doctoral thesis has been related to the development of a finite mixture model using α -stable distributions. This proposal, based on the expectation-maximization (EM) and the Hidden Markov Random Fields (HMRF) algorithms, has also been used as the basis for the development of an automatic segmentation system for MRI images that also models the intensity relationships between neighboring voxels. As demonstrated from our experimental results, this approach is computationally much more efficient than the Bayesian alternative, and it achieves a better fit of the final segmentation. Moreover, since the α -stable distributions can better explain phenomena such as the asymmetry of input data or their impulsivity, the number of components needed to explain any input distribution is reduced, which simplifies the interpretability of the final results. This Chapter closes with a comparison that justifies the application of intensity normalization models with α -stable distributions. These models increase the inter-class separability during the study and evaluation of dementia-related functional images.

Characterization of functional imaging using isosurfaces

Not only with the incorporation of more accurate methods of analysis we could achieve a better characterization of diseases such as dementia. Sometimes it is necessary to include new markers to explain its evolution. Following this idea, another contribution of this dissertation has been the development of a computer-aided diagnosis system for Parkinson's disease through a morphological analysis using isosurfaces (the 3D equivalent of 2D isolines). According to this proposal, the intensity levels of a cerebral image are stratified into N different levels from which isosurfaces are generated. Then its morphological characteristics (such as area, volume, sphericity or relative position of the center of mass) are extracted and used as input for classification and/or regressions systems to evaluate neurological disorders like Parkinson's disease. In fact, we have used this proposal to evaluate a set of FP-CIT SPECT images from both healthy subjects and patients with Parkinson's disease, and we have compared the results obtained using traditional voxel intensity levels classification. As concluded from the results obtained, our proposal not only achieves equal or better classification results, but also decreases the number of input variables needed to predict

Parkinson's disease (which helps to limit the possible overfitting to the training data of the system), avoids applying any kind of spatial normalization, and improves the overall interpretability of the system. The application of this technique can be extended to the analysis of other pathologies such as Alzheimer's disease (when measuring the thickness of the neuronal cortex), multiple sclerosis (by detecting hyperintensities in structural imaging) or cancer (helping to parcel out tumor masses in the central nervous system).

Multivariate analysis of heterogeneous markers

Although univariate statistical tests such as the t-test remain as the default method for assessing the significance of a marker in clinical research, this trend is changing. In recent years, the number of studies that have begun to incorporate advanced analyses such as multivariate predictive models is growing exponentially. With the release of new databases, as is the case of the DIAN repository for the study of subjects with Dominantly Inherited Alzheimer's Disease (DIAD), there was a growing interest in proposing an initial model with artificial intelligence to estimate the status and/or evolution of patients with this variant of Alzheimer. With this aim, in Chapter 6 we have proposed the first model for the evaluation of DIAD using ML techniques. This study compares control subjects versus patients carrying a mutation in any of their PSEN1, PSEN2 and/or APP genes; with special focus on the presymptomatic stages of the disease. This study concludes that Alzheimer's should not be considered as a disease with a single etiology, and recognizes the existence of two types of PSEN1 mutated gene subjects clearly differentiated. It is expected that these findings may lead to the development of more advanced models to predict DIAD symptomatology while also opening the door to the development of new lines of treatment for common Alzheimer's.

On the other hand, and since many repositories such as DIAN or PPMI have begun to incorporate a large amount of clinical data such as imaging scans, blood tests or lumbar punctures; it has also been interesting to combine these features to create more accurate models of dementia. To this end, a new ensemble learning methodology has been proposed. This proposal is based on classical majority voting technique but including a non-linear weighting of each input that is calculated based on its individual performance. This system requires the existence of a double cross-validation loop. On the one hand, in the nested loop, the performance of each source is calculated separately; and then, in the outer loop, a set of weights is calculated and assigned to each input source. The advantage of this approach, described in Chapter 7, is that it allows unreliable input sources to be automatically discarded.

Conclusions

This thesis includes a series of proposals that attempt to improve the analysis of neurological data for the study of dementia from three different points of view:

1. First, a method for the computation of a mixture model of *alpha*-stable distributions using the EM algorithm has been designed. This method has also been employed for a structural image segmentation proposal through the analysis of white matter and gray matter intensity distributions. The use of α -stable distributions helps to improve the modeling of input data even despite their potential impulsiveness and / or heavy tails.
2. The second point is related to the need for proposing new neuroimaging markers that explain better the status and progression of patients with dementia. To this end, a morphological analysis based on isosurfaces has been proposed. It is used to quantify the loss of dopaminergic transporters in the nigrostriatal pathway related to neurodegeneration in Parkinson's disease. This model counts with the additional advantage of being physically interpretable, and its application can be extended to the analysis of other neurological disorders.
3. The last point deals with analyzing heterogeneous data sources. On the one hand, we have addressed the characterization of the information available in DIAN by studying each source of information separately to determine the reliability of each type of marker at different stages of the disease. In this context, the proposed model endorses a theoretical model published years ago on Alzheimer's. On the other hand, an ensemble learning method using nonlinearly weighted majority voting has been proposed. As justified from the results obtained when analyzing the PPMI repository, this proposal can automatically discard unreliable input data types and might be used for both classification and regression scenarios.

In short, the evolution of machine learning in neuroscience requires the development of new and more accurate mathematical models, a further characterization of the biological processes at the central nervous system using new markers, and the combination of all the relevant sources of information that can be provided to the learning system.

Índice general

| | |
|--|-----------|
| 1. Interés, objetivos y estructura de la tesis | 3 |
| 1.1. Interés del estudio | 3 |
| 1.2. Objetivos y metas | 6 |
| 1.3. Estructura del documento | 7 |
| 1.4. Listado de contribuciones científicas asociadas a la tesis doctoral | 12 |
| 2. Estado del arte | 15 |
| 2.1. Técnicas de imagen médica | 15 |
| 2.1.1. Imagen de resonancia magnética (MRI) | 16 |
| 2.1.2. Tomografía Computarizada por Emisión de Fotón Único (SPECT) | 16 |
| 2.1.3. Tomografía por Emisión de Positrones (PET) | 17 |
| 2.2. Aplicaciones clínicas de la neuroimagen | 17 |
| 2.2.1. Evaluación clínica de la Enfermedad de Alzheimer | 18 |
| 2.2.2. Evaluación clínica de la Enfermedad de Parkinson | 20 |
| 2.3. Limitaciones de la neuroimagen clínica convencional | 20 |
| 2.4. Modelos de análisis automático para neuroimagen | 21 |
| 2.4.1. Introducción | 21 |
| 2.4.2. Esquema general de un modelo de clasificación/regresión con aprendizaje máquina | 23 |
| 2.4.3. Preprocesado de imágenes cerebrales | 25 |
| 2.4.4. Métodos estadísticos para la extracción de características | 35 |
| 2.4.5. Selección de características | 37 |
| 2.4.6. Algoritmos de clasificación y regresión | 42 |
| 2.5. Aplicaciones de las distribuciones alfa-estables en neuroimagen | 54 |
| 2.6. Caracterización de patrones morfológicos en neuroimagen | 54 |
| 2.7. Análisis multivariante de marcadores heterogéneos con inteligencia artificial | 55 |
| 2.8. Análisis multivariante de marcadores heterogéneos mediante aprendizaje de conjuntos | 57 |

| | |
|---|------------|
| 3. Bases de datos y métodos de evaluación | 61 |
| 3.1. Repositorios utilizados durante la realización de la tesis doctoral | 61 |
| 3.1.1. Parkinson’s Progressions Markers Initiative (PPMI) | 61 |
| 3.1.2. Imágenes del Hospital Virgen de la Victoria de Málaga | 62 |
| 3.1.3. Dominantly Inherited Alzheimer’s Network (DIAN) | 63 |
| 3.1.4. Internet Brain Segmentation Repository (IBSR) | 67 |
| 3.2. Evaluación de la capacidad de aprendizaje de un modelo de clasificación | 68 |
| 3.3. Tests de permutación | 69 |
| 4. Uso de distribuciones alfa-estables en neuroimagen | 73 |
| 4.1. Introducción a las distribuciones alfa-estables | 73 |
| 4.1.1. Forma general | 73 |
| 4.1.2. Representación | 74 |
| 4.1.3. Estadísticos descriptivos | 74 |
| 4.2. Modelo de mezclas de distribuciones alfa-estables con el algoritmo EM | 75 |
| 4.2.1. Propuesta | 75 |
| 4.2.2. Experimentos para justificar el modelo de mezclas de distribuciones alfa-estables | 77 |
| 4.3. Aplicación del modelo de mezclas de distribuciones alfa-estables a la segmentación de imágenes estructurales | 81 |
| 4.3.1. Justificación de la propuesta | 81 |
| 4.3.2. Modelo propuesto para la mejora de la segmentación de imágenes estructurales | 82 |
| 4.3.3. Experimentos para cuantificar la mejora de los resultados de segmentación | 84 |
| 4.4. Normalización en intensidad de imágenes funcionales con distribuciones alfa-estables | 87 |
| 5. Extracción de características morfológicas en neuroimagen | 91 |
| 5.1. Caracterización del Parkinson mediante isosuperficies | 91 |
| 5.1.1. Análisis de la magnitud del gradiente en imagen funcional | 91 |
| 5.1.2. Propuesta para el análisis de imágenes FP-CIT SPECT mediante isosuperficies para caracterizar la EP | 93 |
| 5.1.3. Implementación del modelo | 93 |
| 5.1.4. Análisis de resultados | 100 |
| 6. Análisis multivariante de biomarcadores heterogéneos para el estudio de la base de datos DIAN | 105 |
| 6.1. Evaluación de la base de datos DIAN | 105 |
| 6.1.1. Justificación del análisis | 105 |

| | | |
|-----------|---|------------|
| 6.1.2. | Propuesta para el análisis de fuentes heterogéneas y definición de experimentos | 107 |
| 6.1.3. | Esquema de clasificación | 108 |
| 6.1.4. | Análisis de resultados | 109 |
| 7. | Análisis multivariante mediante aprendizaje de conjuntos | 121 |
| 7.1. | Metodología para la mejora del voto por mayoría ponderado | 121 |
| 7.1.1. | Justificación | 121 |
| 7.1.2. | Propuesta | 122 |
| 7.1.3. | Implementación | 124 |
| 7.1.4. | Análisis de resultados | 126 |
| 8. | Discusión y Conclusiones | 137 |
| 8.1. | Discusión | 137 |
| 8.1.1. | Uso de distribuciones alfa-estables en el preprocesado de imágenes neuronales | 137 |
| 8.1.2. | Marcadores morfológicos para el análisis de la demencia | 140 |
| 8.1.3. | Análisis de marcadores procedentes de fuentes heterogéneas | 144 |
| 8.2. | Conclusiones (Versión en castellano) | 156 |
| 8.3. | Conclusions (English version) | 159 |
| | Bibliografía | 220 |
| | Abreviaturas y Siglas | 223 |

Índice de figuras

| | | |
|------|---|----|
| 1.1. | Diagrama general de la tesis doctoral. | 11 |
| 2.1. | Vista axial de un sujeto sano (etiquetado como HC) y un sujeto con EA avanzada (etiquetado como AZ) de la base de datos ADNI. Imagen funcional 18F-FDG PET (fila superior) e imagen estructural MRI-1T (fila inferior). | 19 |
| 2.2. | Vista axial de dos imágenes FP-CIT SPECT de la base de datos PPMI correspondientes a un sujeto de control (izquierda) y un paciente con EP (derecha). | 21 |
| 2.3. | Diagrama general de un problema de clasificación/regresión con ML. | 23 |
| 2.4. | Plantilla <code>labels_Neuromorphometrics.nii</code> de <i>SPM12</i> . Ejemplo representativo con cortes axial, sagital y coronal donde cada color representa una etiqueta (región anatómica) diferente. | 32 |
| 2.5. | Diagrama explicativo de un sistema de clasificación con RF. | 47 |
| 2.6. | Arquitectura básica de una red neuronal con dos capas ocultas. | 49 |
| 3.1. | Ejemplo de matriz de confusión para un problema binario con etiquetas 0 y 1. | 68 |
| 3.2. | Representación gráfica de un test de permutación. Como la tasa de clasificación con las etiquetas originales (rojo) es superior al umbral (en verde, y definido en este caso particular a partir del 95 % de resultados de clasificación con re-etiquetado), entonces se considera que el resultado de clasificación es estadísticamente significativo. | 70 |
| 4.1. | Parámetros de la distribución α -stable y su influencia en la forma de la distribución. Por defecto, $\{\alpha = 1, \beta = 0, \gamma = 1, \delta = 0\}$ | 74 |
| 4.2. | Histograma generado a partir de dos distribuciones α -estables con 10000 puntos. Las líneas continuas representan la predicción del conjunto (suma de componentes) mientras que las líneas discontinuas representan cada una de las componentes por separado. | 78 |
| 4.3. | Histograma generado a partir de tres distribuciones α -estables con 10000 puntos. Las líneas continuas representan la predicción del conjunto (suma de componentes) mientras que las líneas discontinuas representan cada una de las componentes por separado. | 80 |

| | | |
|-------|---|-----|
| 4.4. | Número total de casos confirmados de infección por SARS-CoV-2 en Corea del Sur durante el período comprendido entre el 20 de diciembre de 2019 y el 25 de abril de 2020. La línea continua representa la mezcla de distribuciones mientras que las líneas discontinuas hacen referencia a cada componente por separado. | 80 |
| 4.5. | Distribución de intensidades de una imagen MRI perteneciente al repositorio IBSR. | 81 |
| 4.6. | Esquema representativo de un vóxel (blanco) y sus 6 vóxeles vecinos (gris) en un espacio 3D. | 84 |
| 4.7. | Segmentación manual del repositorio IBSR y su comparación utilizando el modelo propuesto de segmentación. | 86 |
| 4.8. | Distribuciones de intensidad de 30 imágenes FP-CIT SPECT del repositorio PPMI antes de aplicar ningún tipo de normalización en intensidad (izquierda) y después de aplicar la normalización en intensidad con distribuciones α -estables (derecha). | 87 |
| 4.9. | Vista axial de las 30 imágenes FP-CIT SPECT del repositorio PPMI antes de aplicar ningún tipo de normalización en intensidad (izquierda) y después de aplicar la normalización en intensidad con distribuciones α -estables (derecha). | 88 |
| 4.10. | Valores z_{Score} , calculados mediante el algoritmo Mann-Whitney-Wilcoxon U-Test y ordenados en sentido decreciente, para los 47104 valores de intensidad de la región del estriado de 626 sujetos incluidos en el repositorio PPMI. | 89 |
| | | |
| 5.1. | Ejemplo del corte axial de una imagen FP-CIT SPECT del repositorio PPMI (izquierda) y la magnitud del gradiente de la imagen (derecha). | 92 |
| 5.2. | Ejemplo con 4 de los patrones del algoritmo Marching Cubes. . . . | 94 |
| 5.3. | Ejemplo de dos isosuperficies calculadas usando el mismo umbral de intensidad para dos imágenes FP-CIT SPECT del repositorio PPMI. Las imágenes han sido previamente normalizadas espacialmente (registro afín) y en intensidad (método con α -estables descrito en 4.4). | 94 |
| 5.4. | Ejemplo de $N = 7$ isosuperficies calculadas para el primer sujeto sano (HC) del repositorio PPMI. | 95 |
| 5.5. | Texto. | 96 |
| 5.6. | Proyecciones sobre los planos X,Y,Z de las isosuperficies en el hemisferio derecho representadas en la Figura 5.3. | 97 |
| 5.7. | Esquema de la red neuronal MLP. | 100 |
| 5.8. | Resultados de clasificación para diferentes números de isosuperficies en función de la región evaluada y métodos de clasificación evaluados. | 101 |
| 5.9. | Variabilidad de resultados de clasificación para los modelos con $N = 20$ isosuperficies que hacen uso del conjunto de características $[L_H, R_H, B]$ | 102 |

| | |
|---|-----|
| 5.10. Curvas ROC para los experimentos que utilizan ν SVC y MLP. | 103 |
| 6.1. Modelo teórico propuesto por Jack et al. | 106 |
| 6.2. Resultados de clasificación (en términos del parámetro <i>balanced accuracy</i>) obtenidos para las comparaciones entre NC y MC considerando todos los MC (arriba) o solo los MC asintomáticos (abajo). | 110 |
| 6.3. Resultados de clasificación (en términos del parámetro <i>accuracy</i>) obtenidos para las comparaciones entre NC, PSEN1, PSEN2 y APP. | 111 |
| 6.4. Resultados de clasificación (en términos del parámetro <i>balanced accuracy</i>) para las comparaciones entre NC y MC_{PSEN1} asintomáticos. | 111 |
| 6.5. Representación del logaritmo del p_{valor} asociado a cada región para las comparaciones de NC frente a MC asintomáticos (izquierda), y de NC frente a MC_{PSEN1} asintomáticos (derecha). | 112 |
| 6.6. Histogramas de los 8 marcadores FDG con mejor p_{valor} según el test de ANOVA cuando se comparan los subconjuntos NC y MC_{PSEN1} asintomáticos. | 113 |
| 6.7. <i>Clustering</i> generado a partir de los marcadores FDG incluyendo NC y MC_{PSEN1} asintomáticos. Representación gráfica incluyendo las 3 primeras componentes de PCA. | 114 |
| 6.8. Tasas de clasificación, en términos del parámetro <i>balanced accuracy</i> , a la hora de comparar los subconjuntos MC_{PSEN1}^{NC-} (Cluster1) y MC_{PSEN1}^{NC+} (Cluster2) frente a los sujetos NC. | 116 |
| 6.9. Resultados de clasificación cuando se comparan los grupos NC y MC_{PSEN1} asintomáticos con diferentes umbrales de la variable EYO. | 117 |
| 6.10. Porcentaje de variabilidad explicada por PCA en función del número de componentes N_{Comp} | 117 |
| 6.11. Curvas ROC asociadas a la comparación NC frente a MC incluyendo todos los casos (fila superior) y NC frente a MC asintomáticos (fila inferior). | 118 |
| 6.12. Curvas ROC asociadas a la comparación NC frente a MC_{PSEN1} incluyendo todos los casos (fila superior) y NC frente a MC_{PSEN1} asintomáticos (fila inferior). | 119 |
| 6.13. Resultados del área bajo la curva calculada para cada experimento de clasificación. | 120 |
| 7.1. Diagrama del modelo de clasificación con <i>ensemble</i> propuesto. | 123 |
| 7.2. Representación gráfica de tres de las propuestas de inventariado para la propuesta de clasificación con <i>ensemble</i> | 127 |
| 7.3. Resultados de los tests de permutación. | 129 |
| 7.4. Resultados de clasificación por tipo de marcador y cuando se aplica el método de <i>ensemble</i> con inventariado cuadrático. | 133 |
| 7.5. Curvas ROC asociadas a cada experimento de clasificación para cada tipo de marcador junto a la curva ROC que se obtiene aplicando el método de <i>ensemble</i> con inventariado cuadrático. | 134 |

| | |
|---|-----|
| 8.1. Representación gráfica de una ROI definida utilizando diferentes umbrales de intensidad para el análisis VAF. | 143 |
| 8.2. Resultados de clasificación (en términos del parámetro <i>balanced accuracy</i>) para diferentes valores del parámetro N_{Comp} utilizando el modelo con PCA (izquierda). Varianza acumulada en función del parámetro N_{Comp} (derecha). | 143 |
| 8.3. Regiones más significativas para el análisis longitudinal de la base de datos de DIAN propuesto por Lockett et al. | 150 |

Índice de tablas

| | |
|--|----|
| 1.1. Trabajos publicados como primer autor durante la realización de la tesis doctoral. | 12 |
| 2.1. Ejemplos de trabajos de análisis longitudinal (no basados en el uso de redes neuronales) para la evaluación y/o predicción de la EP y de la EA. | 52 |
| 2.2. Ejemplos de trabajos de análisis longitudinal (basados en el uso de redes neuronales) para la evaluación y/o predicción de la EP y de la EA. | 53 |
| 3.1. Demográficas del repositorio PPMI para un escenario con balanceo de clases. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar. | 62 |
| 3.2. Demográficas del repositorio FP-CIT SPECT del Hospital Virgen de la Victoria de Málaga. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar. | 62 |
| 3.3. Demográficas del repositorio DIAN. (*: la variable sexo está referida al porcentaje de mujeres en cada subconjunto). | 64 |
| 3.4. Listado de patologías cuyo diagnóstico positivo se considera criterio de exclusión. | 65 |
| 3.5. Sujetos en cada nivel de la escala CDR según su tipo de mutación. | 65 |
| 3.6. Demográficas del repositorio IBSR. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar. | 67 |
| 4.1. Mezcla de 2 distribuciones α -estables. | 78 |
| 4.2. Mezcla de 3 distribuciones α -estables. | 79 |
| 4.3. Porcentaje de vóxeles correctamente etiquetados según el modelo de segmentación propuesto con distribuciones α -estables y su comparación utilizando distribuciones gaussianas. | 85 |
| 5.1. Resultados de la clasificación utilizando la magnitud del gradiente de la imagen funcional en lugar de la imagen original. | 92 |
| 5.2. Clasificadores propuestos para el análisis de características morfológicas en volúmenes FP-CIT SPECT. | 99 |

| | | |
|------|--|-----|
| 5.3. | Resultados de clasificación utilizando un total de $N = 20$ isosuperficies para el hemisferio izquierdo (L_H), hemisferio derecho (R_H), cerebro completo (B), y la concatenación de todos los anteriores ($[L_H, R_H, B]$). | 101 |
| 5.4. | Características morfológicas con menor p_{valor} (promediado entre todos los conjuntos de entrenamiento) al comparar sujetos sanos y pacientes con EP en un escenario con $N = 20$ isosuperficies. | 102 |
| 6.1. | Resumen de resultados de clasificación (en términos del parámetro <i>balanced accuracy</i>) entre NC y MC_{PSEN1} asintomáticos incluyendo el resultado de la resustitución y el límite superior del error empírico según Górriz et al. | 113 |
| 6.2. | Demográficas de los subconjuntos $MC_{\text{PSEN1}}^{\text{NC-}}$ y $MC_{\text{PSEN1}}^{\text{NC+}}$ cuando se utilizan todos los sujetos MC_{PSEN1} o solo los asintomáticos. | 115 |
| 7.1. | Relación de experimentos de clasificación propuestos. Al final no se han incluido los marcadores de tipo Plasma debido a su baja significancia estadística según el test de Welch como se verá más adelante. | 126 |
| 7.2. | Propuestas para el inventanado de los pesos w_j^{INT} usando la predicción de las clases del bucle externo (escenario binario con etiquetas 1 y -1). | 126 |
| 7.3. | Promedio de resultados de la clasificación individual en el bucle anidado. | 130 |
| 7.4. | Resultados del t-Test aplicado a los resultados de clasificación del test de permutación. El estadístico σ hace referencia a la desviación típica del t-Test. | 131 |
| 7.5. | Resultados de clasificación final (en términos del parámetro tasa de acierto) obtenidos al aplicar diferentes métodos de <i>ensemble</i> | 132 |
| 7.6. | Pesos obtenidos para cada tipo de dato cuando se usa el método de inventanado cuadrático. | 132 |
| 8.1. | Comparativa de los resultados de clasificación reportados en trabajos en trabajos anteriores sobre el uso de características morfológicas para el estudio de la EP. | 141 |

Bloque A

Interés, objetivos y estructura de la tesis

Capítulo 1

Interés, objetivos y estructura de la tesis

1.1. Interés del estudio

La incorporación de nuevas técnicas de análisis de datos e inteligencia artificial en el campo de la neurología está dando paso a sistemas de ayuda al diagnóstico (*Computer-Aided-Diagnosis systems*, CAD) cada vez más precisos [1–4]. Las aplicaciones prácticas de estos sistemas no solo permiten comprender mejor los mecanismos que desencadenan determinadas neuropatías como la demencia o validar ciertas hipótesis médicas, sino que también tienen un uso potencial en la práctica clínica convencional [5, 6].

Hasta hace poco, muchos de estos modelos, la mayoría basados en técnicas de aprendizaje máquina (*Machine Learning*, ML), han hecho un uso casi exclusivo de imágenes cerebrales para la caracterización de patrones estructurales y/o funcionales a partir de los cuales determinar el estado o evolución de patologías como la Enfermedad de Alzheimer (EA) o como la Enfermedad de Parkinson (EP) [7–16]. No obstante, la complejidad que han ido añadiendo estos modelos con el tiempo no siempre se ha traducido en una mejora sustancial de las tasas de acierto diagnóstico (clasificación), o pronóstico (regresión), ni en el desarrollo de nuevos tratamientos. Este hecho ha impedido que, en los últimos años, este tipo de soluciones se hayan terminado por convertir en un estándar para el apoyo en el diagnóstico clínico. Por otro lado, en algunos escenarios la falta de interpretación física de muchos de los modelos presentados hasta ahora, ha sido también objeto de críticas por parte de una comunidad médica a veces más interesada en la búsqueda clásica de marcadores diana que en la propia comprensión de las interacciones que un conjunto de variables estadísticamente significativas puede tener sobre una enfermedad [17–19].

La mejora de los métodos de modelado matemático no solo contribuye a una mayor comprensión sobre el funcionamiento de la demencia, sino que

también sirve como base para el estudio de otras patologías neurológicas e incluso del propio cerebro sano. Por ejemplo, el análisis de los valores de intensidad en imágenes estructurales permite crear un mapa de segmentación de los tejidos cerebrales (materia blanca, materia gris y líquido cefalorraquídeo) que, a su vez, puede utilizarse para evaluar el proceso de envejecimiento cerebral, detectar procesos oncológicos en el sistema nervioso central o para determinar si la distribución anormal de estos tejidos puede estar relacionada con alteraciones de la mielina, entre otras muchas aplicaciones [20–24]. Pese a que la segmentación de imágenes es solo un caso particular, este ejemplo puede extrapolarse a cualquier ámbito de estudio dentro del campo de la neurología. Cuanto mejor sea el ajuste a los datos de entrada relacionados con el proceso biológico de interés, mayor será nuestra capacidad para explicar sus mecanismos y más fácil será desarrollar tratamientos y/o mejorar las pruebas diagnósticas actuales. Todo esto se traduce, a largo plazo, en mejores tiempos de respuesta y en una mejora en la calidad de vida de los pacientes neurológicos. En este ajuste entre los modelos matemáticos y los datos de entrada, históricamente, siempre se ha utilizado el teorema central del límite como una justificación teórica del uso de distribuciones gaussianas para el modelado de procesos físicos naturales. Sin embargo, existen multitud de escenarios (como es el caso de la neuroimagen) donde la impulsividad y asimetría de las señales impiden su correcto ajuste mediante distribuciones gaussianas. De hecho, en el campo de la segmentación de imágenes estructurales, existen modelos de mezclas de distribuciones que asumen la normalidad de los datos de entrada pese a que los histogramas de intensidad presentan distribuciones con colas pesadas y una fuerte asimetría [25–27]. Aunque en este tipo de escenarios primero habría que comprobar la normalidad de los datos de entrada mediante un test específico como el test de Anderson-Darling [28] o el test de Shapiro-Wilk [29], en caso de que no se cumplan las condiciones necesarias para este tipo de pruebas será conveniente recurrir a modelos no-paramétricos o a otro tipo de sistemas que asuman una distribución de los datos de entrada diferente. En este contexto, el uso de distribuciones α -estables, [30], gracias a su capacidad de generalización, constituye una herramienta muy versátil en el campo de la neuroimagen y, en concreto, en etapas relacionadas con el preprocesado de imágenes tanto estructurales como funcionales [31].

Incluso cuando se mejora el ajuste de los datos de entrada, a veces la cantidad de información disponible puede no ser suficiente para explicar el desarrollo de una enfermedad. En este escenario, muchos trabajos apuntan a la necesidad de proponer nuevos tipos de marcadores relacionados con la histología de la enfermedad que se quiere analizar [32–36]. Pese a que la dinámica actual está centrada en la búsqueda de mutaciones genéticas, expresión de proteínas e incluso en procesos relacionados con la metabolómica; el análisis de la imagen cerebral también está ayudando a solventar esta necesidad gracias al desarrollo de nuevas características y a la identificación

de patrones que contribuyen a un mejor conocimiento de trastornos neurológicos que incluso quedaban excluidos del análisis de imagen tradicional [37–40]. Sea cual sea la finalidad del estudio, algunos de los trabajos con ML que se están llevando a cabo en la actualidad están consiguiendo no solo profundizar en la patogénesis de la demencia sino que también ayudan a reducir el número de características necesarias para su evaluación. Un ejemplo de esto lo podemos encontrar en el trabajo publicado en [41] donde los autores proponen un análisis morfológico de las zonas con mayores niveles de intensidad en imagen funcional para el diagnóstico del Parkinson. A diferencia del enfoque tradicional basado en un análisis a nivel de vóxeles de intensidad, un análisis morfológico consigue una alta tasa de acierto diagnóstico a la vez que reduce el sobreajuste de los modelos matemáticos a los datos de entrada (haciendo el modelo más generalizable) y mejora la interpretabilidad de los resultados. En este contexto, el uso de isosuperficies de intensidad (entendidas como una generalización tridimensional de las isolíneas) permite no solo modelar el estado y evolución de pacientes con demencia, sino que también abre la puerta a una caracterización más eficiente de lesiones cerebrales incluyendo la posible existencia de metástasis, focos de intensidad asociados al avance de la esclerosis múltiple o traumatismos craneoencefálicos, entre otros; e incluso a cuantificar la actividad metabólica cerebral.

Con la aparición de nuevas bases de datos que incorporan múltiples tipos de biomarcadores (pruebas de imagen, proteínas, lípidos, péptidos, factores de crecimiento, genes, etc.), [42], surge la necesidad de analizar estas fuentes de información para así determinar si consiguen una mejora significativa respecto a las tasas de clasificación/regresión de las propuestas actuales [43]. El análisis individual de estas fuentes de datos¹ permite no solo descartar aquellos tipos de marcadores que no son relevantes desde un punto de vista estadístico o cuya obtención no compensa el coste de tiempo y/o recursos necesarios para su obtención; sino que también permite generalizar su evolución. Si, por el contrario, lo que se pretende es combinar la información de dos o más fuentes de datos heterogéneas, una de las soluciones más comunes en el campo de la inteligencia artificial está basada en el uso de los conocidos como métodos de conjunto (o *ensemble* en inglés) [44, 45]. Mediante un modelo de *ensemble*, la predicción de varios bloques de aprendizaje individuales puede combinarse entre sí consiguiendo una mejora de las tasas de acierto en clasificación/regresión y mejorando la capacidad de generalización del sistema incluso en escenarios con pocas muestras de entrada y una alta dimensionalidad. No obstante, este tipo de mejoras llevan, en la mayoría de casos, asociado un incremento del coste computacional y una falta de interpretabilidad a tener en cuenta en ciertos tipos de problemas.

¹No se refiere a un análisis univariante de cada variable de entrada por separado sino al análisis de una o más variables que comparten ciertas características comunes entre ellas como, por ejemplo, el tipo de prueba clínica con la que se han obtenido.

Pese a la importancia creciente en los últimos años del uso de redes neuronales en el análisis de imágenes cerebrales, y en concreto de redes neuronales profundas (*deep learning*), su aplicación a la práctica clínica ha recibido varias críticas sobre la base de su falta de interpretabilidad física, al número de muestras que requieren para su entrenamiento, a su eficiencia computacional y a su coste de ejecución [46–48]. Cuando se trabaja con redes neuronales, tenemos una serie de pesos que son combinaciones lineales de otras combinaciones de pesos anteriores y que limitan la interpretabilidad de los resultados cuanto mayor es el nivel de complejidad de la red. Por otro lado, este tipo de soluciones requiere un gran número de muestras de entrada, lo que resulta un factor limitante en escenarios de análisis de datos como los procedentes de imagen cerebral ya que el coste asociado a la adquisición de este tipo de pruebas (y a su etiquetado por parte de profesionales bien entrenados) resulta elevado. Si se comparan los resultados de clasificación/regresión respecto a otras soluciones más tradicionales como el uso de máquinas de vectores de soporte (*Support Vector Machines*, SVM) [49], la aplicación de soluciones con *deep learning* puede no resultar eficiente ya que una red neuronal de este tipo funciona como un aproximador universal de funciones [50], así que su implementación para resolver problemas sencillos (como sería el caso del operador suma) puede requerir una cantidad enorme de recursos asociados al cálculo de los parámetros de la red, al almacenamiento de los pesos o al número de ejemplos de entrada necesarios para entrenar la red. Dicho de otra forma, si el entrenamiento del modelo es más complejo que la resolución del problema en sí, entonces no deberían utilizarse este tipo de soluciones.

En definitiva, la búsqueda de nuevos biomarcadores, su correcto modelado y la información (individual o en conjuntos) que estos proporcionan, constituyen tres de las líneas más importantes en el campo de la neurociencia y su aplicación práctica para el diagnóstico clínico y como medio para cuantificar la progresión de patologías como la demencia.

1.2. Objetivos y metas

En el marco descrito anteriormente, esta tesis tiene como objetivo fundamental proponer nuevos tipos de marcadores de imagen para el estudio de los trastornos neurodegenerativos más frecuentes, y proponer algún enfoque que permita combinar la información procedente de diferentes tipos de datos en escenarios con inteligencia artificial. La obtención de estos marcadores deberá tener en cuenta las potencialidades y amplio rango de aplicación que la distribución α -estable tiene en el campo de las aplicaciones biomédicas y especialmente en neuroimagen. Según esto se definen los siguientes objetivos:

- Revisar los trabajos más recientes que hagan uso de distribuciones α -estables en neuroimagen y evaluar su aplicación en el preprocesado de imágenes cerebrales en etapas como la segmentación de imágenes estructurales o la normalización de imágenes funcionales.
- Proponer nuevos marcadores (en este caso características morfológicas a partir de las pruebas de imagen médica) que puedan utilizarse en la evaluación (diagnóstico) de trastornos neurodegenerativos como la EP o la EA utilizando técnicas de ML. En particular, desarrollar marcadores que permitan superar ciertas limitaciones que tienen este tipo de sistemas (como el sobreajuste de los bloques de aprendizaje a los datos de entrenamiento) pero tratando de conservar la interpretabilidad física al menos de aquellas variables que sean más significativas desde el punto de vista estadístico.
- Evaluar, mediante técnicas de ML, la influencia que diferentes tipos de marcadores tienen sobre el curso de las neuropatías de interés incluso en etapas de la enfermedad en las que todavía no se han manifestado los primeros síntomas clínicos.
- Proponer una nueva metodología de aprendizaje de conjuntos (*ensemble*) que permita combinar toda la información procedente de diferentes fuentes de datos heterogéneas² que le de más peso a las variables con mayores tasas de clasificación en detrimento de aquellas con una baja separación entre clases.

1.3. Estructura del documento

Una vez presentado el interés, marcados los objetivos de esta tesis doctoral y siguiendo el diagrama general incluido en la Figura 1.1 donde se ha esquematizado la relación entre todos los bloques y se han señalado las publicaciones asociadas a cada punto (recogidas posteriormente en la Sección 1.4); se expone la siguiente relación de contenidos desarrollados en un total de 7 capítulos como sigue a continuación:

- **Capítulo 2.** La revisión del estado del arte comienza con una breve introducción a las técnicas de imagen médica que más se utilizan hoy en día para evaluar la demencia, describe las aplicaciones clínicas de la neuroimagen convencional y señala cuáles son sus limitaciones. Con la inclusión de los primeros modelos de aprendizaje máquina en el campo de la neurociencia, se ha ido dando paso a sistemas de análisis

²Decimos que varias fuentes de datos son heterogéneas cuando proceden de fuentes de diversa naturaleza. Por ejemplo, análisis de sangre para determinar los niveles de ciertas proteínas, o la intensidad de la captación de un radioligando mediante imagen cerebral funcional.

automático cada vez más precisos, con mayor capacidad de generalización, e incluso capaces (en algunos casos) de predecir la evolución de sujetos en etapas presintomáticas de su enfermedad. Todo ello contribuye a un mayor entendimiento sobre el funcionamiento del cerebro humano y sus dolencias. Junto al diagrama general de un modelo para la evaluación (diagnóstico) y seguimiento (pronóstico) de pacientes con algún trastorno neurodegenerativo, se desarrollan los bloques que componen este sistema y se revisan algunos de los modelos matemáticos que más se emplean en la actualidad para evaluar tanto imágenes cerebrales como datos procedentes de cualquier otro tipo de pruebas clínicas convencionales.

Una vez contextualizado el análisis de datos con inteligencia artificial en el campo de la neuroimagen, se propone una revisión sobre los antecedentes de los trabajos desarrollados en los capítulos posteriores, con especial atención al uso potencial de las distribuciones α -estables (Capítulo 4), la extracción de características morfológicas basadas en isosuperficies (Capítulo 5) y el análisis de marcadores heterogéneos en el estudio de la EA (Capítulo 6) y de la EP (Capítulo 7).

- **Capítulo 3.** La primera parte de este bloque incluye una descripción sobre los repositorios que se han analizado en esta tesis doctoral junto a un resumen sobre los aspectos más importantes acerca de las instituciones que han dado soporte y acceso a cada una de estas bases de datos. La segunda parte del capítulo trata de introducir cómo se han evaluado los modelos de clasificación propuestos en los siguientes capítulos incluyendo tanto la definición de métricas de clasificación como el desarrollo de algunos algoritmos de evaluación.
- **Capítulo 4.** Cualquier tipo de análisis de datos con inteligencia artificial comienza siempre con el preprocesado de los datos de entrada. Esta afirmación también se aplica en trabajos con imágenes cerebrales. Es por ello que, como puede observarse en la Figura 1.1 según el tipo de análisis que se quiere realizar, esta etapa de preprocesado puede incluir entre otros procedimientos una normalización de cada imagen de referencia a un espacio, tamaño e intensidad comunes (facilitando su posterior comparación con otras imágenes); o una etapa de segmentación de los tejidos cerebrales –si trabajamos con imágenes estructurales– para evaluar el grado de afectación neurológica (atrofia) que puede estar provocando la demencia.
En este contexto, la introducción del uso de distribuciones de probabilidad α -estables en neuroimagen constituye una mejora significativa en el análisis de imágenes cerebrales, tanto funcionales como estructurales, gracias al desarrollo de métodos cada vez más eficientes para el preprocesado de este tipo de datos. Basándonos en esta idea, en el

Capítulo 4 se justifica el desarrollo de una metodología para el modelado de mezclas de distribuciones α -estables; y expone dos escenarios en los que la aplicación de este tipo de distribuciones mejora los resultados de los procedimientos existentes. Por un lado se presenta un método de segmentación de imágenes estructurales combinando su ajuste de intensidad con información espacial. Y por otro lado, se explica cómo la normalización en intensidad con distribuciones α -estables en imagen funcional contribuye a una mejor diferenciación entre sujetos sanos y pacientes con EP.

- **Capítulo 5.** Uno de los objetivos del análisis de imágenes médicas es proporcionar nuevos marcadores que permitan caracterizar mejor una determinada enfermedad o proceso biológico. Dentro del procesado de imágenes, la caracterización morfológica de algunos de sus patrones ha permitido medir cuantitativamente la distribución de transportadores dopaminérgicos en el cerebro dando paso así al desarrollo de nuevos sistemas automáticos de ayuda al diagnóstico para la EP. A raíz de esta idea, en el Capítulo 5 se propone la inclusión de una etapa de extracción de características a partir de isosuperficies para modelar un sistema de ayuda al diagnóstico para la EP. Este paso, representado en la Figura 1.1 como una etapa dentro del preprocesado de imágenes debido a que las características que se generan en este bloque no dependen en ningún caso de las etiquetas de los sujetos de entrada, supone una mejora respecto a modelos anteriores por su rendimiento de clasificación y por la interpretación física de sus resultados.
- **Capítulo 6.** El desarrollo de un sistema de ayuda al diagnóstico asistido por computador con ML requiere el entrenamiento de un modelo de clasificación utilizando datos procedentes tanto de sujetos sanos como de sujetos con la enfermedad que se pretende diagnosticar. Siguiendo con el esquema de la Figura 1.1, en este capítulo se desarrolla una aplicación práctica para el análisis de datos heterogéneos relacionados con la EA de variante hereditaria (también conocida como DIAD por sus siglas en inglés, *Dominantly Inherited Alzheimer's Disease*). Este trabajo no solo constituye el primer modelo histórico de la enfermedad con técnicas de ML, sino que además, incorpora un estudio sobre la evolución de los pacientes con DIAD en etapas presintomáticas de su enfermedad. Todo esto pone en relieve la capacidad de los sistemas de aprendizaje automático para identificar patrones difícilmente perceptibles desde el punto de vista de la observación clínica convencional.
- **Capítulo 7.** El auge de las nuevas bases de datos clínicas, que incluyen marcadores procedentes de diferentes tipos de pruebas, ha dado paso al desarrollo de modelos de aprendizaje capaces de combinar el aprendizaje de cada fuente de información de entrada por separado

para generar un modelo más preciso de la enfermedad que se pretende caracterizar. A este tipo de sistemas se les suele conocer como modelos de aprendizaje de conjuntos o *ensemble* por su traducción al inglés. En este capítulo se ha desarrollado una nueva metodología de *ensemble* capaz de combinar el aprendizaje de varios tipos de datos relacionados con la EP de forma que el sistema es capaz de asignar de forma no-lineal un peso mayor a aquellas fuentes de información que más ayudan a diferenciar la enfermedad, y de penalizar el resto de fuentes (menos significativas o explicativas).

- **Capítulo 8.** Finalmente se ofrece una discusión general de los trabajos presentados para el desarrollo de la tesis, así como una relación de conclusiones y líneas de trabajo futuro más interesantes relativas a cada propuesta.

Diagrama general

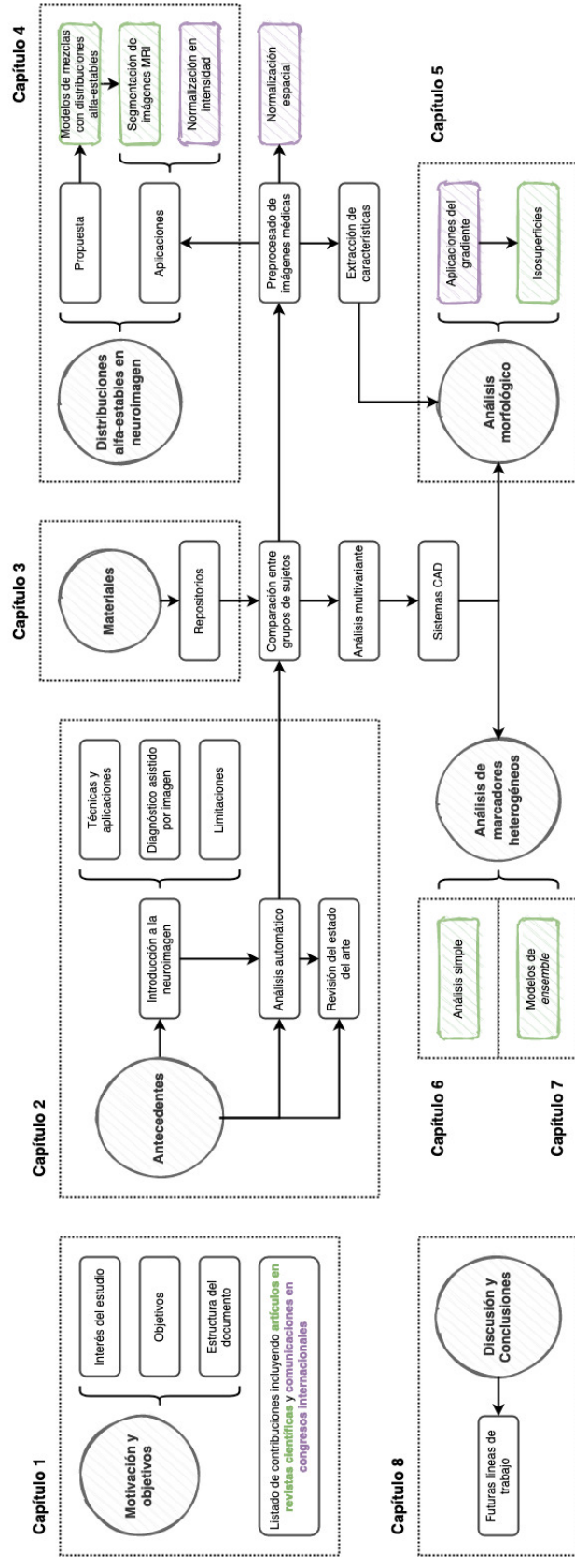


Figura 1.1: Diagrama general de la tesis doctoral.

1.4. Listado de contribuciones científicas asociadas a la tesis doctoral

Los contenidos de esta tesis doctoral se encuentran publicados en revistas internacionales de alto impacto y/o presentados en congresos internacionales. La siguiente tabla recoge la relación de contribuciones y los medios en los que se han presentado estos trabajos:

| Título | Medio | Cita |
|--|--|------|
| A Heavy Tailed Expectation Maximization Hidden Markov Random Field Model with Applications to Segmentation of MRI | Frontiers in Neuroinformatics (Q1) | [51] |
| On a Heavy-Tailed Intensity Normalization of the Parkinson's Progression Markers Initiative Brain Database | IWINAC 2017 | [52] |
| Classification Improvement for Parkinson's Disease Diagnosis Using the Gradient Magnitude in DaTSCAN SPECT Images | SOCO 2018 | [53] |
| Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson's Disease | Frontiers in Neuroinformatics (Q1) | [54] |
| Analysis of I[123]-Ioflupane SPECT intensity isosurfaces to assist the diagnosis of Parkinsonism | NSS/MIC 2018 | [55] |
| Ensemble classification of heterogeneous biomarkers in the diagnosis of Parkinsonism | NSS/MIC 2018 | [56] |
| Comparison Between Affine and Non-affine Transformations Applied to I[123]-FP-CIT SPECT Images Used for Parkinson's Disease Diagnosis | IWINAC 2019 | [57] |
| Autosomal Dominantly Inherited Alzheimer Disease: Analysis of genetic subgroups by Machine Learning | Information Fusion (Q1) | [58] |
| Morphological Characterization of Functional Brain Imaging by Isosurface Analysis in Parkinson's Disease | International Journal of Neural Systems (Q1) | [59] |
| Expectation-Maximization algorithm for finite mixture of α -stable distributions | Neurocomputing (Q1) | [60] |

Tabla 1.1: Trabajos publicados como primer autor durante la realización de la tesis doctoral.

Bloque B

Introducción a la neuroimagen y antecedentes

Capítulo 2

Estado del arte

Este capítulo tiene como objetivo fundamental desarrollar una explicación sobre las modalidades de imagen médica más utilizadas a la hora de evaluar la demencia; exponer en qué consisten las enfermedades de Alzheimer y de Parkinson; e introducir una descripción, lo más detallada posible, acerca del esquema general de un modelo de aprendizaje máquina para la evaluación de estas patologías. Todo ello nos permitirá comprender, en un segundo bloque, el impacto que ha tenido en las últimas décadas el desarrollo del Machine Learning y su aplicación práctica a la neurociencia, conformando así los antecedentes de las propuestas presentadas en esta tesis doctoral.

2.1. Técnicas de imagen médica

El uso de la imagen médica en el campo de la neurología, y su aplicación al análisis del sistema nervioso central, nos proporciona los medios necesarios para poder entender la estructura anatómica del cerebro humano y analizar su función y actividad metabólica.

Aunque existen muchos otros tipos de pruebas de imagen, como es el caso de la resonancia magnética funcional (*functional Magnetic Resonance Imaging*, fMRI) que permite evaluar a nivel de imagen qué regiones cerebrales están activas durante la realización de una acción concreta¹, o como es el caso de la resonancia magnética ponderada por difusión (*Diffusion-Weighted Magnetic Resonance Imaging*) usada para el mapeo de la tractografía nerviosa [61–63]; en esta sección solo nos centraremos en el desarrollo de los aspectos técnicos más relevantes sobre el proceso de adquisición de los tres tipos de imagen médica más frecuentemente utilizados para la evaluación de la demencia.

¹Por ejemplo, al intentar mover una extremidad o durante un ejercicio de articulación de la voz para determinar qué áreas están involucradas en estas acciones.

2.1.1. Imagen de resonancia magnética (MRI)

La imagen de resonancia magnética (*Magnetic Resonance Imaging*, MRI) es un tipo de prueba de imagen no ionizante que no emplea ningún tipo de radiación y que es considerada como una prueba segura en comparación con otras técnicas como los rayos X o como la tomografía axial computarizada (*computerized tomography*) [64]. Para la adquisición de esta imagen, se utiliza un campo magnético constante de alta potencia que excita los núcleos atómicos de hidrógeno² alineando los momentos magnéticos de sus núcleos según su espín inicial y de forma que los protones precesan³ en torno a un eje situado en la dirección del campo magnético a una frecuencia determinada según la expresión (2.1) y que es conocida como frecuencia de Larmor. En esta expresión, B_0 hace referencia a la potencia del campo magnético y γ_z es un parámetro que depende del núcleo atómico (en el caso del hidrógeno, $\gamma_z = 42.6$ MHz/T).

$$f_{\text{Larmor}} = \frac{\gamma_z}{2\pi} B_0 \quad (2.1)$$

Al aplicar un fuerte campo de radiofrecuencia a la frecuencia f_{Larmor} , la energía del pulso es absorbida por los núcleos atómicos excitando sus espines nucleares a estados de energía más altos. Cuando se interrumpe esta sobreexcitación, los núcleos vuelven a su estado de equilibrio (relajación) emitiendo una parte de la energía que ahora sí puede ser captada por un receptor de radiofrecuencia. Durante esta fase de relajación se establecen dos intervalos de tiempo⁴ en el que los átomos se alinean respecto al plano longitudinal o respecto al plano transversal creando dos imágenes. La combinación de ambas (mediante su ponderación) nos permite diferenciar la materia blanca (*White Matter*, WM) de la materia gris (*Grey Matter*, GM) si se asigna un peso mayor al resultado del primer intervalo, o el CSF del resto del cerebro para el segundo.

2.1.2. Tomografía Computarizada por Emisión de Fotón Único (SPECT)

En el caso de la prueba de Tomografía Computarizada por Emisión de Fotón Único (*Single-Photon Emission Computed Tomography*, SPECT), a cada paciente se le inyecta un radiofármaco –con cierta afinidad hacia algún tipo de molécula transportadora (radioligando)– cuya emisión natural de

²Se puede buscar la resonancia de cualquier otro tipo de átomo pero los de hidrógeno son más abundantes y consiguen una mejor relación señal a ruido (*Signal to Noise Ratio*, SNR).

³Movimiento de rotación del eje inclinado de giro de un giróscopo.

⁴Realmente hay tres intervalos pero el intervalo final que se usa más para evaluar el volumen sanguíneo y conseguir un ajuste más fino del líquido cefalorraquídeo (*CerebroSpinal Fluid*, CSF) así que suele englobarse dentro del segundo intervalo.

fotones gamma puede ser captada con una cámara receptora que va girando alrededor del sujeto tomando varias adquisiciones en diferentes planos y ángulos [65]. A partir de estas capturas⁵, y en cada plano de interés, se realiza una reconstrucción espacial para generar una única matriz bidimensional (conocida como sinograma) empleando el procedimiento descrito por **Radon** en [66] derivado del teorema de Fourier.

La imagen SPECT final muestra una distribución de intensidad que será mayor en aquellas regiones donde la concentración del radioligando sea más alta, permitiendo así cuantificar la distribución del biomarcador objetivo.

2.1.3. Tomografía por Emisión de Positrones (PET)

Las imágenes de Tomografía por Emisión de Positrones (*Positron Emission Tomography*, PET) tienen como origen la detección del par de fotones gamma de 511keV (cada uno) resultantes de la aniquilación de un positrón y un electrón (partícula y antipartícula) en un ciclotrón [65, 67]. Este par de fotones se genera en sentidos opuestos y su medición requiere que ambos detecten de manera simultánea por dos detectores de centelleo. Estos detectores, junto a sus correspondientes tubos fotomultiplicadores se colocan frente a la fuente de emisión de positrones de forma que las señales captadas por estos pasan una serie de amplificadores y circuitos que llevan al evento de coincidencia capaz de detectar la aniquilación en la línea de los detectores⁶.

La adquisición de las imágenes puede llevarse a cabo siguiendo varios métodos aunque el más común es la composición bidimensional descrita anteriormente para SPECT. De hecho, y de manera similar a como ocurre en SPECT, la distribución de intensidad de una imagen PET muestra focos de intensidad mayores en aquellas regiones con mayor concentración de aniquilaciones positrón/electrón ligadas a una mayor presencia de radioligando inyectado.

2.2. Aplicaciones clínicas de la neuroimagen

Los avances alcanzados en las últimas décadas en el campo de la neuroimagen, que se podría definir como la disciplina que estudia el sistema nervioso a través de la obtención y análisis de imágenes, han permitido profundizar la interpretación que tenemos sobre la mente humana y entender muchos de los mecanismos involucrados en el desarrollo de sus dolencias.

Tradicionalmente, el uso de imágenes estructurales y funcionales, e incluso su combinación, han servido como herramienta de apoyo al diagnóstico

⁵Normalmente separadas con una resolución angular de 2 grados.

⁶En realidad se utilizan múltiples detectores en anillo para captar la mayor cantidad posible de pares de fotones.

de diferentes patologías. En concreto, la imagen MRI se suele utilizar en el análisis de lesiones intracraneales (traumatismos craneoencefálicos, daños cerebrales por hipoxia, ictus, lesiones neuronales asociadas a la esclerosis múltiple o aneurismas entre otros) [68–72], el diagnóstico de trastornos neurodegenerativos como el Huntington [73, 74], o para determinar la existencia de tumores cerebrales (incluyendo astrocitomas, glioblastomas, gliomas, meningiomas, metástasis cerebrales, etc.) [23, 24, 75]. Por otro lado, las imágenes de tipo funcional SPECT y PET son ampliamente utilizadas en el estudio de procesos metabólicos del cerebro. Estas herramientas resultan fundamentales a la hora de estudiar patologías como la EP o la EA, aunque en las últimas décadas también han demostrado su utilidad en campos como el de la psicología cognitiva, la psiquiatría e incluso en el estudio del comportamiento [76–78].

2.2.1. Evaluación clínica de la Enfermedad de Alzheimer

La EA está descrita en [79] como un trastorno neurodegenerativo progresivo e irreversible, está caracterizada por un deterioro selectivo de la memoria. Es la causa más frecuente de demencia a nivel mundial y, si bien se dispone de tratamientos que pueden mejorar algunos de sus síntomas, no existe en la actualidad ninguna cura o terapia que permita alterar el curso de la enfermedad. Entre los síntomas más frecuentes se describen: pérdida de autonomía, deterioro visuoespacial, problemas de lenguaje y síntomas conductuales [80–82].

A nivel de imagen, la evaluación de pacientes con sospecha de Alzheimer se lleva a cabo preferiblemente con una prueba de resonancia magnética [83]. Los hallazgos radiológicos más frecuentes (Figura 2.1) incluyen atrofia generalizada y/o focal, lesiones en la WM, reducción del volumen del hipocampo⁷ y atrofia del lóbulo temporal medio [84–87]. La imagen funcional PET con el radioligando F18-Fluorodesoxiglucosa, más comúnmente llamada 18F-FDG PET, permite localizar aquellas regiones del cerebro que presentan hipometabolismo mientras que su variante SPECT mide las zonas con hipoperfusión [88–90]. Estas áreas se localizan principalmente en el precúneo y en los córtex parietal lateral y temporal posterior. Los trazadores del beta amiloide ($A\beta$) F18-florbetapir, F18-flutemetamol, F18-florbetaben y Pittsburg Compound B (11C-PiB); también tienen un uso muy extendido en el estudio de la EA debido su capacidad de unión a las lesiones de amiloide en el cerebro [91–94]. El principal inconveniente de este tipo de pruebas es que, en muchos casos, muestran más el proceso de envejecimiento natural de los pacientes que una medida de la gravedad de su enfermedad [95]. Con la reciente inclusión de nuevos radioligandos, como es el caso de los trazadores de proteína τ , se espera incrementar la tasa de acierto del diagnóstico por

⁷También influenciado por la edad.

imagen y se puedan identificar nuevos patrones para el diagnóstico precoz de la enfermedad [96, 97].

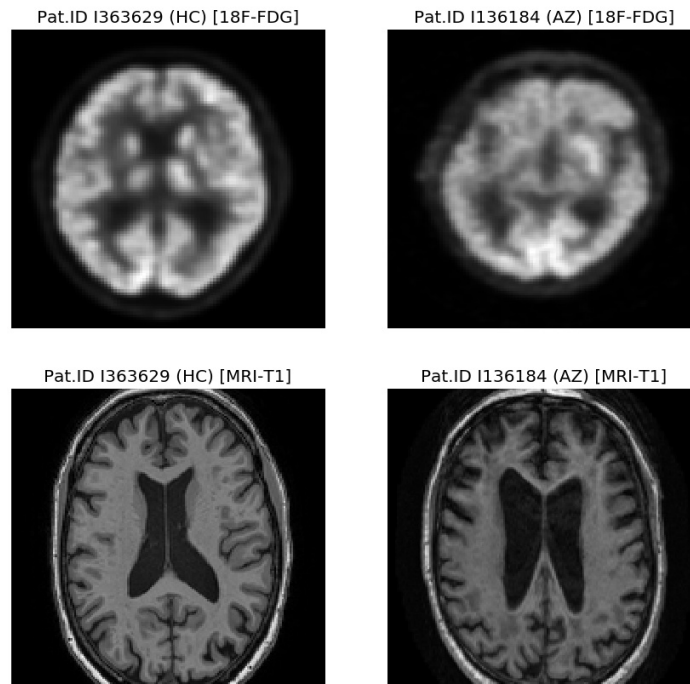


Figura 2.1: Vista axial de un sujeto sano (etiquetado como HC) y un sujeto con EA avanzada (etiquetado como AZ) de la base de datos ADNI. Imagen funcional 18F-FDG PET (fila superior) e imagen estructural MRI-1T (fila inferior).

Una de las líneas de investigación sobre la EA más importantes en la actualidad está centrada en el estudio de su variante hereditaria (DIAD) debida a una mutación en el gen que regula la proteína precursora de amiloide (*Amyloid Precursor Protein*, APP) [98], en el gen de la Presenilina-1 (PSEN1) [99] o en el gen de la Presenilina-2 (PSEN2) [100]. Aunque variante de la enfermedad solo representa el 1% del total de casos de EA a nivel mundial, tiene un gran impacto en la investigación de la demencia debido a que los individuos que presentan alguna mutación en cualquiera de estos tres genes siempre terminará desarrollando la enfermedad. Este hecho permite estudiar la evolución de los pacientes con alguna de estas mutaciones incluso en etapas presintomáticas de la demencia [101, 102].

Aunque la enfermedad es bastante similar a la EA común (*Late Onset Alzheimer's Disease*, LOAD) en cuanto a presentación clínica [103–106], existen algunas diferencias relacionadas con la edad de los sujetos en el debut de su enfermedad⁸, la propia condición hereditaria e incluso el desarrollo de

⁸En el caso de LOAD los primeros síntomas suelen presentarse en adultos > 65 años,

patologías asociadas [107]. Pese a estas diferencias, la comunidad científica espera que las conclusiones que se puedan extraer sobre la variante hereditaria puedan extrapolarse a la variante común, lo que abriría las puertas al desarrollo de nuevas terapias que puedan ralentizar o incluso prevenir el proceso neurodegenerativo [42, 82, 108–111].

2.2.2. Evaluación clínica de la Enfermedad de Parkinson

Por su parte, la EP, definida también como un trastorno neurodegenerativo crónico e irreversible, tiene como causa la muerte progresiva de las neuronas dopaminérgicas en la vía nigroestriada por alteraciones que aún se desconocen [112, 113]. Los síntomas más frecuentes de la EP son temblor, rigidez muscular, bradiquinesia y pérdida de autonomía; pero también existen alteraciones cognitivas que se acentúan con el tiempo [114].

Para apoyar el diagnóstico de la EP, una de las herramientas más utilizadas en la práctica clínica es el uso de imágenes I^[123]-Ioflupano de tipo SPECT [115]. Estas imágenes, conocidas en muchos contextos como FP-CIT o DaTSCAN (nombre comercial), hacen uso de un radioligando con una fuerte afinidad por los transportadores presinápticos de dopamina (*presynaptic Dopamine Transporters*, DATs) [116, 117]. Como los sujetos con EP manifiestan una reducida captación de los DATs debido a la falta de neuronas dopaminérgicas y terminales, las imágenes de estos pacientes se caracterizan por un patrón de captación tenue y de forma redondeada mientras que los sujetos sanos presentan patrones en forma de letra “c” (núcleo caudado y putamen) mucho más iluminadas y siempre simétrico [118]. Este comportamiento en imagen sirve no solo para ayudar al diagnóstico diferencial de la enfermedad sino que también permite estimar la evolución longitudinal de los pacientes [119, 120].

Como ejemplo, en la Figura 2.2 se han representado dos cortes axiales de dos sujetos: uno sano (izquierda) y otro con diagnóstico de EP (derecha). Destaca la forma redondeada y pequeña de la región de captación que se ha obtenido en el paciente con EP. En etapas avanzadas de la enfermedad, el bajo nivel de captación en la zona del estriado lo puede hacer prácticamente indistinguible del resto del cerebro.

2.3. Limitaciones de la neuroimagen clínica convencional

Uno de los problemas de la neuroimagen clínica convencional es la necesidad de disponer de uno o varios evaluadores con experiencia. Incluso contando con estos, el riesgo de error en el diagnóstico manual puede ser alto. Por poner un ejemplo, en la revisión sistemática y meta-análisis llevados

mientras que los casos de DIAD empiezan a mostrar síntomas entre los 45 – 65 años.

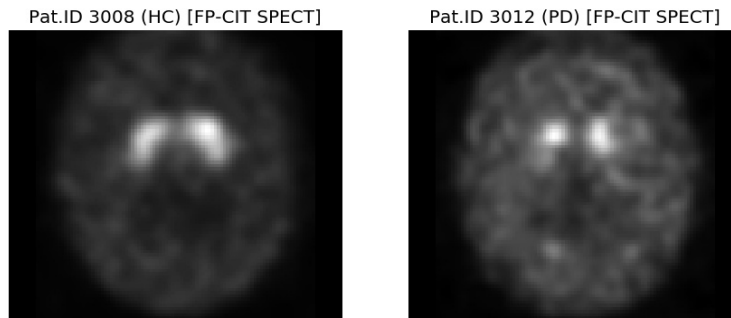


Figura 2.2: Vista axial de dos imágenes FP-CIT SPECT de la base de datos PPMI correspondientes a un sujeto de control (izquierda) y un paciente con EP (derecha).

a cabo en 2016 por **Rizzo et al.** [121], se determinó que la precisión en el diagnóstico de la EP durante los últimos 25 años se situaba en torno al 80%⁹. En las conclusiones de este trabajo, también compartidas en [122, 123], se manifiesta que aunque son las etapas tempranas de la enfermedad (así como sus variantes atípicas) las más limitan la capacidad de diagnóstico, sería interesante incorporar nuevas herramientas de apoyo al diagnóstico que permitan alcanzar resultados más precisos incluso en escenarios con personal poco entrenado.

2.4. Modelos de análisis automático para neuroimagen

2.4.1. Introducción

Desde el punto de vista de la investigación, el uso de la neuroimagen siempre ha estado justificado por la búsqueda de patrones estructurales y/o funcionales, en la forma de marcadores de imagen, capaces de caracterizar una gran variedad de procesos biológicos en el cerebro y que pueden ayudar al diagnóstico temprano de ciertas enfermedades así como a la medida cuantitativa de algunas de sus propiedades [124, 125].

La automatización del análisis de imágenes médicas permite reducir los costes tanto de personal experto como de tiempo asociado al proceso diagnóstico, a la vez que ayuda a localizar patrones difícilmente cuantificables a simple vista, dando lugar a diagnósticos más precisos [126]. En este sentido, la aparición de modelos basados en el uso de estadística avanzada y/o aprendizaje máquina (*Machine Learning*, ML), ha conseguido en los

⁹Promedio de resultados obtenidos por personal experto entrenado y personal con poca experiencia. Estos resultados se deben, en gran parte, a que la EP en realidad comprende un espectro de patologías conocido como “Síndromes Parkinsonianos”.

últimos años impulsar nuestro conocimiento sobre el cerebro humano a la vez que está permitiendo desarrollar nuevas herramientas para su uso como apoyo al diagnóstico, o para predecir la evolución de diferentes procesos biológicos. En función del problema que pretenden resolver, tenemos:

- **Modelos de segmentación.** El problema de segmentación tiene como objetivo fundamental delimitar las regiones de WM, GM, CSF y cráneo en una imagen estructural; aunque también se puede hablar de segmentación a la hora de delimitar lesiones neurológicas como en el caso de la esclerosis múltiple [127–130], o la existencia de masas tumorales en el sistema nervioso central [131, 132]. Este problema afecta a imágenes individuales y se encuentra limitado por factores como la resolución de la imagen original, el nivel de ruido durante la adquisición de la prueba o la corrección del movimiento del sujeto, entre otros.
- **Modelos de normalización.** Cuando se quieren comparar varios sujetos entre sí, es necesario realizar algún tipo de transformación (deformación) que permita hacer comparables estas pruebas, ya sea a nivel espacial, permitiendo la coincidencia de las diferentes estructuras cerebrales; como a nivel intensidad (o rangos de intensidades), sobre todo para pruebas que no siempre conservan las mismas condiciones iniciales (diferentes modelos de adquisición de imágenes, cantidad de radioligando que se le inyecta a cada paciente, tasa de absorción del radioligando, calibración de los equipos de adquisición, etc.) [133, 134].
- **Modelos de extracción de características.** Ya sea mediante la aplicación de alguna transformación matemática o simplemente mediante la selección de algunas de las características discriminativas de las pruebas de imagen de entrada, tienen como objetivo fundamental la caracterización de los sujetos de entrada a partir de un conjunto de variables resultado [135, 136].
- **Modelos de clasificación.** Tienen como objetivo fundamental el desarrollo de un modelo de aprendizaje (a ser posible automático) que permita identificar patrones característicos de una enfermedad o proceso biológico de interés, a partir de los datos de entrada de un conjunto representativo de sujetos pertenecientes a dos o más clases. Una vez el modelo de clasificación ha sido entrenado, este puede predecir a qué clase (diagnóstico) pertenecerá un nuevo sujeto (desconocido) cuyas características sean evaluadas por el sistema [5, 7, 137–139].
- **Modelos de regresión.** En este caso el modelo de aprendizaje (a ser posible automático) tratará de explicar la evolución (normalmente en términos de tiempo) de un conjunto de sujetos de entrada y predecir el curso que tendría cualquier nuevo sujeto (desconocido) a través de sus características en un instante de tiempo determinado [43, 140].

Aunque se podría haber hecho esta división atendiendo a un criterio diferente, cabe destacar en este tipo de escenarios un detalle relacionado con la cantidad mínima de muestras (sujetos) que son necesarias para cada tipo de análisis. Así, mientras que la mayoría de propuestas de segmentación de imágenes sólo requieren el análisis sujeto a sujeto de imágenes estructurales, los modelos que hacen uso de técnicas de inteligencia artificial para el diagnóstico asistido por computador así como los modelos de evolución longitudinal, precisan de una gran cantidad de datos de entrada capaces de explicar el proceso biológico de interés.

Los siguientes apartados resumen las etapas más importantes del análisis de datos de neuroimagen con ML (Sección 2.4.2), e introducen una visión general sobre los procedimientos más frecuentemente utilizados en la actualidad para cada bloque en problemas tanto de clasificación como de regresión (Secciones 2.4.3 hasta 2.4.6).

2.4.2. Esquema general de un modelo de clasificación/regresión con aprendizaje máquina

En el campo de la neuroimagen, la mayor parte de modelos de aprendizaje automático (fundamentalmente modelos de clasificación o de regresión) están basados en el uso de esquemas de aprendizaje supervisado como el representado en la Figura 2.3.

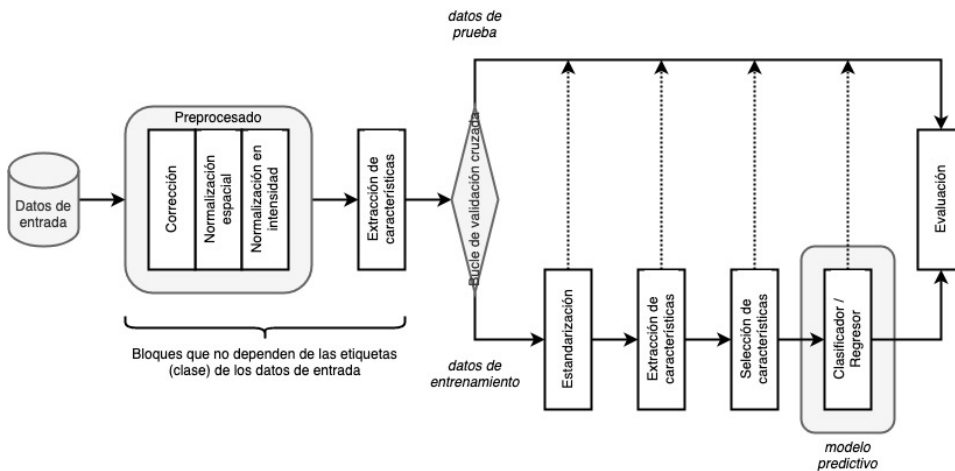


Figura 2.3: Diagrama general de un problema de clasificación/regresión con ML.

Si se tienen en cuenta solo datos procedentes de imágenes cerebrales¹⁰, cada bloque que conforma el sistema completo cumple con la siguiente función:

¹⁰Normalmente, las tablas con columnas de datos procedentes de pruebas clínicas requieren las mismas o incluso menos etapas.

- **Etapa de Preprocesado.** Hace referencia al acondicionamiento de los datos de entrada necesarios para extracción de algunas de sus características. A nivel de imagen, el preprocesado desempeña un papel fundamental en la detección de enfermedades cerebrales porque mejora la calidad de la imagen mediante procedimientos como la corrección movimiento (imágenes estructurales), el suavizado o la corrección de ruido [124]. Pero también permite que las muestras de entrada puedan ser comparables entre sí en términos de posición y tamaño (normalización espacial); intensidad (normalización en intensidad); e incluso de distribución de probabilidad de alguna de sus propiedades.

Durante esta etapa también pueden incluirse métodos de extracción de características siempre y cuando estos no dependan de las etiquetas de los datos de entrada, evitando así el posible sesgo que puede surgir al validar un modelo con datos derivados de muestras ya conocidas.

- **Bucle de validación cruzada (*cross-validation loop*).** A partir del cual se pueden generar métricas con las que determinar la capacidad del modelo para predecir las etiquetas de datos que todavía no se hayan utilizado durante la estimación del mismo [141–143]. Mediante este bucle, los datos de entrada se separan en varios subconjuntos:

- Datos de entrenamiento (*training*): conjunto de datos que permiten ajustar el modelo de aprendizaje.
- Datos de validación (*validation*): conjunto de datos que permiten obtener una estimación –inicialmente no sesgada– del ajuste del modelo a los datos de entrenamiento durante el ajuste de sus hiperparámetros. Su uso es más frecuente en escenarios con redes neuronales.
- Datos de prueba (*test*): conjunto de datos con los que se obtiene una evaluación no sesgada del ajuste del modelo final a los datos de entrenamiento.

- **Estandarización.** Para que todas las variables de entrada puedan ser comparables entre sí. Se refiere al reescalado de los datos mediante (2.2) y/o su normalización.

$$\hat{x} = \frac{x - \mu_x}{\sigma_x} \quad (2.2)$$

- **Extracción de características.** Que, a diferencia de las que se extraen fuera del bucle de validación cruzada, sí pueden depender de las etiquetas de los datos de entrenamiento. En esta etapa se incluyen los métodos de reducción de la dimensionalidad como es el caso del análisis de componentes principales (*Principal Component Analysis*, PCA)

[144, 145] o el algoritmo de los mínimos cuadrados parciales (*Partial Least Squares*, PLS) [146].

- **Selección de características.** Si el número de características de entrada de un clasificador es demasiado elevado en comparación con el número de muestras de entrada, se podría incurrir en un sobreajuste del modelo de predicción a los datos de entrenamiento. Para evitar esto, normalmente se utiliza una bloque de selección de características que asigna una puntuación (*score*) a cada característica de entrada y selecciona aquellas características con mejor puntuación [147]. Se suelen utilizar tests de hipótesis nula como por ejemplo t-Test, el test de Welch, ANOVA o el test de Mann-Whitney Wilcoxon [148–152].
- **Clasificador/Regresor.** En referencia al aprendizaje supervisado, se trata de una función matemática ($f(\cdot)$) que permite caracterizar una variable de salida, \mathbf{Y} , en función de un conjunto de variables de entrada \mathbf{X} según la asignación $\mathbf{Y} = f(\mathbf{X})$.

Si la variable de salida \mathbf{Y} constituye una variable categórica con un número finito de valores posibles (clases), entonces a la función de mapeo resultante se le denominará clasificador; mientras que si el rango de valores de \mathbf{Y} es continuo, al modelo resultante se le denomina regresor.

- **Evaluación.** En cada iteración del bucle de validación cruzada, el clasificador/regresor ajustado con los datos de entrenamiento se utiliza para estimar las etiquetas de los datos de prueba. Si se comparan estos valores estimados con los resultados reales de los datos de entrenamiento, se puede cuantificar el grado de precisión del modelo entrenado o, dicho de otra forma, su capacidad para predecir \mathbf{Y} [153].

2.4.3. Preprocesado de imágenes cerebrales

La cantidad de bloques que conforman la etapa de preprocesado en escenarios de análisis de neuroimagen con inteligencia artificial, está sujeta a las necesidades concretas que plantea cada problema. Aunque algunos bloques son fijos para cualquier tipo de análisis, como por ejemplo ocurre con la corrección de ruido, otros elementos como los relativos a la normalización o a la extracción de características dependerán de si se están comparando varias muestras entre sí, o de si los tipos de características de entrada ya disponibles son suficientes como para modelar el problema subyacente. Según este hecho, se pueden señalar los siguientes bloques de preprocesado como los más relevantes para el análisis de datos de neuroimagen:

Corrección de ruido

En comparación con el resto de técnicas de imagen médica, las pruebas MRI son particularmente sensibles a los efectos del ruido durante su adquisición [154, 155]. De acuerdo con este hecho, la explicación que se presenta en este apartado se centrará en comentar los aspectos más relevantes de la corrección de las fuentes de ruido en imagen estructural.

Durante la reconstrucción de una imagen MRI, normalmente se aplica una transformada de Fourier bidimensional que descompone en magnitud y fase el resultado de la prueba evaluada durante los intervalos de relajación de los núcleos atómicos sobreexcitados dando lugar a mapas que ayudan a diferenciar los diferentes tejidos cerebrales. Ya que en este proceso de adquisición pueden aparecer fuentes de ruido¹¹ que afectan de manera no simétrica al resultado de la magnitud y de la fase, su modelado y corrección constituye una de las primeras etapas del preprocesado de este tipo de imágenes [155]. En el trabajo publicado en [156], los autores afirman que la magnitud de la intensidad de una imagen MRI en presencia de fuentes de ruido resistivas e inductivas, puede modelarse a partir de distribuciones de Rice (cuando el valor de la SNR cumple cierto umbral) o de distribuciones gaussianas (en caso contrario). Sobre esta observación, justificada primero en equipos con una sola bobina pero cuya aplicación puede extenderse también a otros tipos de escáneres con más bobinas, se han ido desarrollando varias propuestas que tratan de corregir la reconstrucción de las imágenes estructurales mediante la eliminación, o al menos la atenuación, de los efectos que estas fuentes de ruido tienen sobre el resultado final del prueba aunque siempre limitadas por la necesidad de preservar el equilibrio entre la atenuación de la fuente y la preservación de las características de la imagen [157]. Entre las soluciones más frecuentemente citadas en la bibliografía actual, encontramos trabajos basados en técnicas de filtrado [158, 159]; en la aplicación de estimadores el como los de máxima verosimilitud (*Maximum Likelihood*) [160, 161]; transformaciones en el dominio de Wavelet [162, 163]; algoritmos de minimización del error cuadrático medio [164]; e incluso redes neuronales [165].

Por otra parte, y aunque la reconstrucción de imágenes MRI presupone que tanto el dispositivo de adquisición como el paciente al que se le está realizando la prueba están en completo reposo, esto en realidad nunca ocurre. Los movimientos que provocan la propia respiración, el pulso sanguíneo, el flujo convencional del CSF en el sistema nervioso central e incluso las pequeñas vibraciones mecánicas a las que están sometidos los detectores; hacen que la imagen se desenfoque limitando así su resolución y reduciendo la SNR de la prueba [154]. Aunque existen muchos algoritmos capaces de corregir este tipo de ruido –y que abarcan desde simples modelos estadísticos

¹¹Normalmente asociadas a la resistencia de las bobinas de los detectores o a las pérdidas inductivas en los mismos.

como el trabajo presentado en [166] hasta redes neuronales profundas (*Deep Learning*) como en [167]–, resulta prácticamente imposible determinar un único procedimiento estándar para la corrección del movimiento debido a que cada propuesta es capaz de modelar y corregir de manera eficaz solo un número limitado de efectos en la imagen final. Esto tiene como consecuencia que la elección del método de corrección quede sujeto, en muchos casos, a una evaluación previa (manual) de cada adquisición [168].

Normalización espacial

Partiendo de un conjunto de imágenes de entrada ya corregidas en cuanto a nivel de ruido (o cuyo nivel de ruido es lo suficientemente pequeño como para poder omitir su corrección) el siguiente paso en la etapa de pre-procesado será aplicar algún tipo de registro espacial a la imagen. Gracias a este registro, o normalización, los resultados de cada prueba clínica pueden compararse de forma directa con los de cualquier otra imagen o plantilla.

Aunque el uso de un enfoque de normalización afín –entendida como aquel tipo de normalización que solo emplea transformaciones como traslaciones, giros, reescalado y cortes– preserva las estructuras anatómicas del cerebro, es más común encontrar trabajos que aplican también transformaciones no-lineales por su mayor ajuste y por la posibilidad de comparar los vóxeles de cada imagen utilizando los mismos ejes de coordenadas y en la misma posición como en [13, 138, 169, 170]. No obstante, la cantidad de sesgo que introducen este tipo de soluciones puede alterar potencialmente el resultado final del diagnóstico, sobre todo en escenarios en los que la información sobre el estado de una enfermedad guarda algún tipo de relación con la morfología de una región cerebral como ocurre en el estudio de la EP referido a la región del estriado en imagen funcional FP-CIT SPECT [57].

Existen varias aplicaciones *software* que incorporan funciones y procedimientos para el análisis de secuencias de datos de imágenes cerebrales incluyendo la aplicación de diferentes métodos de normalización espacial. La más destacada de todas es la herramienta *Statistical Parametric Mapping (SPM)* disponible a través del enlace <https://www.fil.ion.ucl.ac.uk/spm/> [171]. En su última versión, *SPM12*, la normalización espacial de las imágenes estructurales se basa en el procedimiento de “segmentación unificada” –descrito en [25]– que combina un método de segmentación, una corrección de sesgo y una normalización espacial iterativa de los tres tejidos cerebrales (WM, GM y CSF) deformados mediante una combinación lineal con mil funciones de base de coseno y una mezcla de distribuciones gaussianas que modela la intensidad de cada tejido [172]. No obstante, y como la mayor parte de repositorios con imágenes funcionales no incluyen una imagen MRI asociada a cada adquisición que pueda utilizarse como referencia para calcular esta transformación espacial, lo que se suele hacer es minimizar la diferen-

cia media cuadrática entre una plantilla de referencia funcional y diferentes versiones deformadas de la imagen que se quiere normalizar como se explica en [173]. Aunque este es el método de referencia, existen alternativas como el método de registro anatómico difeomórfico mediante álgebra de Lie (más conocido como DARTEL) propuesto por **Ashburner** en [174]; o el método de normalización espacial descrito en [175] e implementado en la librería de análisis *Advanced Normalization Tools* (ANTs) [176, 177].

Normalización en intensidad

En imagen funcional, el resultado de una normalización en intensidad es una transformación de las características de intensidad de una imagen para producir un contraste de tejido diferente [178]. A la vista de esto, la normalización garantiza que las diferencias entre sujetos (o entre distintas captaciones realizadas sobre un mismo sujeto) se deben a razones fisiológicas (o de funcionamiento) del propio sujeto y no a factores tales como la calibración de la cámara gamma con la que se está realizando la adquisición de la imagen, la cantidad de radioligando inyectado en el paciente o su tasa de absorción [52, 179]. Al margen de estas ventajas, la normalización en intensidad constituye un paso clave del análisis de imágenes cerebrales con aprendizaje profundo (*Deep Learning*, DL) [178, 180].

Aunque existen varios procedimientos desarrollados para la normalización en intensidad de imágenes funcionales –como es el caso de la normalización tradicional al máximo de intensidad como la que se aplica en [169, 170]–, el más común de todos ellos es el método *Binding Ratio* (BR) [181]. Esta técnica está basada en el uso de una medida cuantitativa de la cantidad de radiotrazador en una región específica del cerebro, C_{VOI} , respecto a la que se obtiene en otra región no-específica, C_{N} según la relación descrita en (2.3):

$$BR = \frac{C_{\text{VOI}} - C_{\text{N}}}{C_{\text{N}}} = \frac{C_{\text{VOI}}}{C_{\text{N}}} - 1 \quad (2.3)$$

Pero aunque este procedimiento ha sido un estándar muy común de escenarios de análisis de imágenes cerebrales, presenta el inconveniente de estar sujeto a un alto nivel de subjetividad derivado de la propia definición de las regiones específicas y no-específicas de cada imagen, además de estar considerado como un método propenso a errores [134]. Por poner un ejemplo, si en el análisis de la EP mediante imágenes FP-CIT SPECT no existen grandes diferencias entre la densidad de transportadores presinápticos de dopamina en la región específica (estriado) y la no-específica (córtex occipital), entonces no se consigue un buen margen de separación entre clases. Este comportamiento implica que el método no siempre sea sensible a los cambios en la distribución de intensidad de cada región del cerebro, lo cual limita la capacidad de caracterización de las imágenes [182].

Extracción del cerebro (*skull stripping*)

Para obtener una imagen clara del cerebro sin elementos externos que puedan interferir en su posterior análisis, hace falta introducir un paso intermedio encargado de eliminar todo aquello que no pertenezca a un tejido del sistema nervioso central. La mayoría de procedimientos encargados de esta tarea se han apoyado tradicionalmente en el propio registro espacial de la imagen, en el uso de atlas anatómicos (como el *Automated Anatomical Labelling*, [183]) e incluso aplicando geometría como en [184–189]. Aunque también han aparecido soluciones basadas en ML como en [190], si alguna de las condiciones iniciales de los métodos de registro y/o geometría no se cumple, entonces lo que se recomienda es recurrir a técnicas más avanzadas como el uso de redes neuronales convolucionales (*Convolutional Neural Networks*, CNN) como en [191, 192], incluso en escenarios con imágenes no estructurales [193].

Segmentación de tejidos cerebrales

La segmentación de una imagen estructural es un proceso que permite parcelar (dividir) el cerebro en sus tres componentes principales: WM, GM y CSF. El análisis de estas partes, en especial de la GM y de la WM permite estudiar a mayor nivel de detalle neuropatologías como la Esclerosis Múltiple, la EP o la EA, y resulta clave en el estudio del proceso de envejecimiento del cerebro humano.

A nivel de clínica tradicional, la segmentación manual de imágenes MRI siempre se ha visto como un procedimiento tanto costoso (sobre todo desde el punto de vista de vista temporal) como impreciso (debido a que su resultado dependía en gran medida de la experiencia y conocimiento previo del evaluador responsable). Con el fin de vencer estas limitaciones, la comunidad científica ha ido desarrollando en las últimas décadas modelos de segmentación automática cada vez más eficientes e incluso con más aplicaciones que la propia evaluación del sistema nervioso central [194].

Históricamente, los primeros trabajos de segmentación automática de imágenes MRI consistían en establecer umbrales de intensidad o en delimitar los contornos de cada región mediante diferenciales de intensidad [195–197]. Aunque también aparecieron publicaciones que incorporaban operadores morfológicos, por ejemplo mediante el uso de técnicas de erosión y dilatación de regiones de intensidad [198], y trabajos que usaban regiones de crecimiento [199]; la mayoría de los sistemas basados en medidas directas de la intensidad fueron reemplazados por modelos de *clustering* [200, 201], mezclas de distribuciones de intensidad (histogramas) [25, 27] o cadenas de Markov para el cálculo de estas mezclas [202, 203].

Debido a que, cuando se trabaja a nivel de histograma, la segmentación de imágenes estructurales se ve alterada por efectos tales como variaciones en el contraste de los tejidos, o los efectos de sombra (*shading*) por el ruido de radiofrecuencia debido a las variaciones en el campo magnético durante la captura de la imagen; las propuestas de segmentación fueron mejorando y surgieron los primeros modelos con clasificadores estadísticos basados en el uso de algoritmos como el EM, los modelos de máxima verosimilitud (*Maximum Likelihood*) y los modelos con campos aleatorios de Markov (*Markov Random Fields*, MRF) [204–208]. Estos mejoran la mezcla de distribuciones, añaden información espacial y resultan (en general) en modelos que estiman con mejor nivel de precisión las regiones de materia blanca y materia gris del cerebro.

Al mismo tiempo que van surgiendo estas propuestas, aparecen diversos trabajos de segmentación basados en el uso de técnicas de aprendizaje máquina. Como ejemplos, se pueden encontrar publicaciones que hacen uso de algoritmos como *K-Nearest Neighbor* (KNN) [209, 210], árboles de decisión [211], máquinas de vectores de soporte (*Support Vector Machine*, SVM) [212] e incluso modelos de segmentación con *Random Forests* (RF) [213]. La importancia cada vez mayor de este tipo de sistemas hizo que pronto empezaran a aparecer las primeras propuestas basadas en el uso de redes neuronales. Los primeros modelos de segmentación usaban redes de tipo SOM (*Self-Organizing Map neural networks*, [214]) junto a vectores de entrada con resultados estadísticos relativos a la intensidad de cada voxel y de sus vecinos (información espacial) [215]. Al poco tiempo, estos modelos ya mejoraban las propuestas anteriores modificando las características de entrada, incluyendo nuevas variables en el estudio y/o modificando el esquema de clasificación total e incluso evaluando diferentes formas de hacer el mapeo de las redes [216–218].

Aunque en los últimos años han seguido apareciendo propuestas de segmentación usando *clustering* [219], modelos de mezclas de distribuciones gaussianas [220, 221], MRF [222–224], y RF [225, 226]; en la literatura actual se pueden encontrar decenas de trabajos relativos al uso de redes neuronales profundas. Entre las más utilizadas, destacan aquellas basadas en el uso de redes CNN como las propuestas *fully connected* presentadas en [227]; los modelos con ventana deslizante (*patch-based CNN*) descritos en [228–231]; redes con estructuras U-Net como en [232, 233] o el trabajo publicado por **Luna & Park**, ganadores del *Challenge MICCAI2018* [234]; e incluso esquemas con estructuras M-Net como el que se incluye en [235].

A nivel de librerías *software*, tanto el procedimiento descrito para *FreeSurfer* en [236] como el propuesto para *SPM* en [171], constituyen los dos modelos de referencia para la segmentación de imágenes estructurales en la actualidad. Tomando como referencia el trabajo publicado por **Palumbo et al.** en [237], estos procedimientos se resumen como sigue a continuación:

- En *FreeSurfer v6.0*, el proceso de segmentación abarca un total de 31 etapas entre las que se incluye: la corrección de la no-uniformidad de los valores de intensidad¹² [239]; el realineamiento espacial de la imagen [240]; un procedimiento de *skull stripping* para eliminar todos los elementos de cada imagen que no pertenezcan al sistema nervioso central (cráneo principalmente) [186]; una etapa en la que se aplica una normalización al espacio de un atlas en el que el valor de cada coordenada indica la probabilidad de pertenencia a cada tejido (WM o GM); la aplicación de un modelo de mezclas de distribuciones gaussianas con MRF y EM para modelar la relación espacial de las intensidades entre estructuras anatómicas; y un paso de parcelación en el que se identifican un total de 37 regiones anatómicas diferentes [241]. Aunque la herramienta también ofrece la posibilidad de aplicar correcciones manuales al resultado de segmentación final, no es frecuente encontrar trabajos que intentan aplicar este tipo de correcciones por la cantidad de recursos humanos y de tiempo que requiere este paso y porque las diferencias no suelen ser significativas [242].
- En el caso de *SPM12*, primero se realinea la imagen respecto a su comisura anterior (evitando así problemas que puedan surgir posteriormente con el registro espacial), y se aplica un modelo de mezclas de distribuciones gaussianas modificado que consta de tres pasos: 1) una corrección de la no-uniformidad de los valores de intensidad; 2) un registro espacial de la imagen al espacio de unos mapas de probabilidad a priori para cada tipo de tejido; y 3) la aplicación de un modelo de Bayes para calcular la probabilidad a posteriori para cada tipo de tejido pero derivada de la intensidad de la imagen de entrada. Este procedimiento iterativo (dentro del proceso de “segmentación unificada”, [25]), establece que son necesarias dos componentes gaussianas para caracterizar cada tipo de tejido. Cuando el sistema converge (porque ya no hay cambios significativos de la pertenencia de cada vóxel a cada tejido), se aplica una operación de modulación que ayuda a preservar la cantidad de puntos en cada tejido y a corregir los posibles defectos derivados de la normalización espacial (ajuste fino).

Parcelación o etiquetado de regiones cerebrales

El procedimiento de parcelación de una imagen estructural tiene como objetivo asignar una etiqueta anatómica a cada coordenada del cerebro. Normalmente este procedimiento se basa en un modelo de probabilidad estimado anteriormente a partir de un conjunto de datos etiquetados manualmente y

¹²Debido principalmente a la falta de homogeneidad espacial del campo magnético estático B_0 , lo que provoca variaciones en la intensidad de la imagen [238].

ajustados teniendo en cuenta la información geométrica del modelo de referencia que se trata de identificar [243, 244]. La aplicación de deformaciones espaciales y la identificación (con la mayor precisión posible) de todos los giros, surcos y regiones de interés (*Region Of Interest*, ROI) de la imagen de entrada son los encargados de conseguir un ajuste fino del mapa de etiquetado.

Aunque tradicionalmente la parcelación se aplica solo a imágenes estructurales, su uso puede extenderse también a otros tipos de imagen siempre y cuando estas se encuentren registradas al espacio original de la imagen estructural. Si es así, lo que se suele hacer es determinar el mapa de etiquetas de la imagen estructural y aplicar dicho mapa directamente a la imagen de interés. En la Figura 2.4 se han representado los cortes axial, sagital y coronal de la plantilla `labels_Neuromorphometrics.nii` de *SPM12*.

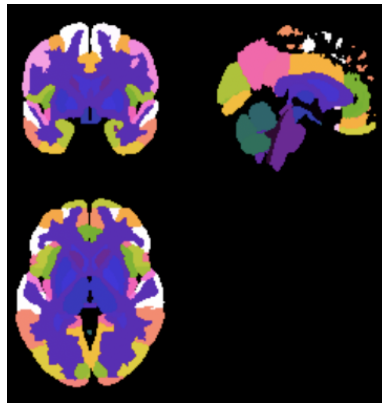


Figura 2.4: Plantilla `labels_Neuromorphometrics.nii` de *SPM12*. Ejemplo representativo con cortes axial, sagital y coronal donde cada color representa una etiqueta (región anatómica) diferente.

Métodos no estadísticos para la extracción de características

Durante el preprocesado de imágenes cerebrales (y por extensión de cualquier otro tipo de dato de entrada) puede incluirse algún bloque de extracción de características que permita caracterizar de manera más eficiente las posibles variables de entrada que posteriormente serán analizadas dentro del bucle de validación cruzada. La condición principal que deben cumplir todas las nuevas características que generen de este bloque es que su cálculo sea totalmente independiente de las etiquetas de los datos de entrada, evitando así el posible sesgo derivado del entrenamiento en el bucle de validación cruzada. Aunque en la bibliografía actual se pueden encontrar cientos de estas propuestas, aquí señalaremos solo algunas de las más citadas y en función del tipo de información que proporcionan:

- **Standardized Uptake Value Ratio (SUVR)** - En imagen funcional. Se trata de marcadores que miden el ratio entre la intensidad de una región del cerebro (región específica) y otra de referencia (región no-específica¹³) como los que se calculan para la normalización en intensidad con BR, pero referidos a cada región de interés. Este tipo de marcadores son muy comunes en bases de datos mundiales como ADNI o DIAN (ver Sección 3.1.3). Entre los trabajos que analizan este tipo de marcador para el diagnóstico automático de trastornos neurodegenerativos destacamos aquí la propuesta presentada por **Son et al.** en [246] donde se justifica el uso de una red CNN entrenada con valores SUVR extraídos de imágenes F18-florbetaben procedentes de sujetos de control, pacientes con Deterioro Cognitivo Leve (DCL) y pacientes con Alzheimer.
- **Cortical thickness** - Solo en imagen estructural. Trata de determinar el grosor del córtex neuronal en diferentes áreas del cerebro [247]. Entre los métodos más citados en la bibliografía consultada, destacan las propuestas implementadas en herramientas como *FreeSurfer*, como la metodología descrita en [248] donde los autores combinan varias etapas de preprocesado¹⁴ para identificar –mediante el uso de estructuras topológicas y deformaciones– las fronteras entre la materia blanca y la materia gris asociadas a cada región de interés; o los trabajos de desarrollados por **Fischl et al.** en [252, 253] que también están basados en una combinación de procedimientos similares a los empleados en [248], pero que tratan de delimitar esta frontera mediante la clasificación de sus valores de intensidad. Aunque este tipo de soluciones todavía forman parte del estándar actual en las últimas versiones de *FreeSurfer* y *Advanced Normalization Tools* [254], también se pueden citar propuestas más recientes que tratan de mejorar los métodos estándar gracias a un ajuste más preciso de ciertas regiones anatómicas de interés como en [255], o que mejoran alguno de los procedimientos internos que llevan al cómputo final del grosor cortical como en [256].
- **Análisis de texturas** - El análisis de texturas aplicado al estudio de la neuroimagen puede contribuir a identificar nuevas características de entrada, definir objetos o regiones de interés e incluso etiquetar patrones de intensidad cuyo análisis posterior puede conducir al desarrollo de nuevos modelos de clasificación/regresión [257]. Desde la aparición

¹³Por ejemplo, cuando se evalúan imágenes 11C-PiB para el estudio de la EA, se suele utilizar la materia gris del cerebelo como región no específica para el cálculo de los valores SUVR [245].

¹⁴Incluyendo el registro de las imágenes MRI al espacio definido por el atlas de Talairach [249, 250]; la normalización de sus valores de intensidad; la aplicación de un método de *skull stripping* similar al definido en [251]; y la definición de unos umbrales de intensidad que permiten modelar una versión segmentada preliminar de la imagen.

de los primeros métodos de extracción de características basados en el análisis de texturas, como el publicado por **Haralick et al.** en [258] donde se propone el uso de matrices de co-ocurrencia para el análisis de los valores de gris de una imagen, han ido surgiendo varias propuestas que han tratado de poner en práctica estas transformaciones también en el campo de la neuroimagen incluyendo, entre otros, los siguientes:

- **Zhang et al.** en [259]. Para cada ROI de interés, los autores combinan varios parámetros procedentes de los histogramas de intensidad, de medidas de la no-uniformidad de cada región, del análisis del valor absoluto del gradiente; y varios parámetros derivados del análisis de texturas de Haralick; utilizando imágenes MRI para evaluar la EA.
- **Martinez-Murcia et al.** en [169, 260]. Se propone el desarrollo de dos sistemas CAD con imágenes FP-CIT SPECT para evaluar la EP clasificando parámetros derivados del análisis de Haralick.
- **Ding et al.** en [261]. Donde los autores combinan el método *Voxel-Based Morphometry* (VBM) descrito por **Ashburner & Friston** en [262], junto a una etapa de filtrado de imágenes estructurales para la caracterización de sujetos con EA.
- **Maani et al.** en [257]. Incluye un análisis de texturas basado en un mapa estadístico similar al análisis de Haralick pero aplicado a la proyección de las regiones de intensidad de GM y WM.
- **Li et al.** en [48]. Los autores desarrollan un sistema CAD para la evaluación de la conversión entre el DCL y la EA con marcadores procedentes del análisis de Haralick utilizando imágenes 18F-FDG PET.
- **Vaithinathan et al.** en [263]. Donde se justifica el desarrollo de un sistema CAD con imágenes MRI para la evaluación de la EA con características como el momento central, la homogeneidad, la entropía y el contraste.
- **Cai et al.** en [264]. Se desarrolla un sistema CAD para la EA con características extraídas en tres niveles: marcadores relacionados con los promedios de intensidad, desviación, simetría, curtosis y entropía de varias ROIs; características derivadas del análisis de Haralick; y variables relativas a las relaciones espaciales entre niveles de intensidad de gris similares.

- **Mapas de densidad de intensidades** - Tratan de cuantificar los patrones de atrofia en imagen MRI para la evaluación de patologías como la EA mediante medidas de la densidad de WM, GM y del líquido cefalorraquídeo; en diferentes regiones del cerebro. Como ejemplos más destacados cabe mencionar los trabajos publicados en [265–267].

2.4.4. Métodos estadísticos para la extracción de características

Con el fin de evitar el posible sesgo derivado del entrenamiento de los modelos de aprendizaje automático con datos procedentes de sujetos de prueba, los métodos que se incluyen en la etapa de preprocesado no pueden depender las etiquetas de los datos de entrada. Es por esto que, siguiendo el esquema representado en la Figura 2.3, en este apartado se describen algunos de los métodos de extracción de características que, por su naturaleza a la hora de trabajar con dichas etiquetas o por su similitud respecto a otros modelos que sí lo hacen, deben incluirse como una etapa intermedia dentro del bucle de validación cruzada.

Métodos para la reducción del espacio dimensional mediante proyecciones ortogonales

Hasta la fecha, la práctica totalidad de repositorios de imágenes cerebrales para el estudio de la demencia han estado caracterizados por su pequeño tamaño muestral (cientos o pocos de miles de imágenes en los mejores casos), una alto número de variables de entrada (asociadas a la cantidad de vóxeles de intensidad en cada prueba de imagen clínica) y, en muchas ocasiones, una alta correlación entre estas variables de entrada [268]. En un escenario de estas características, el riesgo de efectos adversos como el sobreajuste de los modelos de aprendizaje a los datos de entrenamiento es muy elevado¹⁵. Para evitarlo, algunas propuestas de análisis de datos de neuroimagen han tratado de aplicar alguna técnica de reducción de la dimensionalidad que les permitiera, a partir de un número más reducido de variables transformadas, explicar la práctica totalidad de la información contenida en estos datos de entrada minimizando –en la medida de lo posible– el riesgo de incurrir en el posible sobreajuste de los modelos de entrenamiento.

¹⁵Para justificar este hecho supóngase que se quiere analizar el conjunto de imágenes FP-CIT SPECT del repositorio PPMI (Sección 3.1.1) para desarrollar un sistema CAD para la evaluación de la EP basado en los valores de intensidad de estas imágenes. En total, cada imagen contiene un total de $91 \times 109 \times 91$ (dimensión de cada imagen) valores de intensidad, lo cual multiplica por más de 1000 la cantidad de sujetos incluidos en el estudio.

Uno de los algoritmos de reducción de la dimensionalidad más utilizados en la actualidad es el método de Análisis de Componentes Principales (*Principal Component Analysis*, PCA) propuesto originalmente por **Wold et al.** en [144].

Explicado de manera resumida, PCA trata de aplicar transformaciones ortogonales a un espacio de características de entrada (posiblemente correlacionadas entre sí) para generar un nuevo espacio linealmente no correlacionado que preserva toda la información del conjunto original [145]. Matemáticamente, el procedimiento comienza con una estandarización de los datos de entrada, seguido del cómputo de la matriz de covarianza (Σ), y la descomposición mediante el método SVD (*Singular Value Decomposition*) de la matriz de covarianza que ahora podrá expresarse como un producto de matrices (descrito en la expresión (2.4)) que incluyen: 1) la matriz \mathbf{U} , que contiene los autovectores de $\Sigma\Sigma^T$; 2) la matriz diagonal \mathbf{S} , que contiene los autovalores de $\Sigma\Sigma^T$; y 3) la matriz \mathbf{V}^T , que contiene los autovectores de $\Sigma^T\Sigma$.

$$\Sigma = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (2.4)$$

Como cada autovalor de la matriz \mathbf{S} representa la varianza de su autovector (componente principal), la reducción de la dimensionalidad de los datos de entrada vendrá dada por la selección de las primeras k -componentes (las que aglutinan la mayor parte de la información). Un criterio común, en este sentido, es seleccionar tantas componentes como sea necesario hasta explicar al menos el 90% de la varianza total¹⁶ de los datos de entrada.

Este método de reducción de la dimensionalidad ha sido utilizado en múltiples escenarios de análisis de datos de neuroimagen incluyendo entre los más destacados el modelo CAD propuesto en [135] para la evaluación de la EA con imagen funcional 18F-FDG; los análisis estadísticos descritos por **Campbell et al.** en [269] con los que los autores evalúan la distribución de intensidades en imágenes 11C-PiB PET adquiridas a sujetos con EA; el sistema CAD propuesto por **Salvatore et al.** en [5] para la evaluación de la EP frente a la parálisis supranuclear progresiva utilizando imágenes MRI; el modelo descrito en [139] para el análisis de síndromes parkinsonianos utilizando imágenes 18F-FDG; los trabajos publicados en [270, 271] para evaluar la EA con imágenes MRI; o el modelo de *deep learning* presentado por **Martínez-Murcia et al.** en [272] utilizando *autoencoders* convolucionales sobre imágenes 18F-FDG de sujetos sanos, pacientes con DCL y pacientes con diagnóstico confirmado de EA.

Una alternativa a PCA también muy utilizada en neurociencia es la técnica conocida como *Partial Least Squares* (PLS) [146], que trata de crear cada componente ortogonal no sólo mediante el análisis de la covarianza

¹⁶El porcentaje de varianza explicada por cada componente se calcula dividiendo cada autovalor entre la suma total de todos los autovalores.

entre las variables de entrada sino que también intentando maximizar la distancia entre clases de entrada en problemas de clasificación. Aunque esta opción ha sido criticada en algunos contextos por su limitación a la hora de justificar qué peso ha asignado a cada variable de entrada, presenta la ventaja adicional de que cada componente de PLS consigue explicar más información que la que se explicaría con una componente de PCA, lo que se traduce en una mayor reducción del espacio dimensional [273].

Entre las propuestas con ML aplicado al estudio de los trastornos neurodegenerativos más frecuentes utilizando PLS, cabe mencionar el trabajo de **Segovia et al.** en [274] donde los autores desarrollaron un sistema CAD para la evaluación de la EP con imágenes FP-CIT SPECT, procedentes de sujetos sanos y pacientes diagnosticados con la enfermedad (algunos de ellos todavía en etapas tempranas), llevando a cabo un análisis de los hemisferios por separado; la comparativa entre PCA y PLS presentada en [275] utilizando imágenes fMRI para la evaluación de la EA; los sistemas CAD descritos en [276, 277] utilizando imágenes 18F-FDG para el estudio de la EA; o el modelo publicado en [9] con imágenes MRI segmentadas.

Métodos basados en mapas de puntuaciones

En relación con el análisis de texturas comentado en la Sección 2.4.3, aunque son pocos, existen algunos trabajos que requieren de un ajuste de sus parámetros en función de las clases de entrada. En este grupo se puede mencionar el modelo desarrollado por **Hett et al.** en [278] donde los autores tratan de identificar la conversión de pacientes con DCL a EA mediante un análisis de texturas de la región del hipocampo en imágenes MRI generadas a partir de varios filtros de Gabor, [279], y de la aplicación de un sistema de puntuaciones entrenado con sujetos de ambas clases para los valores de intensidad en esta región [280]. Otro trabajo similar es el modelo descrito por **Nanni et al.** en [281] que combina el uso de descomposiciones de Wavelet [282]; filtros de Gabor; descriptores de características como los incluidos en la librería *VLFeat* (disponible en: <https://www.vlfeat.org/>); histogramas de intensidad; un análisis de la covarianza; la evaluación de algunos patrones de intensidad como en [283]; y por último, la aplicación del sistema de calificaciones descrito en [280].

2.4.5. Selección de características

Según **Kohavi & John** en [284], el objetivo principal de un algoritmo de selección de características en un problema de clasificación es identificar un subconjunto de características \mathbf{X}_{opt} a partir del conjunto de datos de entrada $\mathbf{X}_{\text{entrada}}$, tales que el número de variables de \mathbf{X}_{opt} sea el menor posible (evitando así el sobreajuste del modelo de aprendizaje a los datos de

entrenamiento) y que su tasa de clasificación sea máxima. Para conseguir este objetivo, hace falta descartar todas aquellas variables de entrada que no sean relevantes [285]. Y para ello, se puede determinar de forma directa la influencia que tiene cada variable sobre el resultado de la inferencia del bloque de aprendizaje, o se puede intentar asignar una puntuación a cada característica de entrada según una métrica de selección determinada y elegir las variables con mejores puntuaciones [147]. Pero incluso así, esta primera afirmación puede no estar teniendo en cuenta la redundancia de información procedente de variables con una alta correlación entre sí. Es por todo esto, que muchos algoritmos de selección de características tratan de categorizar las variables de $\mathbf{X}_{\text{entrada}}$ en dos categorías: un primer grupo constituido por características con una relevancia “fuerte” en el sentido de que si se descarta una sola de estas variables, el resultado de la clasificación final disminuirá; y un segundo grupo (de “relevancia débil”) en el que si se descarta alguna de las variables de entrada por separado no empeora la clasificación pero que si se descartan en conjunto (no solo una de ellas como en el caso anterior) sí que se manifiesta una pérdida de rendimiento. Si se combinan todas estas condiciones, entonces sí que se podrá afirmar que un algoritmo de selección de características debe ser capaz de eliminar todas aquellas características de entrada que no tengan relevancia (ni fuerte ni débil), y que incluso asigna algún tipo de métrica a cada variable que permite cuantificar de alguna forma el peso que esa variable puede tener en la clasificación final.

En la bibliografía actual podemos encontrar tres tipos de algoritmos de selección de características para aprendizaje supervisado (en aprendizaje no supervisado y en aprendizaje semi-supervisado se pueden evaluar otras opciones, [286]):

- **Métodos de filtro** - Constituyen el enfoque más sencillo para la selección de características ya que consisten en asignar una puntuación, $F(\cdot)$, a cada característica de entrada, $\mathbf{x}^{(j)} \in \mathbf{X}_{\text{entrada}}$, con índice j . Una vez han sido evaluadas todas las características por separado, se seleccionan aquellas con las calificaciones más altas o que superen cierto umbral [287].
- **Métodos de envolvente** - Requieren una etapa de aprendizaje supervisado, a partir de un subconjunto de datos del conjunto original de características, para su ajuste [284].
- **Métodos integrados** - La selección de características se lleva a cabo mediante el proceso de ajuste del modelo. Entre los modelos más utilizados en neuroimagen podemos encontrar modelos de regresión o de máquinas de vectores de soporte como en [288], e incluso métodos basados en árboles de decisión como en [137, 289].

Aunque el ajuste de los métodos de filtro en algunos escenarios puede ser menor, para la realización de esta tesis doctoral se han empleado este tipo de técnicas como referencia por su simplicidad, su velocidad de cálculo y porque son más fácilmente generalizables. En la actualidad existen muchos procedimientos para hacer la selección de características utilizando métodos de filtro. Aquí se resumen algunos de los más utilizados en neuroimagen:

- **t-Test y Test de Welch** - Ambos tests asignan a cada característica de entrada, $\mathbf{x}^{(j)}$, una puntuación dada por la expresión (2.5) donde μ_k (con $k = 1, 2$) representa los valores medios de las muestras pertenecientes a cada clase; σ_k los valores de sus desviaciones; y n_k el número de instancias en cada grupo. La diferencia entre ambos métodos radica en el cómputo de la varianza. Mientras que el t-Test asume que las varianzas de ambas clases son idénticas ($\sigma_1 = \sigma_2$), el test de Welch permite que estas sean diferentes [148].

$$F(\mathbf{x}^{(j)}) = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (2.5)$$

Como ejemplos de trabajos de neuroimagen que hacen uso de tests estadísticos con la forma descrita en (2.5), se pueden citar los modelos desarrollados en [137, 290–293].

- **Fisher's Score** - Referido en trabajos como [263, 274, 294], este algoritmo trata de encontrar un subconjunto de características tales la distancia entre muestras pertenecientes a la misma clase sea mínima, pero que la distancia entre puntos pertenecientes a clases diferentes sea lo más amplia posible [295]. En su versión original simplificada, asigna una puntuación, $F(\cdot)$, a cada característica de entrada $\mathbf{x}^{(j)}$ como sigue en (2.6) donde $k = 1, \dots, c$ representa el índice de la k -ésima clase de los datos de entrada; $\mu_k^{(j)}$ es el valor medio de las muestras pertenecientes a la variable de entrada $\mathbf{x}^{(j)}$ que tienen como etiqueta la clase k -ésima; $\sigma_k^{(j)}$ su desviación; μ_j el promedio de todas las muestras en $\mathbf{x}^{(j)}$ (independientemente de su clase); y donde n_k es igual al número de muestras que pertenecen a la clase con etiqueta k . Una vez se calculan todos los valores $F(\mathbf{x}^{(j)})$, se seleccionan las m características que tengan mayor puntuación.

$$F(\mathbf{x}^{(j)}) = \frac{\sum_{k=1}^c n_k (\mu_k^{(j)} - \mu_j)^2}{\sum_{k=1}^c n_k (\sigma_k^{(j)})^2} \quad (2.6)$$

- **Mann-Whitney-Wilcoxon U-Test** - Constituye un tipo de test estadístico no-paramétrico de hipótesis nula en el que se comparan las medianas de los datos procedentes de dos poblaciones diferentes (clases, $k = 1, 2$) [149]. En este caso las puntuaciones se determinan a partir del estadístico U calculado como el valor mínimo de los estadísticos U_k (es decir, $U = \min(U_1, U_2)$) determinados como sigue en (2.7). En esta expresión, n_k representa el número de muestras que pertenecen a la clase k , y R_k equivale al sumatorio de los rangos de esa clase.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad (2.7)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

A partir de este valor de U , si el número de muestras de entrada es lo suficientemente elevado (normalmente $n_1 + n_2 \geq 25$), puede calcularse el estadístico $Z \sim N(0, 1)$ como sigue en (2.8) que permite establecer un umbral a partir del cual rechazar de la hipótesis nula (p_{valor}).

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0, 1) \quad (2.8)$$

Su uso está extendido en neuroimagen en trabajos como [296–298] por ser un método capaz de trabajar bien con distribuciones con colas pesadas y/o con fuerte asimetría [31].

- **ANOVA** - El método de análisis de la varianza (*ANalysis Of VAriance*, ANOVA), trata de calcular las medias entre los dos conjuntos de datos para comparar sus varianzas (intervarianza) frente a la varianza promedio de cada conjunto de datos por separados (intravarianza) [151].

Una limitación del test es que suele considerar que las clases de entrada están balanceadas (mismo número de sujetos por clase). Cuando esta igualdad no se cumple, se reduce sustancialmente la robustez de la prueba aunque sigue siendo válida.

En un escenario binario y balanceado con N muestras de entrada por clase, y siendo $\mathbf{x}^{(j)}$ la j -ésima característica de entrada con $i = 1, 2, \dots, N$ realizaciones; llamaremos $\overline{\mathbf{x}}^{(j)}_k$ a la media de las realizaciones pertenecientes a la clase k , y $\overline{\mathbf{x}}^{(j)}$ al promedio de valores del conjunto completo:

$$\begin{aligned}\overline{\mathbf{x}^{(j)}}_k &= \frac{1}{N} \sum_{i=1}^N x_{i,k}^{(j)} \quad \text{con} \quad k = \{1, 2\} \\ \overline{\mathbf{x}^{(j)}} &= \frac{1}{2N} \sum_{k=1}^2 \sum_{i=1}^N x_{i,k}^{(j)}\end{aligned}\tag{2.9}$$

Las varianzas de cada clase ($\hat{\sigma}_{\mathbf{x}_k^{(j)}}^2$), así como del conjunto completo ($\hat{\sigma}_{\mathbf{x}^{(j)}}^2$), y la covarianza entre ambos subconjuntos quedan como:

$$\begin{aligned}\hat{\sigma}_{\mathbf{x}_k^{(j)}}^2 &= \sum_i^N (x_{i,k}^{(j)} - \overline{\mathbf{x}^{(j)}}_k)^2 \quad \text{con} \quad k = \{1, 2\} \\ \hat{\sigma}_{\mathbf{x}^{(j)}}^2 &= \sum_{k,i} (x_{i,k}^{(j)} - \overline{\mathbf{x}^{(j)}})^2 \\ \hat{\sigma}_{\text{interconjuntos}}^2 &= \sum_k (\overline{\mathbf{x}^{(j)}}_k - \overline{\mathbf{x}^{(j)}})^2\end{aligned}\tag{2.10}$$

Una vez calculados estos parámetros, el valor del estadístico F se derivará de las expresiones en (2.11) como el cociente en (2.12). Este valor permite rechazar la hipótesis nula de que la media de ambos conjuntos sean iguales.

$$\hat{S}_{\mathbf{x}_k^{(j)}}^2 = \frac{\hat{\sigma}_{\mathbf{x}_k^{(j)}}^2}{2(N-1)}\tag{2.11}$$

$$\hat{S}_{\text{interconjuntos}}^2 = \frac{\hat{\sigma}_{\text{interconjuntos}}^2}{2(N\hat{\sigma}_{\text{interconjuntos}}^2 - 1)}$$

$$F = \frac{\hat{S}_{\mathbf{x}_k^{(j)}}^2}{\hat{S}_{\text{interconjuntos}}^2}\tag{2.12}$$

Algunos ejemplos de trabajos de neuroimagen que han utilizado ANOVA como método de selección de características son las publicaciones presentadas en [299, 300].

- **Relief** - En un problema de clasificación binario, para cada variable de entrada selecciona r muestras (de forma aleatoria) y determina su relevancia, $F(\mathbf{x}^{(j)})$, según (2.13) [301]. Esta expresión está basada en el cálculo de las distancias, $d(\cdot)$, entre las instancias que pertenecen a la misma clase ($x_{i,k}^{(j)}$) y las que pertenecen a la clase opuesta ($x_{i,\neq}^{(j)}$):

$$F(\mathbf{x}^{(j)}) = \frac{1}{2} \sum_{i=1}^r d(x_i^{(j)} - x_{i,k}^{(j)}) - d(x_i^{(j)} - x_{i,\neq}^{(j)})\tag{2.13}$$

En escenarios multiclase, existe un modelo más avanzado conocido como *ReliefF* [302]. Aunque ambos métodos están más orientados a escenarios de clasificación con algoritmos como KNN también se pueden encontrar trabajos en los que se implementan clasificadores como SVM [303].

2.4.6. Algoritmos de clasificación y regresión

Respecto a la elección de los algoritmos de inferencia, este apartado resume las características más importantes de algunos de los métodos más utilizados en el campo de la neurociencia con referencias a trabajos como los que se han desarrollado en esta tesis doctoral.

Métodos de clasificación

- **SVM.** Históricamente se ha justificado que las máquinas de vectores de soporte (*Support Vector Machines*, SVM) constituyen uno de los algoritmos más robustos tanto en escenarios de clasificación como de regresión [304].

A nivel conceptual, para un escenario de clasificación binaria, SVM constituye un modelo lineal que trata de establecer una separación lo más óptima posible entre los dos conjuntos de datos de entrada [305]. Para definir este límite, también conocido como hiperplano de decisión, el clasificador hace uso de funciones discriminantes lineales en un espacio de características multidimensional [49].

Aunque no es objeto de este apartado justificar todo el desarrollo matemático de SVM, en líneas generales los hiperplanos de decisión más simples son aquellos que siguen una forma como la descrita en (2.14) con \mathbf{w}^T un vector de pesos ortogonal al hiperplano que pondera aquellas características con mayor separación entre clases; $\mathbf{x}_i \in \mathbf{X}_{\text{train}}$ (con $i = 1, 2, \dots, Q$) como un vector que representa los vectores de características de entrenamiento (con etiquetas de salida asociadas y_i); y con b como valor umbral que marca el punto de corte con los ejes de coordenadas.

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x}_i + b = 0 \quad (2.14)$$

Como el hiperplano $g(\mathbf{x})$ que separa las clases de entrada (ψ_0 y ψ_1), puede que no sea el único e incluso el más óptimo (de hecho, lo normal es escoger entre todos los hiperplanos posibles aquel que maximiza la distancia entre clases), para determinar \mathbf{w} y b lo que se hace es minimizar una función de costes $\mathbf{J}(\mathbf{w})$ asociada a la distancia entre

clases según (2.15) sujeta a la restricción que aparece en (2.16), con $y_i = 1$ cuando $x_i \in \psi_0$ y -1 en caso contrario:

$$\underset{\mathbf{w}, b}{\text{minimizar}} \quad \mathbf{J}(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.15)$$

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad (2.16)$$

La solución a esta ecuación se puede obtener aplicando multiplicadores de Lagrange (representados en las ecuaciones como λ_i y a los que llamamos “vectores de soporte” cuando $\lambda_i > 0$) y teniendo en cuenta las condiciones de Karush-Kuhn-Tucker.

Cuando las clases son separables linealmente, se llega a la expresión final en (2.17) con $\sum_{i=1}^Q \lambda_i y_i = 0$ que permite obtener los parámetros necesarios para determinar el hiperplano óptimo:

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^Q \lambda_i y_i \mathbf{x}_i \\ b &= y_i - \mathbf{w} \mathbf{x}_i \end{aligned} \quad (2.17)$$

Si las clases no son separables linealmente, entonces será necesario definir una banda en torno al hiperplano de decisión y analizar los valores que caigan dentro y fuera de esta banda de decisión según (2.18):

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq (1 - \xi_i) \quad (2.18)$$

En esta expresión, el valor de ξ_i dependerá de si las realizaciones de los datos de entrenamiento caen fuera de la banda y se etiquetan correctamente ($\xi_i = 0$); caen dentro de la banda y se etiquetan correctamente ($0 < \xi_i < 1$); o caen dentro de la banda pero se etiquetan de manera incorrecta ($\xi_i = 1$).

En cualquier caso, el objetivo ahora será minimizar la nueva función de costes en (2.19) (sujeta a las condiciones en (2.20)) y hacer que el margen de separación entre clases sea lo más amplio posible pero sin incrementar en exceso el valor de ξ_i (evitando las clasificaciones incorrectas):

$$J(\mathbf{w}, b, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^Q \xi_i \quad (2.19)$$

$$\begin{aligned} y_i(\mathbf{w}^T \mathbf{x}_i + b) &\geq 1 - \xi_i \\ \xi_i &> 0 \end{aligned} \quad (2.20)$$

De nuevo, utilizando multiplicadores de Lagrange¹⁷ y aplicando las condiciones de Karush-Kuhn-Tucker, se puede llegar a la expresión de los parámetros \mathbf{w} y b necesarios para construir el hiperplano de separación entre clases con el menor error posible.

Como alternativa a este método, se puede aplicar un mapeo del espacio original de los datos de entrada a un nuevo espacio de dimensión superior que sí permita que las clases sean linealmente separables. Gracias al teorema de Mercer y aplicando lo que se conoce como “*truco del kernel*”, y utilizando la notación Q_s para indicar el número de vectores de soporte de la solución, se puede definir una función $K(\cdot)$ tal que el producto escalar del espacio transformado se puede expresar a partir de una función de similitud en el espacio original. A esta función se le conoce como *kernel* y puede tener muchas formas. Entre las más comunes en escenarios de clasificación con datos de neuroimagen tenemos:

□ **Kernel lineal:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j + C \quad (2.21)$$

□ **Kernel polinómico:**

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma(\mathbf{x}_i \cdot \mathbf{x}_j))^d \quad (2.22)$$

□ **Kernel RBF** (*Radial Basis Function kernel*):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (2.23)$$

Independientemente del *kernel* que se utilice, el problema de clasificación final queda determinado por la expresión (2.24) y sujeto a las restricciones en (2.25):

$$\underset{\lambda}{\text{maximizar}} \left(\sum_{i=1}^Q \lambda_i - \frac{1}{2} \sum_{i,j=1}^Q \lambda_i \lambda_j y_i y_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (2.24)$$

¹⁷El parámetro C limita el número máximo de multiplicadores de Lagrange.

$$\begin{cases} \sum_i^Q \lambda_i y_i = 0 \\ 0 \leq \lambda_i \leq C \\ \mathbf{x} := \psi_0 \iff g(\mathbf{x}) = \sum_{i=1}^{Q_s} \lambda_i y_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + b > 0 \\ \mathbf{x} := \psi_1 \iff g(\mathbf{x}) = \sum_{i=1}^{Q_s} \lambda_i y_i \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + b < 0 \end{cases} \quad (2.25)$$

En la última década, la mayor parte de sistemas CAD desarrollados para el estudio y análisis de la demencia han hecho uso de clasificadores de tipo SVM. Entre estos trabajos, se pueden encontrar soluciones CAD para el diagnóstico automático de la EA con imágenes 18F-FDG PET, como los desarrollados en [48, 135, 296]; modelos de clasificación con imágenes FP-CIT SPECT para el estudio de la EP, [41, 169, 170]; y trabajos con imágenes MRI tanto para la EA como para la EP [5, 267]. Aunque la práctica totalidad de estos modelos hacen referencia a funciones de *kernel* de tipo lineal, en los últimos años también han ido apareciendo otras propuestas con diferentes tipos de funciones. Entre las más frecuentes cabe mencionar los trabajos publicados en [9, 306–310] con funciones *kernel* de tipo RBF.

- **ν SVC.** Se trata de un clasificador que muchos trabajos describen como una variante de SVM no lineal y que hace uso de un parámetro $\nu \in (0, 1]$ en lugar de $C \in [0, \infty)$. Este parámetro permite controlar el ratio de vectores de soporte y su relación con el error de entrenamiento [311].

La función de costes de ν SVC sigue la forma descrita en (2.26) y se encuentra sujeta a la restricciones en (2.27). En ambas expresiones, ρ constituye una variable que permite optimizar el margen de separación entre clases con especial atención a aquellos casos para los que $\xi_i = 0$.

$$J(\mathbf{w}, \xi, \rho) = \frac{1}{2} \|\mathbf{w}\|^2 - \nu\rho + \frac{1}{2} \sum_{i=1}^Q \xi_i \quad (2.26)$$

$$\begin{cases} y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq \rho - \xi_i \\ \xi_i \geq 0 \\ \rho \geq 0 \end{cases} \quad (2.27)$$

De forma similar al desarrollo de SVM, utilizando multiplicadores de Lagrange y aplicando las condiciones de Karush-Kuhn-Tucker, se llega al problema de optimización reflejado en la expresión (2.28) y sujeto a las condiciones en (2.29). La resolución de este problema define el plano de decisión para ν SVC.

$$\underset{\lambda}{\text{maximizar}} \quad \left(-\frac{1}{2} \sum_{i,j=1}^Q \lambda_i \lambda_j y_i y_j \mathbf{K}(\mathbf{x}_i^T \mathbf{x}_j) \right) \quad (2.28)$$

$$\begin{cases} 0 \leq \lambda_i \leq \frac{1}{Q} \\ \sum_{i=1}^Q \lambda_i y_i = 0 \\ \sum_{i=1}^Q \lambda_i \geq \nu \end{cases} \quad (2.29)$$

Aunque este algoritmo de clasificación está mucho menos extendido que SVM, su aplicación práctica al análisis de datos médicos también se encuentra justificada en la actualidad con publicaciones como la presentada en [312] para la evaluación de la enfermedad arterial coronaria.

- **Naïve bayesiano.** Es un algoritmo de clasificación no lineal basado en el uso de probabilidades condicionadas según el Teorema de Bayes descrito en (2.30) para los sucesos A y B conocida la probabilidad condicional $P(B|A)$ [313].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.30)$$

Si se considera que las componentes de los vectores de características de entrada (\mathbf{x}_i con $i = 1, 2, \dots, Q$) son independientes unos de otros, entonces predecir la clase de salida (ψ_k) en función de las características de entrada quedará descrito como sigue en (2.31), siendo esta última expresión equivalente a la que aparece en (2.32):

$$P(\psi_k | \mathbf{x}_1, \dots, \mathbf{x}_Q) = \frac{P(\mathbf{x}_1 | \psi_k) P(\mathbf{x}_2 | \psi_k) \dots P(\mathbf{x}_Q | \psi_k) P(\psi_k)}{P(\mathbf{x}_1) P(\mathbf{x}_2) \dots P(\mathbf{x}_Q)} \quad (2.31)$$

$$P(\psi_k | \mathbf{x}_1, \dots, \mathbf{x}_Q) \propto P(\psi_k) \prod_{i=1}^Q P(\mathbf{x}_i | \psi_k) \quad (2.32)$$

El objetivo final de la clasificación es escoger la clase que tenga mayor probabilidad (criterio de maximización a posteriori). Según esto, la forma final de la clasificación quedará finalmente descrita como sigue a continuación:

$$\psi_{\text{estimada}} = \underset{\psi_k}{\text{argmax}} \left(P(\psi_k) \prod_{i=1}^Q P(\mathbf{x}_i | \psi_k) \right) \quad (2.33)$$

Junto a SVM, los clasificadores de tipo Naïve Bayesian constituyen uno de los algoritmos de clasificación más utilizados para el análisis de datos de sujetos con demencia. Como ejemplos, se pueden citar aquí los modelos descritos en [314, 315] para el estudio de la EP, o los trabajos presentados en [316–320] sobre la evaluación de la EA.

- RF.** Aunque esta técnica también se utiliza para la selección de características, *Random Forests* constituye un método de clasificación basado en un aprendizaje de conjuntos con voto por mayoría (Sección 2.8) con múltiples árboles de decisión [321]. Según estos modelos, como el representado en la Figura 2.5, cada árbol se entrena de forma independiente al resto de árboles y con un conjunto aleatorio de características de entrada. Como cualquiera de estas características puede pertenecer simultáneamente a varios árboles de decisión (método de aleatorización con reemplazo), la varianza de la clasificación final suele ser reducida.

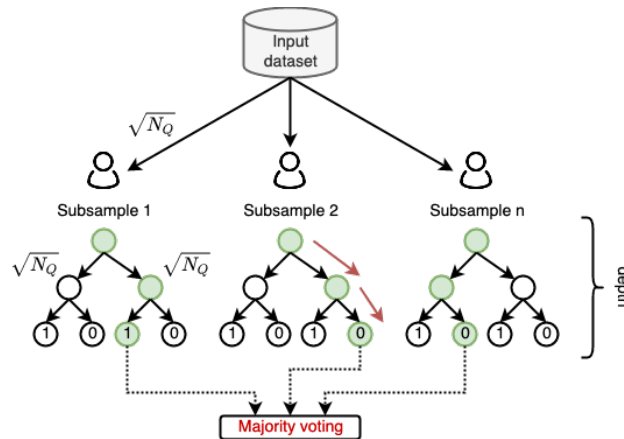


Figura 2.5: Diagrama explicativo de un sistema de clasificación con RF.

La recomendación general a la hora de diseñar un modelo de clasificación con RF, es que en cada nodo de decisión tenga en cuenta $n_q = \sqrt{N_Q}$ características de entrada, siendo N_Q el número total de características a la entrada del sistema completo. La profundidad (*depth*) o número de etapas de decisión, modela la complejidad del sistema (incrementando el coste computacional) pero consiguiendo un mejor nivel de ajuste a los datos de entrada. Al margen del número de árboles de decisión por cada capa que también modela la complejidad del sistema, existe un parámetro conocido como *gini impurity* que modela la influencia que cada decisión tiene sobre la predicción de cada etiqueta. Esto es, para un escenario de clasificación binaria, representa la probabilidad, p_j , de etiquetar una muestra como perteneciente a la clase ψ_0 ($j = 0$) o ψ_1 ($j = 1$) según (2.34):

$$I_G = 1 - \sum_{j=0}^1 p_j^2 \quad (2.34)$$

Como interesa que la probabilidad de fallo sea la menor posible, el algoritmo seleccionará iterativamente aquellos nodos cuyo valor de I_G sea el menor posible. Si para una ramificación dada se determina que el valor de I_G disminuye considerablemente al tener en cuenta una característica concreta, se intentará asignar un peso mayor al resto de clasificaciones que involucren a dicha variable.

Respecto a los modelos de clasificación de neuroimagen con RF para el análisis de la demencia con ML, entre las publicaciones más citadas en los últimos años caben destacar las propuestas desarrolladas en [137, 139, 289, 299, 303, 322, 323], tanto para la evaluación de la EP como de la EA.

- Redes neuronales.** Constituyen una herramienta que intenta modelar el comportamiento de las neuronas biológicas mediante la suma de señales de entrada (estímulos), $x_i \in \mathbf{x}$, ponderadas según unos pesos, w_i , y generando una señal a la salida resultado de una función de activación. En la Figura 2.6 se han incluido un diagrama básico de una red neuronal. Como puede observarse, la unidad fundamental de la arquitectura descrita es el perceptrón (neurona). Una red neuronal completa es aquella que incluye una serie de capas con varios nodos de computación (neuronas) interconectados entre sí. El modelo descrito en esta figura incluye una capa de entrada (sin neuronas), dos capas ocultas y una capa de salida.

Como ya se comentó en la Sección 1.1, las redes neuronales actúan como un aproximador universal de funciones [50]. Por tanto, para cualquier función continua, $f(\mathbf{x})$, y un error acotado $\varepsilon > 0$, siempre existirá una representación en forma de red neuronal, $g(\mathbf{x})$, tal que la diferencia entre la red neuronal y la función de interés quede acotado por el error ε según (2.35).

$$\|f(\mathbf{x}) - g(\mathbf{x})\| < \varepsilon \quad (2.35)$$

La cantidad de capas de la red, el número de neuronas por capa y la interconexión de las neuronas entre sí, son las propiedades que confieren a la red su capacidad para ajustarse a cualquier modelo de datos de entrada. Existen múltiples tipos de arquitecturas con las que se están desarrollando modelos de aprendizaje para la evaluación de la demencia con redes neuronales [324]. Entre las más citadas se incluyen:

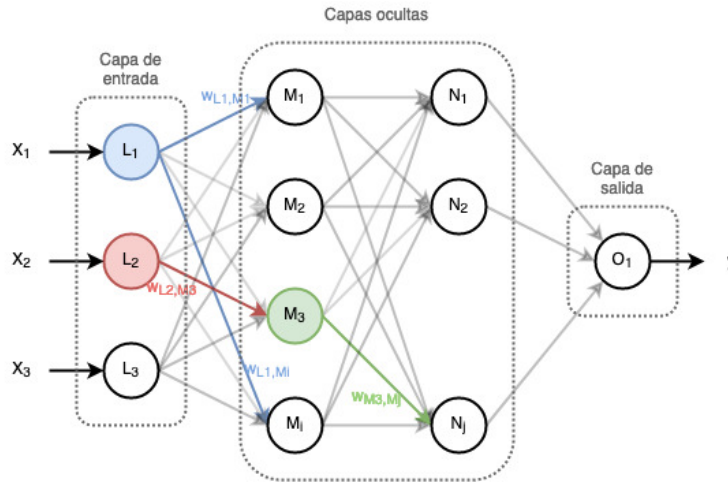


Figura 2.6: Arquitectura básica de una red neuronal con dos capas ocultas.

- **Redes neuronales artificiales** (*Artificial Neural Network*, ANN). Representa el esquema fundamental de una red neuronal profunda. Como ejemplos de su aplicación en problemas de análisis de imágenes cerebrales, se pueden señalar los trabajos publicados en [325, 326] para la evaluación de la EA; o el modelo presentado por **Tang et al.** en [327] para la predicción de la progresión de pacientes con EP.
- **Autoencoders.** Propuestos originalmente por **Hinton** en [328] como una forma de reducir la dimensionalidad de los datos de entrada, se trata de redes capaces de aprender una representación (codificación) del espacio de características de entrada, y construyendo a la salida una representación del espacio de entrada en función del aprendizaje adquirido por la red. En neuroimagen podemos encontrar algunos ejemplos de redes con *autoencoders* para el diagnóstico y supervisión de la demencia como pueden ser los modelos presentados en [329] para la identificación de sujetos con DCL; o los trabajos publicados en [330–332] para la EA.
- **Redes neuronales convolucionales** (*Convolutional Neural Networks*, CNN). Las redes neuronales tradicionales hacen uso de vectores como espacio de características de entrada. Por el contrario, una red neuronal convolucional puede trabajar directamente a nivel de imagen gracias al uso de filtros convolucionales. Como el aprendizaje de este tipo de arquitecturas es capaz de aprovechar la información estructural de cada vóxel así como de sus vecinos, resulta el tipo de red neuronal más frecuentemente utilizada en escenarios de clasificación con imágenes cerebrales [180]. Como ejemplos de este tipo de soluciones podemos señalar aquí: las es-

estructuras descritas en [333, 334], donde los autores hacen uso de redes convolucionales con imágenes estructurales para diferenciar entre sujetos sanos y pacientes con diferentes tipos de demencia; el sistema CAD propuesto en [335] para la caracterización de pacientes con EP utilizando imágenes FP-CIT SPECT; los trabajos presentados en [16, 336] para la identificación de pacientes con DCL o EA; y el modelo de clasificación presentado en [180] para evaluar sujetos sanos frente a pacientes con EA aplicando arquitecturas de redes neuronales convolucionales ya conocidas como *LeNet* y *AlexNet* [337, 338].

- **Redes neuronales recurrentes** (*Recurrent Neural Networks*, RNN). En este tipo de arquitecturas las salidas de una o varias capas posteriores pueden realimentar la entrada de algunas capas anteriores de la red. Esta propiedad le confiere a la red una capacidad de memoria que ayuda a procesar secuencias de datos longitudinales. De hecho, aunque su aplicación principal es el análisis longitudinal, también se pueden encontrar sistemas de clasificación con RNN como el publicado en [339] para el estudio del estado de sujetos sanos, con DCL o con EA; el modelo de **Cui et al.** presentado en [340] para la predicción del estado de la EA; o el estudio focalizado en el análisis de la región del hipocampo descrito en [341] para el estudio de la EA.
- **Transfer learning**. Describe una metodología según la cual en vez de entrenar una red neuronal desde cero, lo que se hace es utilizar un modelo de aprendizaje ya entrenado con un conjunto de datos a gran escala (con dominio o tarea de origen diferente) y adaptar dicho modelo al dominio o tarea de interés. Respecto a los modelos que hacen uso de esta técnica se pueden mencionar los trabajos presentados en [342–345] para la evaluación de sujetos con DCL o EA; o el sistema CAD para la EP incluido en [346].

Métodos de regresión y análisis longitudinal

Aunque los métodos tradicionales basados en técnicas inferencia estadística resultan útiles a la hora de modelar algunas relaciones entre ciertas características de entrada procedentes de dos o más conjuntos de sujetos, los tests de hipótesis que tratan de modelar la dependencia de estas variables de entrada sobre el efecto de otras de salida, así como sus asociaciones de probabilidad, pueden resultar en muchos casos insuficientes a la hora de predecir la evolución de una enfermedad [347].

Conforme los repositorios con imágenes clínicas han ido evolucionando, e incluyendo más sujetos y más visitas en sus estudios, se ha hecho evidente la necesidad de desarrollar nuevos tipos de análisis, cada vez más eficientes

y con mayor nivel de ajuste, que permitieran modelar mejor la progresión de enfermedades como la demencia. No obstante, la implementación de estos métodos precisa de la inclusión de una dimensión temporal que conlleva ciertas dificultades: por un lado, el balanceo de las clases de entrada referido a que en muchos casos no todos los sujetos de entrada incluyen el mismo número de visitas y/o la misma cantidad de pruebas clínicas recogidas en cada una de estas visitas; y por otro lado, el problema de la alineación temporal, entendida como las condiciones que tienen que cumplir los tiempos entre visitas de cada sujeto [348].

Debido a la naturaleza de los repositorios con los que se ha trabajado en esta tesis doctoral (descritos en la Sección 3.1), no ha resultado conveniente realizar ningún tipo de análisis de regresión con base en la falta de datos para muchas de las muestras de entrada (problema del balanceo), y también en función del tiempo entre visitas de los sujetos de entrada (problema de la alineación temporal). Pese a este detalle, se han querido incluir aquí una breve relación de trabajos que sí que han tratado de caracterizar longitudinalmente la evolución de enfermedades como la EA o la EP. Por simplificar esta lista de trabajos, se ha optado por dividir esta relación en dos partes: un primer registro con propuestas de análisis longitudinal que no requieren el uso de ningún tipo de red neuronal (incluido en la Tabla 2.1¹⁸); y un segundo listado con trabajos que sí que utilizan redes neuronales (Tabla 2.2).

¹⁸El término “*Sparse learning*” hace referencia a un conjunto de métodos de aprendizaje que tratan de determinar un modelo de clasificación/regresión con el mayor rendimiento posible pero utilizando el menor número posible de características de entrada como ocurre con el modelo desarrollado en [349].

| Cita | Método principal | Descripción | Imagen | Dataset | Objetivo |
|-------|--|--|--------------|---------|---------------------------|
| [350] | Regresión con SVM | Predicción de cambios clínicos en pacientes con DCL | MRI; 18F-FDG | ADNI | Conversión DCL-EA |
| [351] | Correlación temporal y <i>Sparse learning</i> | Evaluación de la correlación temporal entre las características de entrada y modelo de regresión mediante Sparse Learning para la EA | MRI | ADNI | EA |
| [352] | Regresión con Lasso [353] | Predicción de la progresión de la EA mediante tests neuropsicológicos y biomarcadores | MRI | ADNI | Conversión Control-EA |
| [354] | <i>Sparse learning</i> con RF | Modelo para la predicción de las puntuaciones de varios tests neuropsicológicos relacionados con la EA | MRI | ADNI | EA |
| [43] | Ensemble de modelos lineales con aprendizaje bayesiano | Análisis longitudinal de varios conjuntos preseleccionados de marcadores relacionados con la EP | FP-CIT | PPMI | EP |
| [355] | Regresión con SVM | Modelo para la predicción de la conversión DCL-EA a partir de las puntuaciones de tests neuropsicológicos y de características extraídas en imagen MRI | MRI | ADNI | Conversión Control-DCL-EA |
| [356] | <i>Sparse learning</i> | Modelo de regresión lineal con sparse learning para caracterizar la conversión entre DCL y EA | MRI | ADNI | Conversión DCL-EA |
| [357] | Regresión bayesiana | Análisis longitudinal y multimodal de biomarcadores relacionados con la EA | MRI; AV45 | ADNI | Conversión Control-EA |
| [358] | RF y Elastic-Net [359] | Predicción del estado de sujetos con EP según la expresión de ciertas citoquinas | - | PPMI | EP |
| [140] | Regresión bayesiana y RF | Predicción de la progresión de la EP utilizando los resultados de varios tests neuropsicológicos | - | PPMI | EP |

Tabla 2.1: Ejemplos de trabajos de análisis longitudinal (no basados en el uso de redes neuronales) para la evaluación y/o predicción de la EP y de la EA.

| Cita | Método principal | Descripción | Imagen | Dataset | Objetivo |
|-------|---------------------------------------|---|--------------|---------|---------------------------|
| [360] | <i>Transfer learning</i> | Modelo para la predicción de la conversión DCL-EA a partir de imágenes estructurales, funcionales y de marcadores del líquido cefalorraquídeo | MRI; 18F-FDG | ADNI | Conversión DCL-EA |
| [361] | <i>Sparse learning</i> | Modelo para la identificación y seguimiento del DCL y la EA | MRI | ADNI | DCL, EA |
| [362] | CNN | Modelado longitudinal de la variación de materia gris en diferentes regiones cerebrales | MRI | ADNI | EA |
| [340] | <i>Recurrent Neural Network</i> [363] | Sistema para el análisis longitudinal de marcadores estructurales en sujetos para la clasificación de sujetos con EA | MRI | ADNI | EA |
| [364] | <i>Siamese Neural Network</i> [365] | Predicción de la sintomatología de la EA | MRI | ADNI | EA |
| [366] | <i>Transfer learning</i> | Sistema para el diagnóstico de la EA con información longitudinal | MRI | ADNI | Conversión Control-DCL-EA |
| [367] | CNN | Sistema para el diagnóstico de la EA con información longitudinal | 18F-FDG | ADNI | Conversión Control-EA |
| [368] | <i>Recurrent Neural Network</i> [363] | Predicción de la variación de marcadores MRI | MRI | ADNI | EA |
| [369] | RNN | Modelo para la predicción de la conversión DCL-EA | MRI | ADNI | Conversión DCL-EA |
| [370] | CNN | Modelo para la predicción de la conversión DCL-EA | MRI | ADNI | Conversión DCL-EA |

Tabla 2.2: Ejemplos de trabajos de análisis longitudinal (basados en el uso de redes neuronales) para la evaluación y/o predicción de la EP y de la EA.

2.5. Aplicaciones de las distribuciones alfa-estables en neuroimagen

El uso de distribuciones α -estables ha demostrado ser una herramienta muy eficaz a la hora de modelar datos con fuerte impulsividad y/o colas pesadas para aplicaciones como la ingeniería eléctrica, la economía o la astrofísica entre otros [371]. Pese al uso potencial que tienen este tipo de distribuciones en el campo de la neuroimagen [31], hasta sólo se ha propuesto su aplicación en modelos de segmentación de imágenes estructurales [372]; en el estudio de la señalización neuronal [373]; y en el desarrollo de técnicas de normalización en intensidad más eficientes respecto a técnicas basadas en un cociente de intensidades promedio entre una región específica y una región no-específica del cerebro, [374], en aplicaciones prácticas como el modelo CAD desarrollado por Anita et al. en [310] para la evaluación de la EP.

2.6. Caracterización de patrones morfológicos en neuroimagen

La búsqueda de nuevos biomarcadores que permitan modelar enfermedades neurodegenerativas constituyen un desafío no sólo para el desarrollo de nuevos sistemas de apoyo al diagnóstico más eficaces, sino que también como forma de medir cuantitativamente la pérdida de neuronas en el sistema nervioso central o cualquier otro cambio fisiológico en él [125]. Hasta ahora, muchos de los trabajos que han intentado proponer características novedosas para la caracterización de enfermedades como la EP o la EA, lo han hecho a nivel de marcadores genéticos y bioquímicos [43, 82, 375–384]. Por poner algunos ejemplos concretos, ahora sabemos que la existencia de ciertos niveles del péptido β -amiloide ($A\beta_{4,2}$), o de las proteínas α -sinucleína y τ están fuertemente relacionados con la presencia de este tipo de trastornos [385–388]. Sin embargo, y pese a todo el esfuerzo invertido, todavía no hay un claro consenso sobre cómo funcionan muchos de estos mecanismos y cómo influyen en el desarrollo de las patologías como la EP o la EA [378]. Esto, unido al coste tanto de personal especializado para la extracción de muestras, como de equipos y laboratorio para su posterior análisis, ha impulsado todavía más el interés de la comunidad científica por encontrar otros tipos de biomarcadores menos invasivos y que no requieran de una unidad de laboratorio especializada.

A nivel de imagen, y en relación al hallazgo radiológico mencionado en la Sección 2.2.2 sobre la forma que presenta la región del estriado en sujetos sanos y en pacientes con EP¹⁹, se puede entender que la propia morfología

¹⁹Las afirmaciones presentadas aquí son igualmente válidas para el estudio de la EA

de la región con más captación del radioligando I^{123} -Ioflupano pueda servir como un marcador indicativo de la progresión de la enfermedad incluso en etapas más incipientes de esta cuando todavía no hay un claro deterioro cognitivo [389–392]. De acuerdo con esta idea, la comunicación presentada en [393] (así como su ampliación publicada en [41]) propone el uso de características morfológicas extraídas a partir de la proyección de los valores de mayor intensidad de volúmenes FP-CIT SPECT sobre los planos (x, y, z) como variables discriminativas de la progresión de la EP. A partir de las conclusiones de este trabajo y relacionándolas con las defendidas en la publicación posterior con título *Classification Improvement for Parkinson's Disease Diagnosis Using the Gradient Magnitude in DaTSCAN SPECT Images* [53] donde se justifica el uso de la magnitud del gradiente como una mejora frente al análisis tradicional vóxel a vóxel (*Voxels As Features*, VAF, [394]), se llega al punto de partida marcado en esta tesis doctoral respecto a la caracterización de patrones morfológicos en neuroimagen basados en el uso original de isolíneas (con un trabajo publicado en [395]) así como de su extensión utilizando isosuperficies (Capítulo 5), como herramientas que ayudan a diferenciar sujetos sanos de pacientes con EP con dos ventajas adicionales: 1) que se reduce el número de variables necesarias para caracterizar la enfermedad (evitando el posible sobreajuste de los métodos de clasificación/regresión a los datos de entrenamiento); y 2) que la solución es físicamente interpretable a simple vista.

2.7. Análisis multivariante de marcadores heterogéneos con inteligencia artificial

Tradicionalmente, los estudios clínicos que han tratado de identificar qué características eran relevantes a la hora de caracterizar una enfermedad o un proceso biológico, han basado su criterio en el resultado de tests estadísticos univariantes como el t-Test [396]. Aunque todavía podemos encontrar muchos trabajos basados en este tipo de análisis en el campo de la medicina²⁰, cuando se quieren analizar escenarios más complejos resulta indispensable determinar cómo varios factores (independientes) pueden influir en la variabilidad de otras características (dependientes) para así conseguir un modelo más preciso de la enfermedad. A este tipo de análisis se le conoce como análisis multivariante y nos permite revelar patrones que no serían detectables por métodos univariantes [397]. Durante esta transición, la neurociencia ha pasado de reportar casos clínicos aislados, asociados a una condición anormal de los marcadores de uno o varios pacientes, a realizar

pero en ese caso referidas a la región del córtex y utilizando otro tipo de radioligandos como el 18F-florbetapir o el 18F-FDG.

²⁰Justificados en muchos casos por la necesidad de analizar un conjunto pequeño de datos de entrada o porque son más fáciles de interpretar.

mapas estadísticos cuantitativos capaces de caracterizar patrones generalizables sobre una enfermedad. Incluso con la incorporación de técnicas como el ML, que han mejorado notablemente nuestro conocimiento sobre enfermedades como la demencia y que nos pueden proporcionar herramientas precisas para el diagnóstico y pronóstico de la enfermedad [2], la aplicación práctica de estos métodos se ha visto limitada (en algunos casos) por ciertos supuestos sobre los datos de entrada que no siempre son válidos [398]. Una de estas condiciones tiene que ver con la homogeneidad de los datos de entrada. Y es que, si se quiere explicar un proceso biológico, como el de la demencia, a partir de los datos de entrada de un conjunto de sujetos de entrenamiento para predecir alguna característica en otro sujeto (o grupo de sujetos) diferente, hará falta que las variables de entrada compartan algún tipo de criterio común como, por ejemplo, que el tipo de prueba con el que se han adquirido estas variables sea el mismo. Si esta condición no se cumple, el resultado de un problema de clasificación/regresión puede quedar sujeto a la existencia de patrones arbitrarios que podrían alterar el proceso de aprendizaje contribuyendo, en consecuencia, a la falta de generalidad del sistema. Es más, incluso en los escenarios de análisis en los que se pretende combinar el aprendizaje procedente de dos o más fuentes heterogéneas, se deberá partir siempre en la medida de lo posible del análisis de cada una de estas fuentes por separado.

Por otro lado, y aunque el análisis estadístico tradicional sigue siendo el primer recurso a la hora de estudiar cualquier enfermedad así como sus variantes, la tendencia actual es que todos los modelos de análisis de datos tradicionales evolucionen a herramientas que ayuden a predecir el desarrollo de cada patología [324, 399]. Al igual que ocurrió en su momento con repositorios como ADNI (para el estudio de la EA) o como PPMI (para la caracterización de sujetos con EP) [400, 401], con la aparición de nuevas bases de datos como DIAN (Sección 3.1.3), también surge la necesidad de proponer un primer análisis con inteligencia artificial que confirme la capacidad predictiva de los biomarcadores que se recogen en el repositorio y que permita, a la comunidad científica, profundizar en la comprensión de esta variante de la EA. El desarrollo que se incluye en el Capítulo 6 trata de proponer este primer análisis aplicando técnicas de ML como las descritas en la Sección 2.4, y constituye el primer modelo predictivo sobre la enfermedad publicado hasta la fecha. Es por este motivo que, aunque en este bloque no se haya podido incluir una revisión de propuestas con inteligencia artificial sobre DIAN, sí que se incluye una comparativa en la Sección 8.1 acerca de los resultados obtenidos en esta tesis doctoral y los documentados en publicaciones anteriores basadas en el uso de técnicas de estadística tradicional como las que se han presentado en [102, 402–406].

2.8. Análisis multivariante de marcadores heterogéneos mediante aprendizaje de conjuntos

Llamamos modelo de aprendizaje de conjuntos (o *ensemble learning model*) a aquellos sistemas de aprendizaje máquina en los que se combina el entrenamiento de varios modelos más simples (bloques) para resolver un mismo problema. La hipótesis de partida de este tipo de sistemas es que cada bloque por separado tiene un sesgo (*bias*) y/o una variabilidad (grados de libertad) superiores que los que presenta el modelo de conjunto [44, 45, 407]. Aunque alcanzar una solución cuyo sesgo y varianza se reduzcan de manera simultánea puede resultar contradictorio, ya que por un lado se intenta resolver la complejidad subyacente de los datos de entrada y por otro lado se intentan reducir los grados de libertad del modelo con el fin de evitar una alta variabilidad [45], siempre se intenta que haya un buen equilibrio entre ambos. Es más, que un clasificador tenga un buen rendimiento respecto a un conjunto de datos de entrenamiento no siempre implica que vaya a generalizar bien. Y de hecho, incluso cuando se comparan varios clasificadores similares entre sí, estos pueden tener una capacidad de generalización diferente si los datos de entrenamiento no son lo suficientemente representativos. Es por todo esto que en escenarios con fuentes de datos heterogéneas, siempre se recomienda utilizar metodologías de *ensemble* para reducir la variabilidad que las fuentes de datos de entrada tienen a la salida sin incrementar el nivel de sesgo.

En la actualidad, se pueden encontrar multitud de trabajos que exploran diferentes configuraciones de *ensemble* con diferentes algoritmos de clasificación/regresión para el análisis de enfermedades como la EP o la EA [408]. Normalmente estas propuestas pueden agruparse según tres tipos de meta-algoritmos (o como una combinación de estos) [143], como los que se desarrollan a continuación:

- **Bagging** - En este modelo cada bloque de aprendizaje se entrena en paralelo (de forma independiente) y con el mismo conjunto de muestras de entrada pero utilizando distintas porciones de información [409]. Los resultados de todos los bloques se combinan siguiendo algún tipo de métrica como el promedio de resultados o el voto por mayoría (*Majority Voting*, MV). En la mayoría de casos, los modelos de *bagging* consiguen resultados con poca variabilidad gracias a la aplicación de técnicas como *bootstrapping*²¹.

Como ejemplos de modelos de *ensemble* con *bagging* caben citar varias propuestas de modelos CAD para el diagnóstico de la EP o de la EA

²¹Técnica estadística consistente en un submuestreo del conjunto inicial de datos en muestras de menor tamaño seleccionadas de forma aleatoria y con reemplazo. Cada uno de estos subconjuntos se ajustan a cada uno de los bloques de entrenamiento y sus predicciones son combinadas mediante el promediado de sus resultados.

utilizando SVM [308, 410–413]; RF [137, 289, 299, 322, 323, 414–419]; e incluso modelos de segmentación basados en clasificación con *ensemble* como en [420].

- **Boosting** - En este caso, los bloques se entrenan de manera secuencial y adaptativa de forma que los resultados de cada bloque dependen de los resultados del bloque anterior. Gracias a que cada bloque se ajusta intentando dar más importancia a los datos de entrada que hayan sido mal etiquetados en las etapas anteriores, los modelos basados en *boosting* consiguen resultados con poco sesgo. En función de cómo se realiza el cálculo de los coeficientes asociados a cada bloque de aprendizaje, se suele hablar de dos tipos de modelos de *boosting*: los modelos con estructura adaptativa (*AdaBoost*) y los modelos con estructura gradiente (*GradientBoost* en algunos trabajos) [45, 421, 422].

Sobre el análisis de datos de neuroimagen empleando técnicas de *boosting* se pueden citar trabajos que hacen uso tanto de modelos con estructura adaptativa [323, 415, 418, 423, 424], como de modelos con estructura gradiente [425–427].

- **Stacking** - En un modelo de *ensemble* con *stacking* cada bloque se entrena en paralelo (de manera independiente) y con el mismo conjunto de datos, pero combinando los resultados de todos los bloques de entrenamiento mediante algún tipo de meta-modelo (no una métrica como ocurre con los modelos de *bagging*). Esto hace que cada predicción esté basada en la predicción que se haya obtenido en etapas anteriores, y que se consigan resultados con poco sesgo. En este sentido, resulta interesante comentar que la mayor parte de los modelos con redes neuronales son en realidad modelos de *ensemble* con *stacking* [325, 361, 428–435].
- **Propuestas con modelos híbridos** - Donde destacan modelos con redes neuronales (*stacking*) que combinan su aprendizaje final mediante una metodología de MV (*bagging*) como por ejemplo en [232, 436–439]; aunque también se pueden encontrar algunas propuestas que combinan estructuras híbridas entre *boosting* y *bagging* como en [265].

Bloque C

Introducción de las bases de datos y técnicas de evaluación

Capítulo 3

Bases de datos y métodos de evaluación

Para el desarrollo y evaluación de los métodos propuestos durante la tesis doctoral se han utilizado una serie de repositorios clínicos con pruebas (principalmente de imagen estructural y/o funcional) procedentes de varios centros internacionales. Este capítulo trata de resumir las líneas más importantes de cada repositorio junto a información sobre sus demográficas (Sección 3.1) así como introducir las métricas de clasificación (Sección 3.2) a las que haremos referencia más tarde en el Bloque D.

3.1. Repositorios utilizados durante la realización de la tesis doctoral

Las bases de datos que se indican a continuación son de acceso libre o de acceso libre bajo petición previa a la institución responsable de cada una de ellas. Todos los centros afirman que estos datos cumplen con las normativas éticas y legales necesarias para su aprobación incluyendo el consentimiento informado de los sujetos incluidos en cada estudio que además cumple con los estándares y las recomendaciones éticas presentadas en la Declaración de Helsinki y Tokio de la Asamblea Médica Mundial sobre investigación clínica en seres humanos [440].

3.1.1. Parkinson's Progressions Markers Initiative (PPMI)

El repositorio del Parkinson's Progression Markers Initiative (PPMI) se encuentra disponible a través de su página web oficial (www.ppmi-info.org). La iniciativa fue creada por la fundación Michael J. Fox Foundation junto a un consorcio de entidades públicas y privadas¹, tiene como objetivo funda-

¹<http://www.ppmi-info.org/fundingpartners>.

mental identificar nuevos biomarcadores que permitan medir la progresión de la EP y desarrollar nuevos tratamientos [401, 441].

Como la base de datos está actualizándose continuamente, incluyendo nuevos participantes en el estudio², en esta sección nos referiremos solo a la actualización de abril de 2019 para una muestra balanceada con 193 participantes *de novo*³ por cada clase e independientemente de si alguno de los casos de EP ha sido etiquetado como SWEDD⁴.

| | Hombres | | Mujeres | | Ambos | |
|--------------|---------|-------------|---------|-------------|-------|-------------|
| | % | Edad | % | Edad | % | Edad |
| Sanos | 33.2 | 62.0 ± 10.7 | 16.8 | 58.9 ± 11.9 | 50 | 60.9 ± 11.2 |
| EP | 32.9 | 62.2 ± 9.9 | 17.1 | 60.3 ± 9.7 | 50 | 61.6 ± 9.9 |

Tabla 3.1: Demográficas del repositorio PPMI para un escenario con balanceo de clases. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar.

El PPMI es un estudio multicentro que hace uso de diferentes tipos de pruebas e instrumentos para realizar dichas pruebas. Aunque en esta sección no se incluirá un resumen de los equipos utilizados en cada centro, estos detalles se pueden consultar a través del portal <https://ida.loni.usc.edu/>.

3.1.2. Imágenes del Hospital Virgen de la Victoria de Málaga

Este repositorio, cedido por el Hospital Virgen de la Victoria de Málaga (Málaga, España), contiene un total de 208 imágenes FP-CIT SPECT procedentes tanto de sujetos sanos como de pacientes con EP cuyas demográficas se adjuntan en la Tabla 3.2.

| | Hombres | | Mujeres | | Ambos | |
|--------------|---------|------------|---------|-------------|-------|-------------|
| | % | Edad | % | Edad | % | Edad |
| Sanos | 52.1 | 69.9 ± 9.4 | 47.9 | 68.5 ± 11.0 | 49.7 | 69.2 ± 10.2 |
| EP | 56.8 | 68.8 ± 9.4 | 43.2 | 67.6 ± 10.8 | 50.3 | 68.3 ± 9.6 |

Tabla 3.2: Demográficas del repositorio FP-CIT SPECT del Hospital Virgen de la Victoria de Málaga. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar.

²Criterios de inclusión disponibles a través del enlace: <http://www.ppmi-info.org/wp-content/uploads/2014/06/PPMI-Amendment-8-Protocol.pdf>.

³La etiqueta *de novo* hace referencia a que ninguno de los pacientes con EP incluidos en el estudio recibió ningún tipo de tratamiento en el estudio que pudiera interferir con los resultados de sus pruebas (imagen o cualquier otro marcador clínico) [442].

⁴Pacientes que han sido diagnosticados con EP pero cuyas pruebas de imagen funcional no revelan ningún tipo de déficit dopaminérgico [443].

Todas las imágenes recogidas en el repositorio han sido realizadas tras inyectar una solución intravenosa de 185 MBq (5 mCi) de I^[123]-Ioflupano (radiotrazador) con bloqueo previo del tiroides mediante una solución de Lugol. Tras 3 – 4 horas después de la inyección, se lleva a cabo la captación del radioligando utilizando una cámara gamma modelo *General Electric Millennium* equipada con un cabezal doble y un colimador de propósito general. Para la obtención de cada exploración, se realizaron varias captaciones alrededor de una órbita circular de 360° alrededor del cráneo, a intervalos de 3°, adquiriendo un total de 60 imágenes (de tamaño 128 × 128) por detector con una duración de 35 segundos por intervalo. Una vez se obtienen todos los cortes de cada sujeto, estos se reconstruyeron utilizando un algoritmo de retroproyección filtrada (*filtered back-projection algorithm*) sin corrección de la atenuación y aplicando un filtrado de Hanning con una frecuencia de corte de 0.7.

3.1.3. Dominantly Inherited Alzheimer’s Network (DIAN)

Objetivos Con el fin de recopilar de forma estandarizada información clínica, cognitiva, de neuroimagen, de biomarcadores así como de muestras de tejidos de pacientes con DIAD y de sus familiares directos; la iniciativa DIAN (*Dominantly Inherited Alzheimer Network*) surge en 2008 como un proyecto financiado por el National Institute on Aging⁵ para establecer un registro internacional para el estudio de esta variante de la EA [109, 444]. Los objetivos principales de este proyecto son:

1. Establecer un registro internacional multicentro que realice un seguimiento de pacientes con DIAD (*Mutation-Carriers*, MC) tanto sintomáticos como asintomáticos; así como de sus familiares no portadores (*Non-Carriers*, NC) que sirven como sujetos de control.
2. Comparar sujetos NC con pacientes MC asintomáticos para determinar la evolución de sus marcadores clínicos, cognitivos, de neuroimagen, etc. a lo largo del tiempo.
3. Comparar los fenotipos clínicos y neuropatológicos de sujetos sintomáticos DIAD y LOAD.

Demográficas Cada año, la iniciativa DIAN actualiza su base de datos incluyendo nuevos sujetos, visitas y resultados. La versión del repositorio que se ha utilizado durante la realización de esta tesis doctoral ha sido la undécima (*data freeze 11th*) y sus demográficas (incluyendo los resultados de los tests neuropsicológicos *Mini Mental Status Examination test*, MMSE, [445]) se resumen en la Tabla 3.3.

⁵<https://dian.wustl.edu/about/>

| | Asintomáticos | | Sintomáticos | |
|-------|---------------|------------|--------------|-------------|
| | NC | MC | NC | MC |
| # | 156 | 159 | 12 | 103 |
| Edad | 38.3 ± 10.2 | 34.8 ± 9.2 | 40.8 ± 11.6 | 45.5 ± 10.2 |
| Sexo* | 60.3% | 57.2% | 66.7% | 54.4% |
| MMSE | 26.2 ± 1.2 | 29.0 ± 1.2 | 28.1 ± 1.6 | 22.4 ± 7.0 |

Tabla 3.3: Demográficas del repositorio DIAN. (*: la variable sexo está referida al porcentaje de mujeres en cada subconjunto).

Entre los criterios de inclusión en el estudio se incluyen como condiciones: tener más de 18 años; ser pariente directo de un individuo diagnosticado con DIAD (clínicamente o mediante pruebas genéticas); incluir en el estudio a dos familiares directos que puedan servir como sujetos de control; y hablar de manera fluida al menos uno de los idiomas aprobados por el centro de coordinación de DIAN. Por el contrario, y siguiendo las recomendaciones del centro Addenbrooke’s Hospital - Department of Psychiatry - School of Clinical Medicine - University of Cambridge⁶, con el fin de evitar hallazgos radiológicos que no estén relacionados con DIAD, se han excluido todos aquellos sujetos diagnosticados con alguna de las patologías recogidas en la Tabla 3.4 [446]. Nótese que este criterio de exclusión no se aplica a sujetos con déficit de vitamina B12, depresión, alcoholismo, abuso de drogas, historial de convulsiones, lesiones cerebrales traumáticas sin déficit/disfunción crónica ni problemas de tiroides; debido al gran número de casos reportados con al menos uno de esos diagnósticos, lo que reduce considerablemente el tamaño de la muestra. Adicionalmente, y debido al pequeño tamaño muestral (solo 13 casos) así como a la falta de resultados para la mayoría de sus pruebas, todos los casos de sujetos con un diagnóstico de LOAD también han sido descartados. Una vez aplicados todos los criterios de inclusión/exclusión propuestos para este trabajo, la muestra final quedó compuesta por un total 442 sujetos (41.6% hombres) con una edad con media y varianza de 38.7 ± 11.0 años. En esta muestra se incluyen 265 casos de MC de los cuales 202 correspondían a una mutación en el gen PSEN1, 22 en el gen PSEN2 y 43 en el gen APP.

A nivel de sintomatología, en el estudio se incluye el valor de la escala CDR (*Clinical Dementia Rating*) propuesta por Berg en 1988 [447]. Según este modelo, considerado históricamente como el estándar de facto para la cuantificación del grado de demencia de cualquier individuo [448], la sintomatología de los pacientes con demencia se divide en varios niveles que varían entre una sintomatología leve (CDR = 0.5) correspondiente a pacien-

⁶<https://www.psychiatry.cam.ac.uk/>

| | |
|--------------------------------|---|
| Accidente cerebrovascular | Demencia por alcoholismo |
| Afasia primaria progresiva | Demencia semántica |
| Afasia progresiva no fluida | Demencia transcortical |
| Ataque isquémico transitorio | Demencia vascular |
| Degeneración corticobasal | Enfermedad de Huntington |
| Demencia anómica | Enfermedad de Parkinson (EP) |
| Demencia con cuerpos de Lewy | Hidrocefalia |
| Demencia de origen desconocido | Lesión cereb. traumática (disfunc. crónica) |
| Demencia de origen priónico | Neoplasia en el sistema nervioso central |
| Demencia frontotemporal | Parálisis supranuclear progresiva |
| Demencia logopéica | Síndrome de Down |

Tabla 3.4: Listado de patologías cuyo diagnóstico positivo se considera criterio de exclusión.

tes con algún déficit cognitivo pero que todavía son autónomos, hasta un progreso muy severo de la enfermedad ($CDR = 3$) en el que los pacientes requieren atención y cuidados permanentes. La Tabla 3.5 incluye un resumen de los valores del campo CDR para los sujetos incluidos en este trabajo.

| | Asintomático | Leve | Moderado | Avanzado | Severo |
|-------|--------------|------|----------|----------|--------|
| NC | 173 | 0 | 0 | 0 | 0 |
| APP | 29 | 9 | 4 | 1 | 0 |
| PSEN1 | 122 | 51 | 20 | 5 | 4 |
| PSEN2 | 20 | 2 | 0 | 0 | 0 |

Tabla 3.5: Sujetos en cada nivel de la escala CDR según su tipo de mutación.

Biomarcadores Aunque el repositorio incluye miles de características, la cantidad de pruebas realizadas a cada sujeto (sobre todo las referentes a marcadores genéticos, de sangre y/o plasma) es reducida. Según esto, en una fase inicial de este trabajo se decidió seleccionar solo algunos de los marcadores clínicos más utilizados en el estudio y evaluación del AD. Estos marcadores se pueden agrupar en 2 categorías:

- **Marcadores que no proceden de pruebas de imagen** [*Non-Imaging Biomarkers*, NIB]. La mayoría procedentes de muestras de CSF extraídas mediante punción lumbar [449]. Incluyen, entre otros,

la expresión de los alelos⁷ del gen que codifica la apolipoproteína-E (APOE); recuentos de los péptidos $A\beta_{40}$, $A\beta_{42}$ así como el ratio entre ambos [450]; y mediciones de las proteínas τ y $p\text{-}\tau$, entre otros.

- **Marcadores de imagen** - Entre los que se incluyen:
 - Marcadores de imagen estructural [MRI]. Siguiendo los mismos procedimientos que se describen para la base de datos ADNI (Alzheimer's Disease Neuroimaging Initiative) [451], a cada sujeto se le realiza una resonancia magnética craneal mediante una secuencia en 3D de imágenes T1-MPRAGE [452], utilizando un equipo modelo *Siemens BioGraph mMR PET-MR 3T* o *Siemens Trio 3T* (dependiendo del centro donde se realiza la prueba) con un tiempo de adquisición de entre 5 – 6 min. Como resultado, se obtiene un conjunto de imágenes, con tamaño de vóxel de $1.1 \times 1.1 \times 1.2\text{mm}$, que posteriormente son procesadas con la herramienta *FreeSurfer, v5.3* para su corrección, ajuste, normalización, segmentación y el cálculo del grosor del córtex; siguiendo los procedimientos descritos en [186, 236, 241, 243, 244, 247, 248, 251–253, 453–456]. Tras este procesado, las imágenes MRI se registran al espacio de un atlas anatómico como el atlas de Talairach [250], y se determina el promedio de intensidad de cada una de las regiones anatómicas de interés parceladas por dicha plantilla. Para más detalles sobre este procesado, también se puede consultar la información disponible en <http://surfer.nmr.mgh.harvard.edu/>.
 - Marcadores de imagen funcional [FDG]. Este tipo de marcadores se utilizan para el estudio del metabolismo cerebral cuantificado a través de la captación del radioligando F18-Fluorodesoxiglucosa (18F-FDG). En concreto, a cada paciente se le inyectan unos 5 mCi de trazador 18F-FDG y a los 40 minutos de la inyección se le realiza una prueba de imagen que dura en torno a 20 minutos para evaluar la distribución del radiotrazador. A las imágenes que se van generando para cada paciente, se les aplica una corrección de movimiento y el resultado se registra espacialmente a su imagen MRI [457, 458]. Una vez generadas, cada imagen FDG se parcela utilizando el mismo mapa de transformación que el aplicado a las imágenes MRI resultando en un conjunto de valores SUVR para cada región anatómica de interés. Este procedimiento es idéntico al que se suele llevar a cabo con el repositorio ADNI [459].
 - Marcadores de imagen funcional [PIB]. Permiten analizar la distribución del β -amiloide ($A\beta$) en el cerebro gracias al radiotrazador Pittsburgh compound-B [91]. En este caso, cada imagen se genera

⁷Un alelo es cada una de las posibles versiones que tiene un gen. Un individuo hereda dos alelos para cada gen, uno del padre y el otro de la madre.

40 – 70 minutos después de la inyección con 15 mCi de ^{11}C -PiB, y el tiempo de la adquisición dura en torno a 30 minutos. A partir de aquí, el procesado que se le aplica a cada imagen es idéntico al descrito para las imágenes FDG.

3.1.4. Internet Brain Segmentation Repository (IBSR)

Disponible a través del enlace <https://www.nitrc.org/projects/ibsr/>, se trata de una base de datos fue diseñada para poder evaluar y comparar modelos de análisis de neuroimagen a nivel internacional con especial atención al desarrollo de modelos de segmentación. El conjunto de datos de este repositorio incluye 20 imágenes MRI (así como sus respectivas versiones segmentadas proporcionadas por el MGH Morphometric Analysis Center) disponibles a través del portal <http://www.cma.mgh.harvard.edu/ibsr/>. Las demográficas asociadas a este estudio han sido incluidas en la Tabla 3.6.

| | Hombres | | Mujeres | | Ambos | |
|----------------------|---------|----------------|---------|----------------|-------|----------------|
| | % | Edad | % | Edad | % | Edad |
| Participantes | 50 | 29.0 ± 5.4 | 50 | 29.1 ± 4.4 | 100 | 29.0 ± 4.8 |

Tabla 3.6: Demográficas del repositorio IBSR. El campo Edad viene dado en términos de la media de edad de los sujetos y su desviación estándar.

Aunque las imágenes proceden de dos centros diferentes (cada centro aportó 10 imágenes), las características de los equipos utilizados para su adquisición son muy similares. En el primer centro, se utilizó un equipo *Siemens Magnetom 1.5T* de 1.5 T con tiempo de relajación entre adquisiciones de 40 ms, eco de 8 ms, ángulo de inclinación de 50° , campo de visión de 30 cm y separación entre capas de 3.1 mm; mientras que en el segundo centro se utilizó un equipo modelo *General Electric 1.5T*, también de 1.5 T pero con un tiempo de relajación de 50 ms, eco de 9 ms, ángulo de inclinación de 50° , campo de visión de 24 cm y separación entre capas de 3.0 mm. El tamaño de los cortes de cada imagen fue, en ambos casos, de 256×256 .

Una vez obtenidas todas las imágenes, estas fueron registradas espacialmente a un espacio de coordenadas tridimensional definido a partir de la posición media de las comisuras anterior y posterior en el plano sagital a la altura de la comisura posterior como puntos de referencia [460]. El proceso de segmentación final fue llevado a cabo siguiendo el procedimiento descrito en [195].

3.2. Evaluación de la capacidad de aprendizaje de un modelo de clasificación

Métricas de clasificación

Como ya se comentó en la Sección 2.4.2, el aprendizaje supervisado trata de caracterizar una función de mapeo ($f(\cdot)$) que permita predecir el resultado de una variable de salida (\mathbf{Y}) a partir de unos datos de entrada (\mathbf{X}) según $\mathbf{Y} = f(\mathbf{X})$. En escenarios de clasificación, como los evaluados en esta tesis para el análisis de los trastornos neurodegenerativos más frecuentes, se utilizan matrices de confusión (con la forma descrita en la Figura 3.1) que tratan de contabilizar la cantidad (y tipo) de aciertos (y fallos) en las predicciones para así evaluar la capacidad de generalización de cada modelo predictivo.

| | | Clase estimada | |
|--------------|---|---------------------|---------------------|
| | | 1 | 0 |
| Clase actual | 1 | True positive (TP) | False negative (FN) |
| | 0 | False positive (FP) | True negative (TN) |

Figura 3.1: Ejemplo de matriz de confusión para un problema binario con etiquetas 0 y 1.

A partir de una matriz de confusión se pueden derivar muchos tipos de métricas [461]. Entre las más populares en trabajos sobre neuroimagen, destacan la tasa de acierto (o *accuracy*), que mide la proporción de aciertos de clasificación sin tener en cuenta ni el acierto de cada clase por separado ni el tipo de error que se está cometiendo respecto de las muestras mal clasificadas; la sensibilidad, que refleja la tasa de positivos (muestras pertenecientes a la clase 1) que han sido etiquetados como verdaderos; la especificidad, que hace referencia al ratio de negativos (muestras pertenecientes a la clase 0) etiquetados como verdaderos; la precisión, que indica la proporción de veces que un dato fue etiquetado como positivo; y el parámetro *balanced accuracy* que no es más que un promedio entre la sensibilidad y la especificidad. Las expresiones matemáticas de cada uno de estos parámetros, calculados a partir del número de casos incluidos en la matriz de confusión, se encuentran resumidas en la expresión (3.1).

$$\begin{aligned}
 \text{Tasa de acierto (Accuracy)} &= \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \\
 \text{Sensibilidad (Sensitivity)} &= \frac{T_P}{T_P + F_N} \\
 \text{Especificidad (Specificity)} &= \frac{T_N}{T_N + F_P} \\
 \text{Precisión (Precision)} &= \frac{T_P}{T_P + F_P} \\
 \text{Balanced Accuracy} &= \frac{\text{Sensitivity} + \text{Specificity}}{2}
 \end{aligned} \tag{3.1}$$

En aquellos escenarios en los que la importancia de un tipo de fallo de clasificación prevalezca sobre otro, habrá que prestar mayor atención a los parámetros relacionados con la sensibilidad y/o con la especificidad según convenga [153]. Esto se explica con un ejemplo práctico: supongamos un escenario hipotético en el que se desarrolla un sistema CAD para la detección de melanomas. La repercusión que puede tener en un paciente con melanoma un diagnóstico negativo (falso negativo) siempre será más grave que si a una persona sana se le diagnostica un posible tumor. En el primer caso, al paciente que está desarrollando el tumor no se le aplica ningún tipo de tratamiento, mientras que al sujeto sano pero con diagnóstico positivo se le pueden seguir realizando pruebas (como por ejemplo una biopsia) con las que confirmar su posible diagnóstico.

Curvas ROC

En escenarios de clasificación, una de las herramientas gráficas más útiles para evaluar rápidamente la relación entre la sensibilidad y la especificidad de los resultados de clasificación es el uso de las conocidas como curvas ROC (*Receiver Operating Characteristic curves*) [142]. Estas curvas constituyen una representación de la relación entre la sensibilidad y la métrica relacionada con la especificidad ($1 - \frac{T_N}{T_N + F_P}$), para diferentes umbrales de clasificación. Cuanto más cercana es la curva a la coordenada (0, 1) de la gráfica, mejor será el modelo de clasificación. El área bajo estas curvas (*Area Under the Curve*, AUC) se suele utilizar también como referencia del rendimiento global del modelo. Cuanto mayor sea este valor (más próximo a 1), mejor será el rendimiento del sistema [462].

3.3. Tests de permutación

En el contexto de un modelo de aprendizaje supervisado como los que se desarrollan en esta tesis doctoral, un test de permutación constituye un tipo de prueba no paramétrica que nos permite evaluar la significancia estadística de los resultados de clasificación/regresión generados siguiendo un esquema con validación cruzada como el representado en la Figura 2.3 [463]. Explicado de forma resumida, en cada una de las iteraciones del test de permutación, los

datos de entrada se re-etiquetan de forma aleatoria; se utilizan como entrada de nuestro modelo predictivo; y el resultado de su clasificación, se incorpora a un histograma a partir del cual se determina si la tasa de clasificación original (la que se obtendría sin re-etiquetar los datos de entrada) es estadísticamente significativa según un umbral definido a partir de la proporción de resultados de clasificación con re-etiquetado (Figura 3.2).

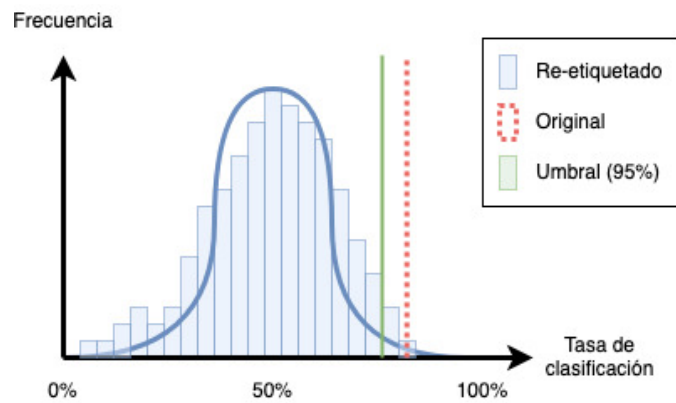


Figura 3.2: Representación gráfica de un test de permutación. Como la tasa de clasificación con las etiquetas originales (rojo) es superior al umbral (en verde, y definido en este caso particular a partir del 95 % de resultados de clasificación con re-etiquetado), entonces se considera que el resultado de clasificación es estadísticamente significativo.

Para que el test de permutación sea exacto, este debería considerar todas las posibles combinaciones de etiquetas aleatorias para los datos de entrada. En un escenario con cientos de muestras de entrada, esto constituiría un coste computacional excesivo, por lo que se suele elegir un número de iteraciones alto pero que no llega al límite marcado por el número de muestras de entrada. A esta versión aproximada del test de permutación se le suele denominar como simulación de Monte Carlo.

Bloque D

Contribuciones de la tesis doctoral. Métodos y resultados

Capítulo 4

Uso de distribuciones alfa-estables en neuroimagen

Uno de los puntos fundamentales de esta tesis doctoral es proponer nuevos métodos basados en el uso de distribuciones α -estables que puedan aplicarse al campo de la neuroimagen. En este capítulo se propone un modelo de mezclas de distribuciones α -estables (Sección 4.2), una aplicación práctica de este modelo para la segmentación de imágenes estructurales (Sección 4.3) y una justificación sobre las ventajas de utilizar estas distribuciones durante la etapa de normalización en intensidad (Sección 4.4).

4.1. Introducción a las distribuciones alfa-estables

4.1.1. Forma general

Una función de probabilidad se dice que sigue una distribución de tipo α -estable cuando su función característica ($\varphi(\omega)$) tiene una forma como la descrita en (4.1). En esta expresión se puede observar que aparecen cuatro parámetros que caracterizan a la distribución como son su exponente característico o impulsividad ($\alpha \in (0, 2]$), un parámetro de asimetría ($\beta \in [-1, 1]$), un factor de escala o dispersión ($\gamma \in (0, \infty)$) y un parámetro que representa la localización de la distribución ($\delta \in (-\infty, \infty)$) [464].

$$\varphi(\omega) = \begin{cases} e^{-|\gamma\omega|^\alpha [1 - i \operatorname{sign}(\omega)\beta \tan(\frac{\pi\alpha}{2})] + i\delta\omega} & (\alpha \neq 1) \\ e^{-|\gamma\omega|[1 + i \operatorname{sign}(\omega)\frac{2\beta}{\pi} \log(|\omega|)] + i\delta\omega} & (\alpha = 1) \end{cases} \quad (4.1)$$

Como la densidad de probabilidad $P_\varphi = \int \varphi(\omega) d\omega$ no tiene solución analítica (excepto para casos muy particulares¹), su valor tiene que determinarse mediante métodos numéricos [30].

¹Por ejemplo, cuando $\alpha = 2$, la distribución resultante es de tipo Gaussiano con desviación típica $\sigma = \frac{\gamma}{\sqrt{2}}$; cuando $\alpha = 1$ y $\beta = 0$, la distribución resultante es de tipo Cauchy; si $\alpha = 0.5$ y $\beta = 1$, entonces la función resultante será de tipo Lèvy.

4.1.2. Representación

Al disponer de más parámetros para ajustar su forma, la distribución α -estable consigue un mejor ajuste de su forma a datos de entrada sobre todo cuando la naturaleza de dichos datos es de tipo impulsivo. Para representar este ajuste, se ha incluido en la Figura 4.1 las formas de diferentes distribuciones de tipo α -estable en función de diferentes valores para cada uno de sus parámetros.

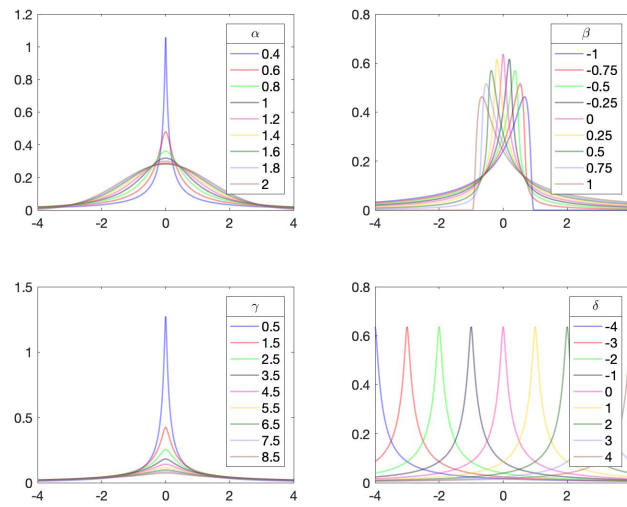


Figura 4.1: Parámetros de la distribución α -estable y su influencia en la forma de la distribución. Por defecto, $\{\alpha = 1, \beta = 0, \gamma = 1, \delta = 0\}$.

4.1.3. Estadísticos descriptivos

Aunque la media de la distribución α -estable no está definida para valores de $\alpha \leq 1$, para escenarios en los que α toma valores mayores que 1, sí que puede calcularse la media de la distribución (avg_φ) como sigue en (4.2).

$$\text{avg}_\varphi = \delta - \beta\gamma \tan\left(\frac{\pi\alpha}{2}\right) \quad (4.2)$$

De forma similar, la varianza de la distribución (var_φ) solo se puede obtener para valores de $\alpha = 2$. Como para este caso particular, la distribución α -estable es equivalente a una distribución Gaussiana, su varianza vendrá determinada por (4.3).

$$\text{var}_\varphi = 2\gamma^2 \quad (4.3)$$

4.2. Modelo de mezclas de distribuciones alfa-estables con el algoritmo EM

4.2.1. Propuesta

Un modelo de mezclas de distribuciones de probabilidad es aquel que permite estimar, a partir de una serie de datos multimodales, muestras obtenidas a partir de dos o más subpoblaciones diferentes. Matemáticamente, los modelos de mezclas de distribuciones representan una distribución de probabilidad resultado de una suma ponderada de k componentes (distribuciones de parámetros θ_j con $j = 1, 2, \dots, k$) ocultas en una mezcla como se indica en (4.4) donde $p_Y(y)$ representa la función de densidad de probabilidad del vector de datos \mathbf{y} , y donde w_j se corresponde con el peso de cada componente.

$$p_Y(\mathbf{y}) = \sum_{j=1}^k w_j f(\mathbf{y}|\theta_j) \quad \text{donde} \quad 0 \leq w_j \leq 1 \forall j \quad \sum_{j=1}^k w_j = 1 \quad (4.4)$$

Aunque los modelos más utilizados en neuroimagen son los modelos de mezclas de distribuciones gaussianas (*Gaussian Mixture Model*, GMM) [25–27, 465, 466], en la actualidad existen multitud de modelos de mezclas de distribuciones para diferentes tipos de datos de entrada [467–469]. En muchas aplicaciones reales, las componentes ocultas de estos modelos pueden estar sesgadas, ser asimétricas e incluso presentar formas de distribución con colas pesadas. Para este tipo de escenarios, una de las propuestas que consiguen un mejor ajuste de los datos de entrada está basada en el uso de distribuciones de tipo α -estable [30].

Hasta la fecha, todas las propuestas para un modelo de mezclas de distribuciones α -estables han basado la inferencia de sus parámetros en la teoría bayesiana [470, 471]. La resolución de estas propuestas mediante análisis de Monte Carlo, emplean diferentes tipos de técnicas de muestreo que consiguen un buen ajuste de los datos de entrada pero que son computacionalmente muy costosas. En este contexto, y como alternativa al método bayesiano, la publicación con título *Expectation-Maximization algorithm for finite mixture of α -stable distributions* [60], propone el uso del algoritmo de maximización de la esperanza (*Expectation-Maximization algorithm*, EM), [472], para determinar los parámetros de un modelo de mezclas finitas de distribuciones α -estables como sigue a continuación:

1. **Inicialización.** El algoritmo EM es un proceso iterativo a partir de la inicialización de los parámetros de cada una de sus k componentes. Si se desconoce el valor de k , este se puede estimar aplicando la misma teoría que se utiliza para estimar el número de componentes de un

modelo de mezclas de distribuciones gaussianas [473]. En este caso, lo normal es utilizar el Criterio de Información de Akaike [474] o el Criterio de Información Bayesiano [475]. Dado k , la inicialización de los parámetros α -estables de cada componente se puede aproximar a partir de la distribución Gaussiana equivalente calculada mediante el algoritmo GMM. Para ello, y aplicando las propiedades descritas en el apartado 4.1.3, sean μ y σ los vectores que contienen los valores de la media (μ_j) y la varianza (σ_j) de cada componente Gaussiana que haya devuelto el algoritmo GMM, los parámetros α -estables de cada componente j vendrán determinados por la expresión (4.5) bajo el supuesto inicial de que todas las componentes son simétricas ($\beta_j = 0$).

$$\theta_j = (\alpha_j, \beta_j, \gamma_j, \delta_j) = (2, 0, \sqrt{\frac{\sigma_j}{2}}, \mu_j) \quad (4.5)$$

2. **Paso de expectación.** Un modelo de mezcla de distribuciones puede modelarse como un problema de *missing data* para la estimación de sus parámetros [372]. Por tanto, sea y el vector de observación resultado de un muestreo aleatorio de k subpoblaciones diferentes, se puede definir una variable z que contiene la pertenencia de cada observación y_i (con $i = 1, 2, \dots, N$) a la clase j . Si suponemos que los valores z_i se obtienen a partir de las distribuciones, estos tomarán el resultado j con una probabilidad $p(z_i = j) = w_j$. Por tanto, y condicionalmente a los valores z_i , se considera que las observaciones y_i proceden de la subpoblación j cuando $y_i|z \sim f(\cdot|\theta_j)$. Así pues, en el paso de expectación del algoritmo EM, se calculan los N valores del vector z_i asignando cada y_i a la subpoblación j con una probabilidad igual a (4.6).

$$\frac{w_j f(y_i|\theta_j)}{\sum_{j=1}^k w_j f(y_i|\theta_j)} \quad (4.6)$$

3. **Paso de maximización.** En este caso, una vez determinadas las etiquetas $z_i = j$, se estima el peso w_j de cada componente según (4.7). En esta expresión, $\{\#z_i = j\}$ representa el número de veces que z_i tiene como valor el índice de la componente j . Esto es, cada peso se calcula a partir de la proporción de las muestras $z_i = j$ en N .

$$w_j = \frac{1}{N} \{\#z_i = j\} \quad (4.7)$$

Como la función de densidad de probabilidad de una distribución α -estable no puede calcularse mediante métodos analíticos, los parámetros θ_j de cada componente se calculan numéricamente según el método propuesto en [30].

Una vez calculados θ_j y w_j para cada componente, se ajusta una función de verosimilitud, \mathcal{L} mediante el problema de optimización descrito en [476] donde se toma como referencia el logaritmo de la función de verosimilitud indicada en 4.8 y los valores $\{\#z_i = j\}$.

$$\mathcal{L}(\theta|\mathbf{y}) = \sum_i^N \log f(y_i|\theta_j) \quad (4.8)$$

Las etapas de expectación y maximización se repiten de manera indefinida hasta que se alcanza una convergencia para la pertenencia de los valores $z_i = j$ a cada componente.

4.2.2. Experimentos para justificar el modelo de mezclas de distribuciones alfa-estables

El modelo de mezclas de distribuciones α -estables presentado en [60] ha sido evaluado utilizando varios conjuntos de datos procedentes tanto de datos sintéticos –con parámetros $\theta_j = \{\alpha_j, \beta_j, \gamma_j, \delta_j\}$ y pesos w_j inicialmente conocidos– como de procesos físicos naturales –con parámetros a priori desconocidos–. Este apartado trata de resumir los resultados más destacados de este modelo frente a su equivalente con distribuciones gaussianas.

Modelo con 2 componentes α -estables sintéticas En la Tabla 4.1 se han incluido los valores de los parámetros reales y los valores de los parámetros estimados del modelo de mezclas de distribuciones α -estables representado en la Figura 4.2. Como puede observarse, la estimación basada en distribuciones α -estables consigue un ajuste preciso del histograma gracias a su capacidad para modelar las muestras más alejadas del centro del conjunto de datos. Por el contrario, el modelo de mezclas con distribuciones gaussianas no consigue este nivel de ajuste debido a la sobreestimación de la desviación estándar de una de las componentes del modelo.

Cuando la longitud del vector de observación es pequeña y la distribución tiene muchos picos, un cambio leve en la asignación inicial a cada componente (sobre todo cuando hay valores que están muy alejados del grueso de la distribución) puede afectar drásticamente a la estimación de los parámetros de cada componente. Este efecto se acentúa más al utilizar el algoritmo EM ya que todavía no se dispone de información previa sobre estos parámetros en el modelo bayesiano. Sobre la base de esto, y tal y como se justifica en la comparación los métodos de estimación de parámetros², se recomienda en la etapa de maximización que el ajuste de los parámetros de cada distribución se realice siguiendo el método de máxima verosimilitud.

²Para este trabajo se han comparado los métodos de la función característica empírica descrito en [477, 478], el método de los cuartiles que aparece en [479] y el método de máxima verosimilitud desarrollado en [476].

| Parámetro | Valor real | Valor inicial | Valor estimado |
|------------|------------|---------------|----------------|
| α_1 | 1.40 | 2.00 | 1.48 |
| β_1 | 1.00 | 0.00 | 1.00 |
| γ_1 | 5.00 | 3.77 | 4.87 |
| δ_1 | 0.00 | 0.42 | 0.47 |
| w_1 | 0.50 | 0.50 | 0.35 |
| α_2 | 1.80 | 2.00 | 1.80 |
| β_2 | 0.00 | 0.00 | -0.01 |
| γ_2 | 10.0 | 14.8 | 10.1 |
| δ_2 | 40.0 | 34.7 | 39.4 |
| w_2 | 0.50 | 0.65 | 0.52 |

Tabla 4.1: Mezcla de 2 distribuciones α -estables.

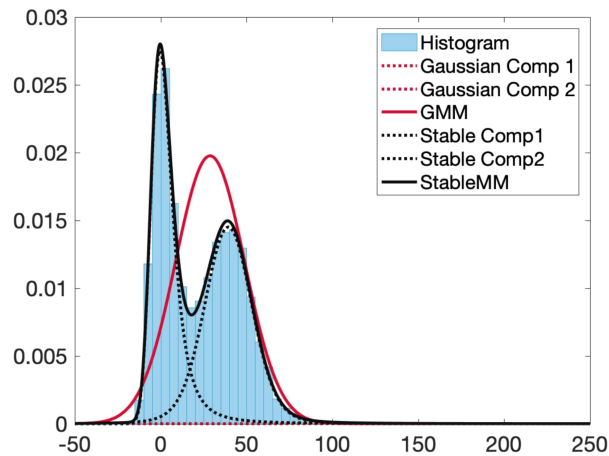


Figura 4.2: Histograma generado a partir de dos distribuciones α -estables con 10000 puntos. Las líneas continuas representan la predicción del conjunto (suma de componentes) mientras que las líneas discontinuas representan cada una de las componentes por separado.

Modelo con 3 componentes α -estables sintéticas Repitiendo el proceso anterior pero esta vez utilizando 3 distribuciones α -estables con parámetros como los que se recogen en la Tabla 4.2, la estimación de las componentes queda reflejada como sigue en la Figura 4.3. Como puede comprobarse, de nuevo el modelo con componentes gaussianas ajusta peor a la forma de las distribuciones debido a los datos más alejados del centro de cada distribución.

| Parámetro | Valor real | Valor inicial | Valor estimado |
|------------|------------|---------------|----------------|
| α_1 | 1.40 | 2.00 | 1.53 |
| β_1 | 1.00 | 0.00 | 1.00 |
| γ_1 | 5.00 | 4.00 | 4.68 |
| δ_1 | 0.00 | 0.54 | 0.08 |
| w_1 | 0.33 | 0.25 | 0.31 |
| α_2 | 1.80 | 2.00 | 1.76 |
| β_2 | 0.00 | 0.00 | 0.08 |
| γ_2 | 10.0 | 19.1 | 11.3 |
| δ_2 | 40.0 | 58.4 | 40.5 |
| w_2 | 0.33 | 0.39 | 0.38 |
| α_3 | 1.60 | 2.00 | 1.62 |
| β_3 | 0.00 | 0.00 | 0.18 |
| γ_3 | 8.00 | 21.1 | 7.60 |
| δ_3 | 80.0 | 52.9 | 80.7 |
| w_3 | 0.34 | 0.36 | 0.31 |

Tabla 4.2: Mezcla de 3 distribuciones α -estables.

Modelos de mezclas en procesos físicos naturales El modelo de mezclas de distribuciones α -estables con el algoritmo EM también ha sido evaluado con datos procedentes de procesos físicos naturales. Uno de los ejemplos incluidos en [60] es el modelado (representado en la Figura 4.4) de la incidencia del coronavirus SARS-CoV-2 en Corea del Sur entre los meses de diciembre y abril³ [481]. Como puede comprobarse, el ajuste utilizando distribuciones α -estables y su equivalente con distribuciones gaussianas es idéntico en ambas componentes (de ahí que ambas propuestas aparezcan solapadas en la figura). Esto se explica por la falta de colas pesadas en ambas distribuciones y la perfecta simetría de los datos en torno al pico de contagios de finales de febrero y la estabilización de las infecciones durante el mes de marzo que implican que $\alpha \approx 2$ y que $\beta \approx 0$ en ambos casos. Como estos valores implican el caso límite en el que las distribuciones α -estables pueden ser aproximadas mediante distribuciones gaussianas, el ajuste de ambas resulta equivalente.

³Corea del Sur fue uno de los primeros países en contener la primera oleada de la epidemia y los datos recogidos entre el 20 de diciembre de 2019 y el 25 de abril de 2020 sirvieron como experiencia para comparar el efecto de las medidas de contención llevadas a cabo en el país y así poder extenderlas a otros países [480].

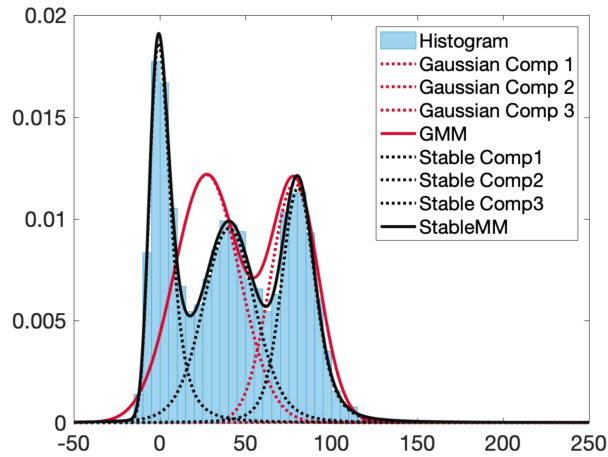


Figura 4.3: Histograma generado a partir de tres distribuciones α -estables con 10000 puntos. Las líneas continuas representan la predicción del conjunto (suma de componentes) mientras que las líneas discontinuas representan cada una de las componentes por separado.

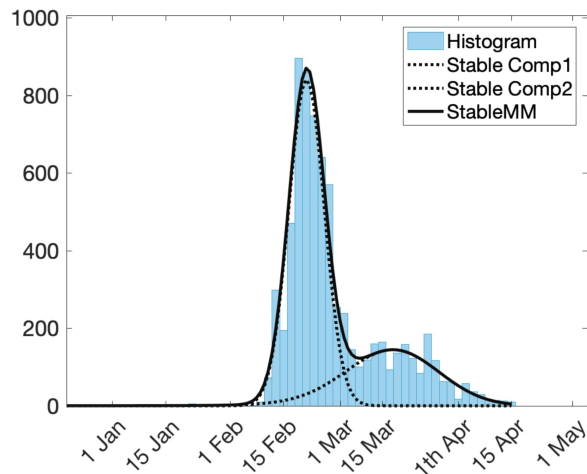


Figura 4.4: Número total de casos confirmados de infección por SARS-CoV-2 en Corea del Sur durante el periodo comprendido entre el 20 de diciembre de 2019 y el 25 de abril de 2020. La línea continua representa la mezcla de distribuciones mientras que las líneas discontinuas hacen referencia a cada componente por separado.

4.3. Aplicación del modelo de mezclas de distribuciones alfa-estables a la segmentación de imágenes estructurales

4.3.1. Justificación de la propuesta

En el contexto de la neuroimagen, si nos centramos en los modelos de segmentación que trabajan a nivel de histograma, la totalidad de los modelos revisados para la tesis doctoral (ver Sección 2.4.3) han partido de la suposición de que la distribución de intensidades en imagen estructural de materia blanca, materia gris y líquido cefalorraquídeo siguen una distribución de probabilidad gaussiana [25, 27, 202, 203, 205–207, 220, 221]. Aunque suponer un modelo de mezclas de distribuciones gaussianas puede estar justificado por el Teorema Central del Límite, en la práctica vemos que este modelo no siempre se ajusta correctamente a los datos de entrada o no lo hace con la suficiente precisión. Tal y como se observa en la Figura 4.5, las distribuciones de intensidad de materia blanca, materia gris y líquido cefalorraquídeo no son simétricas y presentan colas pesadas. Este hecho, puesto en relieve en trabajos como [215], permite entender por qué muchas de las propuestas de segmentación a nivel de histograma no han sido capaces de conseguir un buen ajuste de los datos de entrada dando como resultado soluciones con poco nivel de detalle especialmente en escenarios con ruido.

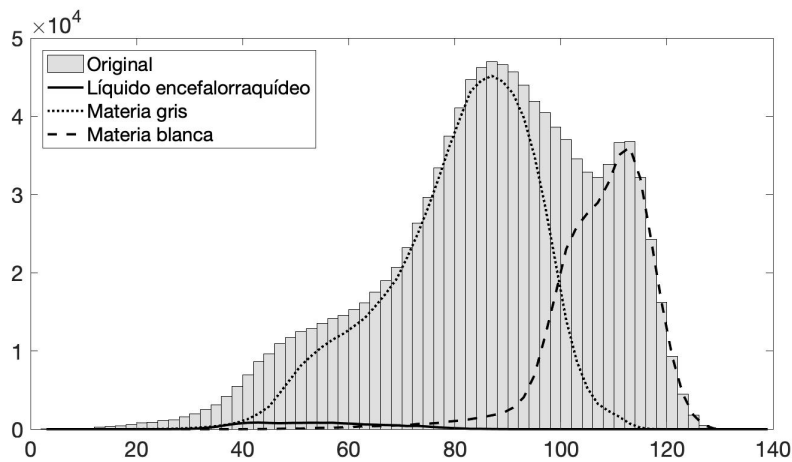


Figura 4.5: Distribución de intensidades de una imagen MRI perteneciente al repositorio IBSR.

Una posible solución al problema descrito en [215] pasaría por utilizar otro tipo de distribución que se ajuste mejor a los datos de entrada. Basándose en esto, en [372] los autores propusieron el uso de una mezcla de distribuciones α -estables para la mejora de los modelos de segmenta-

ción tradicionales ya que este tipo de distribuciones caracterizaban mejor la impulsividad y asimetría de los datos de entrada. Pese a la mejora en los resultados del nuevo modelo propuesto, el propio trabajo llevaba a pensar en la posibilidad de incluir algún tipo de información espacial que pudiera aprovechar mejor la información contenida en las relaciones de intensidad entre vóxeles vecinos para así mejorar el resultado del ajuste del modelo.

4.3.2. Modelo propuesto para la mejora de la segmentación de imágenes estructurales

En el contexto descrito anteriormente, y siguiendo la idea proporcionada en [206], el trabajo con título *A Heavy Tailed Expectation Maximization Hidden Markov Random Field Model with Applications to Segmentation of MRI* [51] propone un modelo de segmentación basado en el uso de campos ocultos de Markov (*Hidden Markov Random Fields*, HMRF) en combinación con el algoritmo EM modelando cada una de las componentes (materia blanca, materia gris y líquido cefalorraquídeo) mediante el uso de distribuciones α -estables. Gracias a HMRF, el modelo matemático que rige la segmentación basada en mezcla de distribuciones, también es capaz de modelar la relación entre vóxeles vecinos (información espacial) consiguiendo un mejor ajuste de la segmentación y mayor protección frente al ruido [215].

Base matemática del modelo La base matemática de la propuesta de segmentación de imágenes estructurales desarrollada para esta tesis doctoral puede resumirse como sigue a continuación:

- Dada una imagen estructural con valores de intensidad \mathbf{Y} sobre la que se quiere determinar la pertenencia de cada voxel (y_i con $i = 1, \dots, N$) a cada tejido del cerebro, el primer paso de la propuesta presentada en [51] consiste en la definición de una nueva matriz \mathbf{Z} con la misma dimensión que \mathbf{Y} y cuyas realizaciones indican la pertenencia de cada vóxel a un tejido concreto⁴.

Asumiendo que dicha pertenencia depende de un criterio de maximización a posteriori (*Maximum A Posteriori probability*, MAP), el objetivo del problema se convierte en determinar una matriz de etiquetas $\hat{\mathbf{Z}}$ (mapeo objetivo) a partir de la probabilidad conjunta de la observación ($P(\mathbf{Y}|\mathbf{Z})$) y de la probabilidad a priori de cada clase ($P(\mathbf{Z})$) según (4.9).

$$\hat{\mathbf{Z}} = \operatorname{argmax}(P(\mathbf{Y}|\mathbf{Z})P(\mathbf{Z})) \quad (4.9)$$

⁴0 para CSF, 1 para GM y 2 para WM aunque por simplicidad el trabajo solo se centra en la estimación de GM y WM.

- A nivel de histograma, el término $P(\mathbf{Y}|\mathbf{Z})$ puede modelarse asumiendo que los valores de intensidad de la imagen siguen un modelo de mezclas de distribuciones de tipo α -estable [372]. Para determinar los parámetros de cada componente ($\theta_j = \{\alpha_j, \beta_j, \gamma_j, \delta_j\}$), se identifican todos los puntos y_i en \mathbf{Y} que cumplan que $z_i = j$, y se ajustan mediante una función de verosimilitud como la descrita en [476] (algoritmo EM).
- Respecto a la información espacial, esta se puede modelar con MRF para $P(\mathbf{Z})$ y $P(\mathbf{Y}|\mathbf{Z})$ según las expresiones (4.10) y (4.11):

$$P(\mathbf{Z}) = \frac{1}{\kappa} \exp(-U(\mathbf{Z})) \quad (4.10)$$

$$P(\mathbf{Y}|\mathbf{Z}) = \frac{1}{\kappa} \exp(-U(\mathbf{Y}|\mathbf{Z})) \quad (4.11)$$

En estas expresiones, el parámetro κ representa una constante de normalización y la función de energía U que permite introducir la información espacial en el modelo EM mediante la igualdad descrita en (4.12).

$$U(\mathbf{Z}) = \sum_{c \in C} V_c(\mathbf{Z}) \quad (4.12)$$

El término $V_c(\mathbf{Z})$ (también conocido como función potencial) constituye una función definida para un conjunto de vóxeles vecinos c sobre el conjunto total de vóxeles en la imagen ($c \in C$) donde el factor β_C funciona como un coeficiente que modela la contribución de energía de cada *clique*⁵, como el representado en la Figura 4.6, sobre la estimación de la clase del voxel objetivo.

$$V_c(z_i, z_j) = \beta_C(1 - I_{z_i, z_j}) \quad \text{con} \quad I_{z_i, z_j} = \begin{cases} 0 & \text{si } z_i \neq z_j \\ 1 & \text{si } z_i = z_j \end{cases} \quad (4.13)$$

Conforme aumenta el valor de β_C , aumenta la contribución que los vóxeles vecinos tienen sobre la estimación de la clase del vóxel objetivo.

- Una vez modelada la contribución de la información espacial, la probabilidad conjunta del modelo quedará descrita según (4.14) siendo los parámetros θ_j dependientes de la variable \mathbf{Z} (los parámetros θ_j se calculan para todos los puntos de \mathbf{Z} que cumplen $z_i = j$).

⁵La palabra *clique* se utiliza frecuentemente en teoría de grafos para designar a un conjunto de vértices adyacentes e interconectados entre sí en un grafo no dirigido.

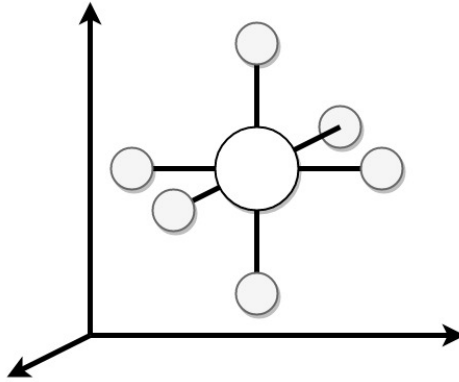


Figura 4.6: Esquema representativo de un voxel (blanco) y sus 6 véxeles vecinos (gris) en un espacio 3D.

$$P(\mathbf{Y}|\mathbf{Z}) = \prod_{i=1}^N f(y_i|\theta_j) \quad (4.14)$$

- Como κ constituye una constante de normalización del modelo, el logaritmo de la función de verosimilitud del modelo EM se puede relacionar con la energía del modelo MRF (expresión (4.15)) y la estimación máxima a posteriori del modelo pasa a ser equivalente a minimizar su energía a posteriori.

$$\log(P(\mathbf{Y}|\mathbf{Z})) \propto -U(\mathbf{Y}|\mathbf{Z}) \quad (4.15)$$

4.3.3. Experimentos para cuantificar la mejora de los resultados de segmentación

La propuesta para la segmentación de imágenes estructurales basada en el uso de los algoritmos EM y HMRF ha sido evaluada utilizando 18 imágenes MRI del repositorio IBSR. La Tabla 4.3 muestra los resultados de aplicar la metodología propuesta con distribuciones α -estables y su comparación empleando distribuciones gaussianas. En la tabla también se incluyen los valores de los coeficientes de similitud de Sørensen-Dice (frecuentemente escrito como DICE), explicados en [482], para los tejidos GM y WM. Tal y como se puede comprobar, el uso de distribuciones α -estables mejora la segmentación de imágenes respecto al uso de distribuciones gaussianas en la mayoría de los casos. La mediana de la tasa de acierto del modelo con distribuciones α -estables fue del 91.4 % mientras que en el caso de utilizar distribuciones gaussianas se alcanza una mediana del 89.4 %. Adicionalmente, en la Figura 4.7, se ha incluido también una comparativa entre la segmentación propuesta del IBSR y el modelo EM-HMRF con distribuciones α -estables.

| Tasa de acierto Stable | Tasa de acierto Gauss. | GM Dice Stable | GM Dice Gauss. | WM Dice Stable | WM Dice Gauss. |
|------------------------------|------------------------------|----------------------|----------------------|----------------------|----------------------|
| 90.1 % | 89.3 % | 0.847 | 0.829 | 0.927 | 0.926 |
| 91.7 % | 92.0 % | 0.899 | 0.893 | 0.920 | 0.928 |
| 89.5 % | 89.4 % | 0.832 | 0.830 | 0.929 | 0.928 |
| 91.6 % | 91.3 % | 0.871 | 0.861 | 0.939 | 0.940 |
| 90.0 % | 89.0 % | 0.846 | 0.822 | 0.924 | 0.927 |
| 91.6 % | 86.6 % | 0.869 | 0.781 | 0.930 | 0.926 |
| 92.0 % | 91.4 % | 0.870 | 0.854 | 0.940 | 0.943 |
| 91.4 % | 88.1 % | 0.840 | 0.789 | 0.944 | 0.934 |
| 89.3 % | 89.1 % | 0.827 | 0.822 | 0.930 | 0.930 |
| 91.9 % | 90.0 % | 0.845 | 0.813 | 0.946 | 0.944 |
| 91.2 % | 89.9 % | 0.854 | 0.833 | 0.941 | 0.939 |
| 91.5 % | 89.3 % | 0.838 | 0.801 | 0.947 | 0.941 |
| 89.0 % | 89.0 % | 0.814 | 0.812 | 0.929 | 0.930 |
| 92.5 % | 92.6 % | 0.903 | 0.902 | 0.932 | 0.934 |
| 91.9 % | 92.3 % | 0.918 | 0.913 | 0.914 | 0.924 |
| 87.9 % | 88.1 % | 0.844 | 0.853 | 0.899 | 0.896 |
| 91.0 % | 90.9 % | 0.888 | 0.898 | 0.919 | 0.910 |
| 87.3 % | 87.3 % | 0.858 | 0.859 | 0.880 | 0.878 |

Tabla 4.3: Porcentaje de vóxeles correctamente etiquetados según el modelo de segmentación propuesto con distribuciones α -estables y su comparación utilizando distribuciones gaussianas.

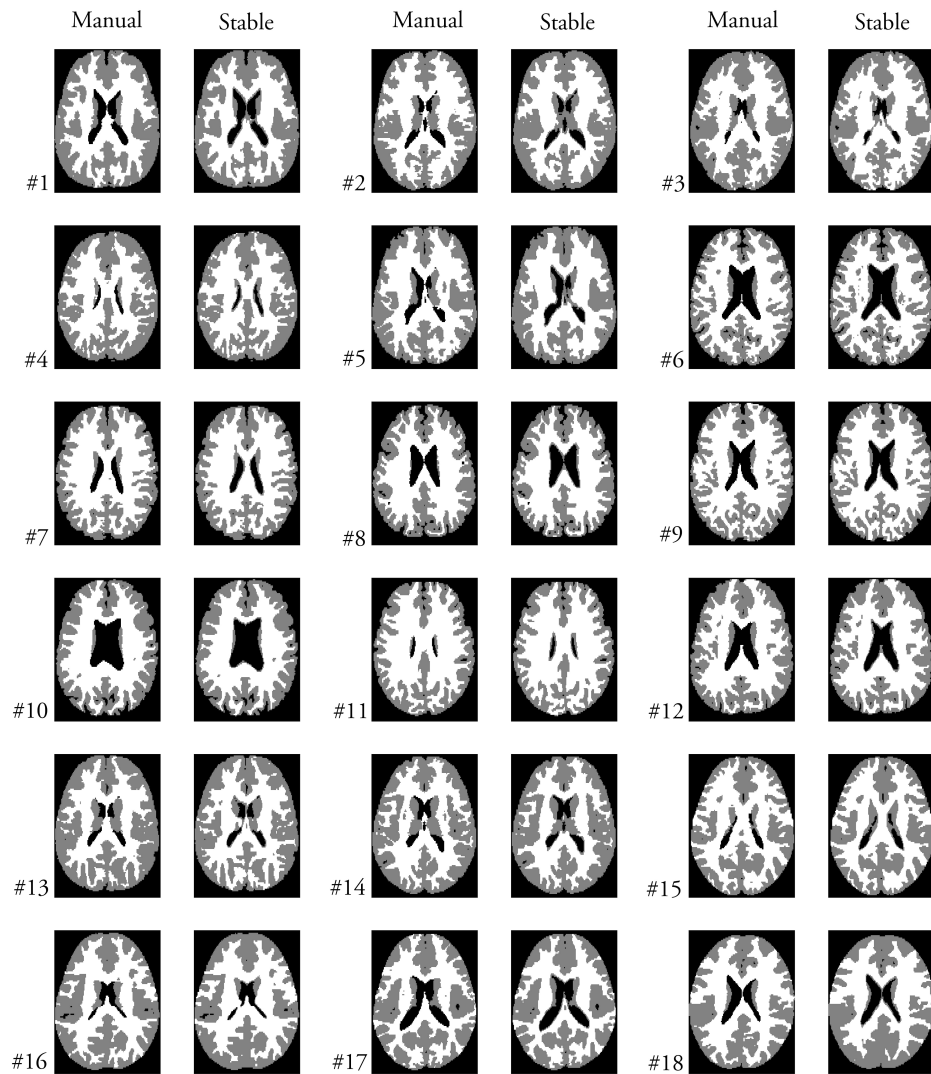


Figura 4.7: Segmentación manual del repositorio IBSR y su comparación utilizando el modelo propuesto de segmentación.

4.4. Normalización en intensidad de imágenes funcionales con distribuciones alfa-estables

Una de las alternativas al método de normalización en intensidad con BR está basada en el uso de distribuciones α -estables [374]. Según este método, y bajo la suposición de que los valores de intensidad para el conjunto de imágenes funcionales a normalizar proceden de variables aleatorias de tipo α -estable con parámetros $\theta_i = (\alpha_i, \beta_i, \gamma_i, \delta_i)$ con $i = 1, \dots, N$ (siendo N el número total de sujetos a normalizar), cada imagen transformada pasa a tener una distribución de intensidad también de tipo α -estable pero de parámetros fijos $\theta_T = (\alpha_T, \beta_T, \gamma_T, \delta_T)$ según la transformación lineal descrita en (4.16), con γ^* y μ^* calculados como el promedio de dispersión y localización del conjunto completo de imágenes.

$$Y = aX + b = \frac{\gamma^*}{\gamma}X + \mu^* - \frac{\gamma^*}{\gamma}\mu \quad (4.16)$$

Siguiendo el procedimiento descrito en la comunicación con título *On a Heavy-Tailed Intensity Normalization of the Parkinson's Progression Markers Initiative Brain Database* [52], se han representado en la Figura 4.8 las distribuciones de intensidad de un conjunto de imágenes del repositorio PPMI y su equivalente transformado siguiendo el método de normalización en intensidad con distribuciones α -estables. Además de la representación incluida en la Figura 4.9, también se ha incluido el corte axial de cada uno de los volúmenes del repositorio PPMI evaluados antes y después de aplicar esta normalización.

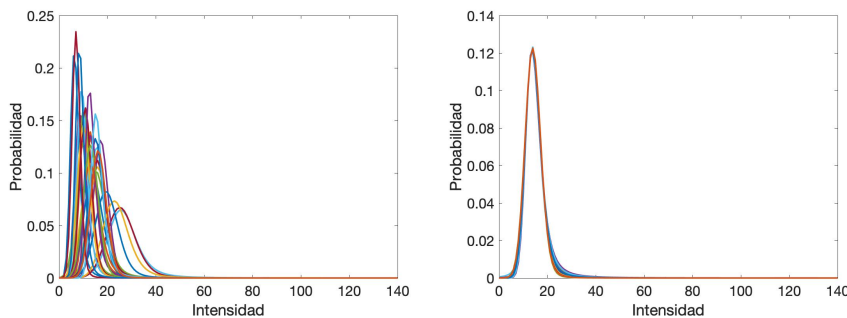


Figura 4.8: Distribuciones de intensidad de 30 imágenes FP-CIT SPECT del repositorio PPMI antes de aplicar ningún tipo de normalización en intensidad (izquierda) y después de aplicar la normalización en intensidad con distribuciones α -estables (derecha).

En el trabajo publicado en [179], los autores compararon el método de normalización en intensidad usando distribuciones α -estables con propuestas como la normalización por media y varianza (*bare normalization*), la normalización integral [171], el método del BR, y la normalización al máximo,

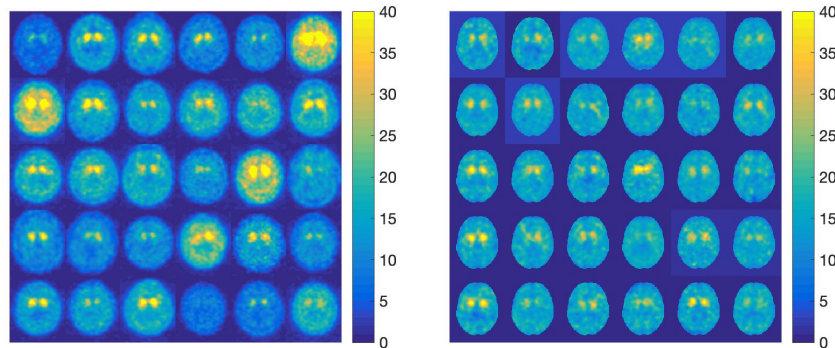


Figura 4.9: Vista axial de las 30 imágenes FP-CIT SPECT del repositorio PPMI antes de aplicar ningún tipo de normalización en intensidad (izquierda) y después de aplicar la normalización en intensidad con distribuciones α -estables (derecha).

entre otros. En todos los escenarios evaluados, la normalización en intensidad con distribuciones α -estables fue la propuesta que consiguió mejores resultados en términos del análisis estadístico propuesto con métodos como la distancia Euclídea, de Hellinger o la divergencia de Kullback-Leibler; para la mayoría de escenarios propuestos. Más tarde, y siguiendo esta propuesta, el trabajo presentado en [52] determinó que la separación entre clases de sujetos sanos y pacientes con EP mediante el método no-paramétrico Mann-Whitney-Wilcoxon U-Test [149], era también más amplia cuando se aplicaba la normalización en intensidad con distribuciones α -estables que cuando no se aplicaba ningún otro tipo de normalización. Para justificar esta afirmación, se ha incluido en la Figura 4.10 una representación del mismo tipo de experimento que el propuesto en [52] aumentando el número de participantes (en este caso 194 sujetos sanos y 432 sujetos con EP usando la primera visita de cada sujeto incluido en el repositorio PPMI) y aplicado una normalización espacial de tipo afín⁶ en vez de la normalización no-lineal propuesta anteriormente.

⁶Imágenes registradas a una plantilla generada *ad-hoc* siguiendo el procedimiento descrito en [483] y registradas a esta plantilla utilizando *SPM12*.

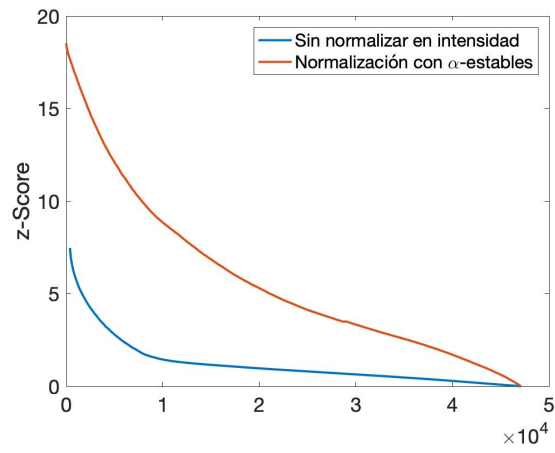


Figura 4.10: Valores z_{Score} , calculados mediante el algoritmo Mann-Whitney-Wilcoxon U-Test y ordenados en sentido decreciente, para los 47104 valores de intensidad de la región del estriado de 626 sujetos incluidos en el repositorio PPMI.

Capítulo 5

Extracción de características morfológicas en neuroimagen

En este capítulo se desarrolla la propuesta para la caracterización de la EP mediante isosuperficies calculadas sobre imagen funcional FP-CIT SPECT. Esta idea surge a partir de un análisis previo de la magnitud del gradiente en imagen funcional (Sección 5.1.1), y su propuesta e implementación (Secciones 5.1.2 y 5.1.3) proporciona mejores resultados que el análisis tradicional vóxel a vóxel (Sección 5.1.4) con la ventaja añadida de que el modelo es totalmente interpretable desde el punto de vista físico.

5.1. Caracterización del Parkinson mediante isosuperficies

Aunque la tendencia actual de la mayoría de ensayos clínicos es determinar nuevas alteraciones genéticas o desarrollar nuevas moléculas capaces de inhibir alguno de los mecanismos de la demencia, la búsqueda de patrones de imagen estructural y funcional puede contribuir a identificar nuevos mecanismos y a caracterizar los ya conocidos [391].

5.1.1. Análisis de la magnitud del gradiente en imagen funcional

Como ya se mencionó en la Sección 2.6, la comunicación presentada en [53] propuso por primera vez el uso de la magnitud del gradiente (calculado mediante el algoritmo de Sobel-Feldman¹ descrito en [484]) en imágenes FP-CIT SPECT, como un indicador más fiable de la progresión de la EP que la propia imagen funcional original (Figura 5.1). El esquema de clasificación presentado hizo primero uso de una etapa de preprocesado en la que las

¹*Sobel-Feldman Operator Algorithm*, (SOA).

imágenes se co-registraron mediante una normalización no-lineal utilizando *SPM12*; luego se aplicó una normalización en intensidad mediante el método descrito en la Sección 4.4 con distribuciones α -estables; y, al final de esta etapa, se delimitó una región de interés centrada en torno a la región del estriado pero promediado entre todas las imágenes. En el bucle de validación cruzada de tipo *Leave-One-Out* (LOO), [141], se seleccionaron los valores de intensidad más significativos según el test estadístico de Mann-Whitney-Wilcoxon, [149], y se emplearon clasificadores de tipo SVM con *kernel* lineal [49, 485].

Aunque el estudio sólo se centró en un análisis a nivel de vóxeles de intensidad, comparando los vóxeles de la zona del estriado y un pequeño margen que permitiera cuantificar correctamente el módulo del gradiente de cada imagen, los resultados alcanzados arrojaron una mejora en la clasificación del 1.25 % en términos del parámetro tasa de acierto cuando se utilizan las imágenes con la magnitud del gradiente. En detalle, los resultados de la clasificación comparando 641 sujetos del repositorio PPMI (incluyendo sujetos sanos y pacientes con EP) quedaron como se indica en la Tabla 5.1.

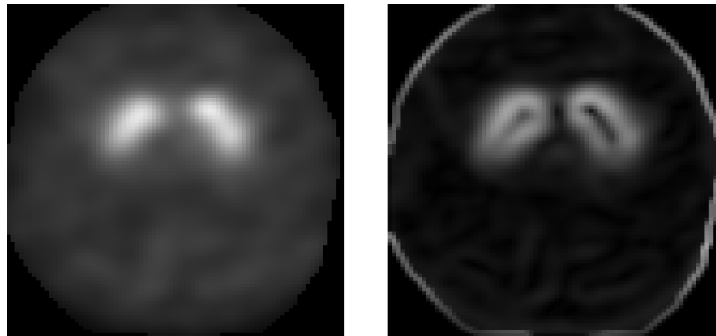


Figura 5.1: Ejemplo del corte axial de una imagen FP-CIT SPECT del repositorio PPMI (izquierda) y la magnitud del gradiente de la imagen (derecha).

| Comparación | Imagen | Accuracy | Sensitivity | Specificity | Precision |
|--------------------|-----------|----------|-------------|-------------|-----------|
| Sanos vs EP | Original | 95.32 % | 96.42 % | 92.78 % | 96.85 % |
| | Gradiente | 96.57 % | 97.32 % | 94.85 % | 97.75 % |

Tabla 5.1: Resultados de la clasificación utilizando la magnitud del gradiente de la imagen funcional en lugar de la imagen original.

Sobre la base de la forma que presenta la región del estriado en la imagen con el gradiente y a los resultados de clasificación para este trabajo, no sorprendió comprobar que las diferencias de intensidad entre la región del estriado son más pronunciadas en pacientes sanos que en pacientes con EP. Es por esto que las diferencias entre vóxeles contiguos tendrían un mayor

impacto en la separación entre clases que los propios valores de intensidad originales. Por tanto, y en relación con lo que se desarrollará en el siguiente apartado, la cuantificación de la forma del estriado en función de la intensidad de la imagen o de la diferencia de intensidades en él (gradiente), constituiría un conjunto de características (morfológicas) claramente indicativas del desarrollo y/o evolución de la enfermedad.

5.1.2. Propuesta para el análisis de imágenes FP-CIT SPECT mediante isosuperficies para caracterizar la EP

Un método que permite parcelar regiones de intensidad, para posteriormente cuantificar su forma, es el uso de isosuperficies (entendidas como una generalización de las isolíneas pero en un espacio tridimensional) [486]. Como todos los puntos que pertenecen a una misma isosuperficie tienen el mismo valor de intensidad, es posible caracterizar espacialmente mediante isosuperficies la distribución de transportadores dopaminérgicos en el cerebro. No obstante, la forma de modelar el umbral de intensidad de cada isosuperficie requiere de un análisis previo. A la vista de esto, la propuesta con título *Morphological Characterization of Functional Brain Imaging by Isosurface Analysis in Parkinson's Disease* [59], desarrolla un sistema CAD basado en la caracterización de isosuperficies obtenidas a partir de imágenes funcionales de tipo FP-CIT SPECT para el diagnóstico de la EP con las siguientes ventajas:

1. Evita la necesidad de aplicar ningún tipo de normalización en intensidad.
2. Sólo precisa de un registro afín que haga comparables en posición y tamaño las características morfológicas de cada volumen FP-CIT SPECT.
3. Evita la necesidad de definir una ROI.
4. Es físicamente interpretable ya que las isosuperficies modelan la distribución espacial de transportadores dopaminérgicos.
5. Se trata de una solución completamente automática.

5.1.3. Implementación del modelo

Algoritmo LMCA Las isosuperficies utilizadas en este trabajo han sido generadas a partir de la última versión del algoritmo *Marching Cubes* (también conocido como *Lewiner Marching Cubes Algorithm*, LMCA) publicado en [487] como una generalización de las obras publicadas en [488, 489]. El procedimiento que sigue este trabajo es la generación de superficies triangulares a partir de una función escalar dada por conjunto patrones como

los representados en la Figura 5.2. La versión actual del algoritmo LMCA utiliza un total de 28 ordenaciones posibles².



Figura 5.2: Ejemplo con 4 de los patrones del algoritmo Marching Cubes.

Para cada valor de intensidad (umbral) que define el límite de una isosuperficie, la función LMCA genera una malla de superficies 2D a partir del volumen (3D) que tiene como frontera ese umbral de intensidad. Al utilizar esta idea sobre volúmenes FP-CIT SPECT, el uso de isosuperficies nos permite cuantificar la distribución de transportadores dopaminérgicos en el cerebro tal y como se muestra en la Figura 5.3 para una sola superficie, o en la Figura 5.4 para $N = 7$ umbrales diferentes en un mismo sujeto.

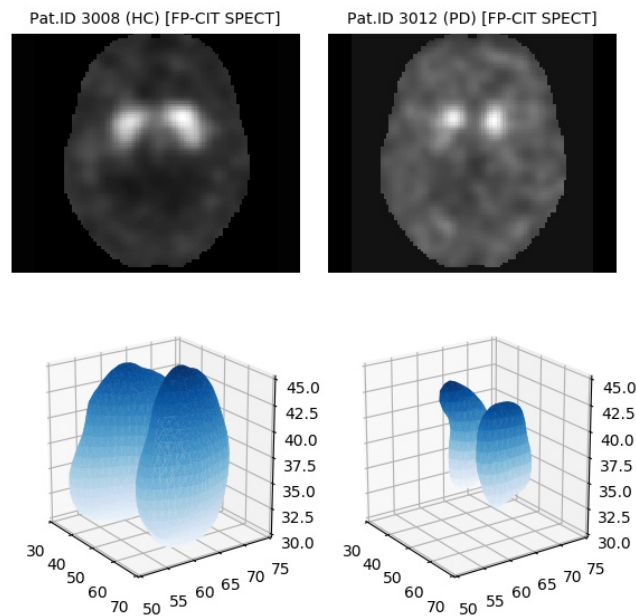


Figura 5.3: Ejemplo de dos isosuperficies calculadas usando el mismo umbral de intensidad para dos imágenes FP-CIT SPECT del repositorio PPMI. Las imágenes han sido previamente normalizadas espacialmente (registro afín) y en intensidad (método con α -estables descrito en 4.4).

²La versión publicada en [488] utilizaba un total de 15 mientras que la presentada en [489] ampliaba este número hasta 33 ordenaciones posibles (aunque algunas de ellas eran topológicamente equivalentes).

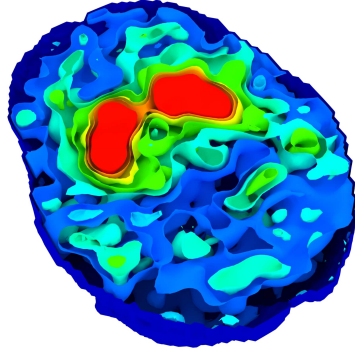


Figura 5.4: Ejemplo de $N = 7$ isosuperficies calculadas para el primer sujeto sano (HC) del repositorio PPMI.

Extracción de características morfológicas La primera etapa para el desarrollo de un modelo de apoyo al diagnóstico de EP basado en las características morfológicas de N isosuperficies generadas para un conjunto de imágenes FP-CIT SPECT, consiste en aplicar un preprocesado a las imágenes de entrada que permita que dichas características sean comparables entre sí (en posición y tamaño). Para ello, y siguiendo las recomendaciones presentadas en *Comparison Between Affine and Non-affine Transformations Applied to I[123]-FP-CIT SPECT Images Used for Parkinson's Disease Diagnosis* [57], se aplica un registro afín de las imágenes de entrada a una plantilla de referencia como la generada siguiendo el procedimiento descrito en [483]. Aunque el enfoque con una normalización en intensidad sigue siendo válido, el uso de características morfológicas permite que no sea necesario este tipo de transformación. A cambio, hace falta determinar qué niveles de intensidad en cada sujeto son comparables con qué niveles de intensidad del resto. Una forma de hacer esto es dividir en N niveles de intensidad posibles el rango de cada imagen y equipararlo con los mismos niveles de otras imágenes como se indica en la Figura 5.5. Siguiendo este esquema, los umbrales de intensidad de cada isosuperficie en cada sujeto pueden calcularse según (5.1) con I_{\min} e I_{\max} los valores de intensidad mínimo y máximo de la imagen respectivamente.

$$I_j = I_{\min} + j \frac{I_{\max} - I_{\min}}{N} \quad \text{con} \quad j = 1, 2, \dots, N \quad (5.1)$$

Una vez se ha fijado el umbral de intensidad de cada isosuperficie, se determina su contorno mediante la función LMCA y se procede con la extracción de las características morfológicas que se describen a continuación³:

³ L_H hace referencia a características calculadas teniendo en cuenta solo vóxeles del

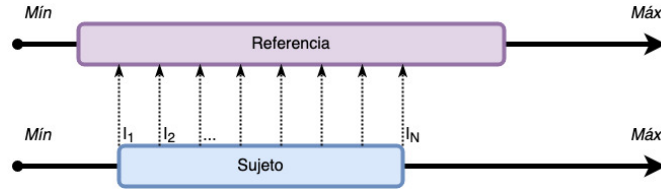


Figura 5.5: Texto.

- **Área y Volumen** $[L_H, R_H, B]$ - Área y volumen de cada isosuperficie.
- **Centro de masas** $[L_H, R_H]$ - Dado un sistema de partículas, su centro de masas representa el punto espacial donde la posición relativa ponderada por la masa de cada partícula suma cero. En este caso, en vez de masa, se usa la intensidad de cada vóxel y su posición. El cálculo del centro de masas en este trabajo se determina utilizando todos los puntos de la isosuperficie. Por tanto, y siendo la intensidad de cada punto del espacio $\mathbf{I}(x, y, z)$, el valor del centro de masas para la isosuperficie I_j que contiene M vóxeles de intensidad vendrá dado por (5.2).

$$\text{CoM}_{I_j} = \frac{\sum_{i=1}^M (x_i, y_i, z_i) * \mathbf{I}(x_i, y_i, z_i)}{\sum_{i=1}^M \mathbf{I}(x_i, y_i, z_i)} \quad \forall \mathbf{I}(x, y, z) \geq I_j \quad (5.2)$$

- **Proyecciones** $[L_H, R_H]$ - Como ya se ha comentado en la Sección 2.6, existen trabajos como los publicados en [41, 393] que han hecho uso de proyecciones sobre los planos axial, sagital y/o coronal como medio para obtener un conjunto de características morfológicas con los que caracterizar la existencia o no de la EP. Siguiendo el modelo publicado en [41], este trabajo ha incluido las proyecciones de cada isosuperficie sobre los planos (x, y, z) y ha generado como características de estas proyecciones medidas las siguientes (en orden alfabético):

- Área de la proyección.
- Centroide (centro de masas).
- Circularidad (redondez) calculada según (5.3).

$$\frac{4\pi \text{Area}}{2 \text{Perímetro}} \quad (5.3)$$

hemisferio izquierdo; R_H si se usan solo vóxeles del hemisferio derecho; y B cuando se usan vóxeles de ambos hemisferios (cerebro completo).

- Diámetro equivalente, D , de un círculo con el mismo área que la región proyectada. Viene dado por (5.4).

$$D = 2\sqrt{\frac{\text{Area}}{\pi}} \quad (5.4)$$

- Excentricidad. Definida como la razón entre la semidistancia focal de la elipse equivalente de la proyección y su semieje mayor.
- Extensión. Razón entre el número de vóxeles de la proyección y el número de vóxeles de una ROI contenida en el plano sobre el que se realiza la proyección.
- Longitud de los ejes mayor y menor de la elipse equivalente que minimiza la distancia entre su contorno y los bordes de la proyección.
- Orientación de la elipse equivalente.
- Perímetro.
- Solidez. Definida como la razón entre el número de vóxeles interiores al contorno calculado con Q_{hull} y el número de vóxeles reales de la proyección.

La Figura 5.6 incluye una representación de las proyecciones axial, coronal y sagital de las isosuperficies representadas en la Figura 5.3.

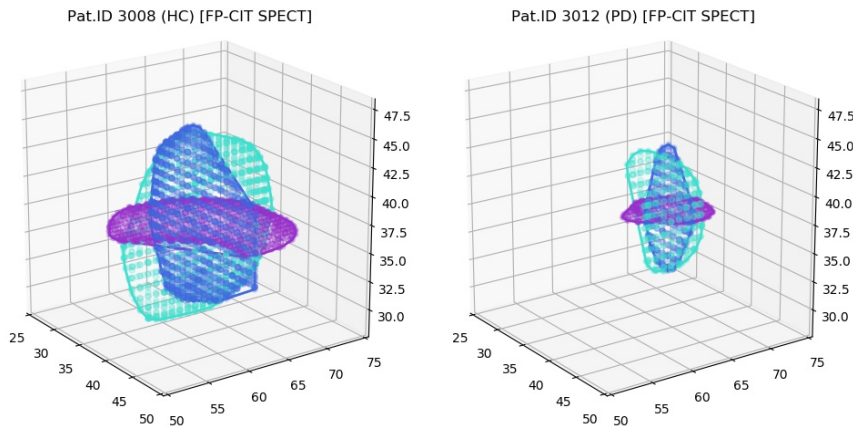


Figura 5.6: Proyecciones sobre los planos X,Y,Z de las isosuperficies en el hemisferio derecho representadas en la Figura 5.3.

- **Características topológicas** $[L_H, R_H]$ - Incluye el *Número de Euler* (χ) del poliedro, [490], y la orientación de la topología. En este caso, χ se ha calculado partir del número de vértices de la isosuperficie (V_p), su número de bordes (E_p) y su número de caras (F_p) según (5.5).

$$\chi = V_p - E_p + F_p \quad (5.5)$$

- **Esfericidad** $[L_H, R_H]$ - Indica el grado de redondez de cada isosuperficie.
- **Coefficiente isoperimétrico** $[L_H, R_H]$ - Para una curva cerrada en un plano, se define como la relación entre el área de la curva (A_r) y el área de una circunferencia que tiene el mismo perímetro que la curva (P_r) según (5.6). En el caso de considerar una región tridimensional, se puede usar el volumen de dicha región (V_r) y su superficie (S_r) como se indica en (5.7).

$$Q_{2D}^{ISO} = \frac{4\pi A_r}{P_r^2} \quad (5.6)$$

$$Q_{3D}^{ISO} = \frac{36\pi V_r^2}{S_r^3} \quad (5.7)$$

- **Indicadores de cantidad** $[L_H, R_H, B]$ - Hace referencia al número de vértices y caras de cada isosuperficie.

Sistema CAD para la evaluación del Parkinson Una vez aplicadas todas las operaciones de extracción de características, se habrán generado un total de 15 características morfológicas si usamos los datos del cerebro completo (B) o 57 valores posibles cuando solo se incluyen las características extraídas de cada hemisferio por separado (L_H o R_H). Generada la lista final, se procede con la clasificación de estos valores siguiendo un modelo de clasificación idéntico al descrito en la Figura 2.3. Para este esquema se utiliza como algoritmo de selección de características el test estadístico Mann-Whitney-Wilcoxon U-Test descrito en [149] como medio para limitar el número de características de entrada en caso de que su valor supere al del número de muestras (sujetos) de entrada⁴. El modelo también hace uso de un esquema de validación cruzada de tipo k -fold con $k = 10$ siguiendo las recomendaciones presentadas en [141, 491]; y como clasificador, algunos de los algoritmos incluidos en la Tabla 5.2 donde también se ha indicado la lista de posibles valores con los que se realiza el ajuste de algunos de sus parámetros [492].

Sobre la implementación de la propuesta con MLP El único método de clasificación que requiere una explicación más detallada es la red neuronal MLP implementada con 3 capas de tamaño $l_{1(\text{size})} = 32$, $l_{2(\text{size})} = 32$ y $l_{3(\text{size})} = 2$. Acompañando el esquema general de esta red incluido en la Figura 5.7, también se proporcionan los siguientes detalles sobre su implementación:

⁴Limitando así el posible sobreajuste (*overfitting*) del modelo de clasificación a los datos de entrada.

| Clasificador | Ajuste de parámetros |
|--|--|
| SVM con <i>kernel</i> lineal [49, 485] | $C \in \{0.1, 1, 10, 100\}$ |
| SVM con <i>kernel</i> RBF (<i>Radial Basis Function</i>) [49, 485] | $C \in \{0.1, 1, 10, 100\}$ $\gamma \in \{0.1, 1, 10\}$ |
| SVM con <i>kernel</i> polinomial [49, 485] | $C \in \{0.1, 1, 10, 100\}$ $d \in \{3\}$ |
| ν SVC [311] | $\nu \in \{0.01, 0.1, 1\}$ |
| Naïve bayesiano [493] | No precisa ajuste. |
| Perceptrón multicapa (<i>Multi-Layer Perceptron</i> , MLP) | Comentado más adelante. |

Tabla 5.2: Clasificadores propuestos para el análisis de características morfológicas en volúmenes FP-CIT SPECT.

1. Aplica una normalización de los bloques de la red (*Batch normalization*) [494].
2. La activación a la salida de las capas l_1 y l_2 es la función ReLU (*Rectifier Linear Unit*); mientras que la función de activación en la capa de salida, l_3 , es log-softmax [495].
3. Durante el entrenamiento, la red aplica una regularización mediante la desactivación aleatoria de diferentes neuronas con probabilidad $p = 0.25$ aplicando la distribución de Bernoulli (técnica de *Dropout*). [496, 497].
4. Para la optimización de los pesos de cada neurona, la red utiliza el optimizador SGD (*Stochastic Gradient Descent*). con una tasa de aprendizaje $\mu_r = 0.1$.
5. La función de pérdidas de la red es la función *Negative Log-likelihood* con expresión general (5.8) siendo $N_b = 16$ el tamaño de cada *batch*; K el número de neuronas en la última capa (número de clases); a_k^L la salida de la neurona k en la capa L (última capa); y el parámetro y un vector de longitud K que contiene un 0 en aquellas posiciones que no coinciden con la clase k que se está evaluando, y 1 en caso afirmativo.

$$\ell(x, y) = -\frac{1}{N_b} \sum_x y^T \ln(a^L) = -\frac{1}{N_b} \sum_x \sum_{k=1}^K y_k \ln(a_k^L) \quad (5.8)$$

6. Durante el entrenamiento, cada vez que la pérdida de validación disminuye, se guarda el estado del modelo. Cuando el resultado de la función de pérdidas en las siguientes 30 iteraciones (*epochs*) no logra

ser menor que el último estado guardado, se detiene el entrenamiento y se restaura el último estado que se haya guardado del modelo (*Early Stopping*) [498].

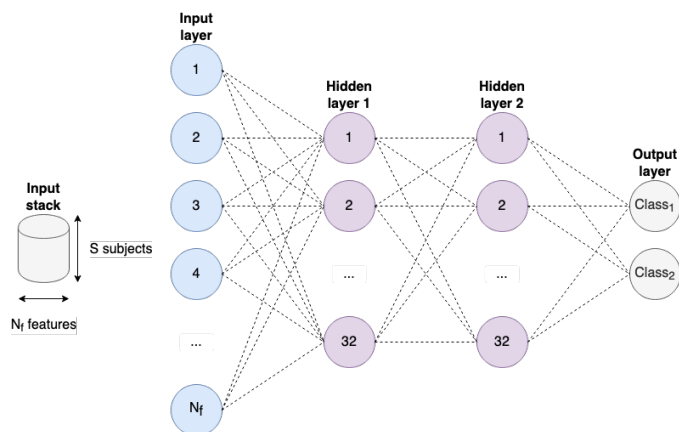


Figura 5.7: Esquema de la red neuronal MLP.

5.1.4. Análisis de resultados

Para probar la eficacia del uso de isosuperficies en el análisis de la EP, se han evaluado un total de 386 imágenes FP-CIT SPECT del repositorio PPMI correspondientes a la primera visita de cada sujeto en el estudio. Una vez preprocesadas las imágenes espacialmente mediante *SPM12*, cada una de ellas se ha dividido en tres conjuntos de datos, correspondientes a su hemisferio izquierdo (L_H), hemisferio derecho (R_H) y cerebro completo (B). Una vez determinados los valores de intensidad máximo y mínimo de cada conjunto, se han calculado los $N \in \{5, 10, 15, 20\}$ umbrales de intensidad según (5.1), se ha determinado la isosuperficie correspondiente a cada uno de estos niveles de intensidad mediante el algoritmo LMCA, y finalmente se han extraído las características morfológicas descritas anteriormente para cada una de estas isosuperficies.

Los resultados de clasificación de todos los experimentos se encuentran incluidos en la Figura 5.8. Los experimentos han sido evaluados utilizando como entrada las características morfológicas procedentes del hemisferio izquierdo (L_H), hemisferio derecho (R_H), cerebro completo (B) y la concatenación de los 3 grupos anteriores ($[L_H, R_H, B]$).

En promedio, los mejores resultados de clasificación se alcanzan cuando $N = 20$. La Tabla 5.3 recoge el detalle de resultados de clasificación (en términos del promedio del parámetro *balanced accuracy*⁵) para este escenario según el clasificador y la región de interés empleados para esta evaluación.

⁵Promedio entre los parámetros especificidad y sensibilidad.

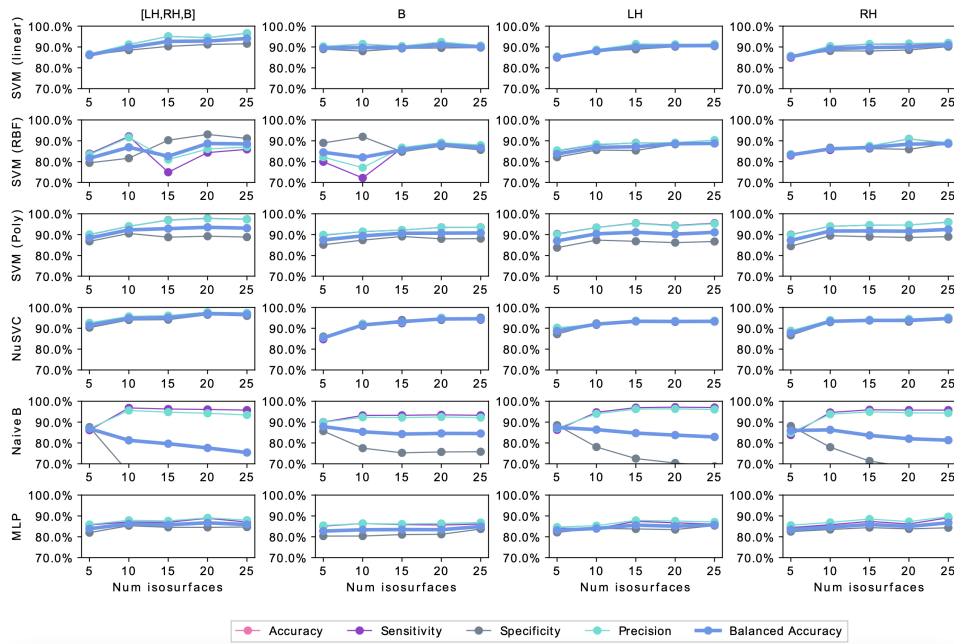


Figura 5.8: Resultados de clasificación para diferentes números de isosuperficies en función de la región evaluada y métodos de clasificación evaluados.

Todos experimentos han sido repetidos 10 veces con diferentes inicializaciones (aleatorias) para determinar la variabilidad de los resultados asociada a las selecciones de los conjuntos de entrenamiento y prueba. La Fig. 5.9 incluye una representación de esta variabilidad para los modelos de clasificación con $N = 20$ isosuperficies.

| Clasificador | L_H | R_H | B | $[L_H, R_H, B]$ |
|------------------|----------------|----------------|----------------|-----------------|
| SVM (Lineal) | 90.59 % | 89.98 % | 90.91 % | 92.83 % |
| SVM (RBF) | 88.56 % | 88.42 % | 88.19 % | 88.70 % |
| SVM (Polinomial) | 90.26 % | 91.64 % | 90.72 % | 93.54 % |
| ν SVC | 93.26 % | 93.75 % | 94.46 % | 97.04 % |
| Naïve bayesiano | 83.76 % | 82.01 % | 84.58 % | 77.60 % |
| MLP | 85.07 % | 84.96 % | 83.42 % | 86.71 % |

Tabla 5.3: Resultados de clasificación utilizando un total de $N = 20$ isosuperficies para el hemisferio izquierdo (L_H), hemisferio derecho (R_H), cerebro completo (B), y la concatenación de todos los anteriores ($[L_H, R_H, B]$).

Para el escenario con $N = 20$ isosuperficies, las características que consiguen una mejor separación entre clases en términos del test estadístico Mann-Whitney-Wilcoxon son los que se indican en la Tabla 5.4. En esta

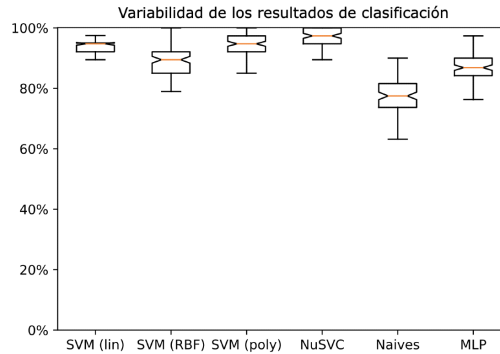


Figura 5.9: Variabilidad de resultados de clasificación para los modelos con $N = 20$ isosuperficies que hacen uso del conjunto de características $[L_H, R_H, B]$.

tabla, las isosuperficies con índices numéricos más altos hacen referencia a umbrales de intensidad mayores. Aunque no se les ha hecho referencia al resto de $N_f = N * (15 + 57 + 57) = 2580$ características en el escenario con más número de variables de entrada, cabe mencionar que 2442 de ellas presentaron un $p_{\text{valor}} \leq 0.05$.

| Característica | Nº Isosuperficie | Región |
|-----------------------------|------------------|--------|
| Perímetro (plano Z) | 10 | L_H |
| Excentricidad (plano Y) | 4 | R_H |
| Esfericidad | 12 | L_H |
| Centro de masas | 12 | B |
| Nº de vértices | 13 | R_H |
| Centroide (plano Y) | 3 | R_H |
| Nº de caras | 13 | L_H |
| Volumen | 13 | L_H |
| Coefficiente isoperimétrico | 5 | B |
| Esfericidad | 13 | R_H |

Tabla 5.4: Características morfológicas con menor p_{valor} (promediado entre todos los conjuntos de entrenamiento) al comparar sujetos sanos y pacientes con EP en un escenario con $N = 20$ isosuperficies.

Aparte de los resultados de clasificación, también se han incluido en este trabajo las curvas ROC (*Receiver Operating Curves*, [142]) asociadas al experimento con mejor tasa de clasificación (ν SVC) y a la red neuronal (MLP). Como se puede observar en la Figura 5.10, el área bajo la curva (*Area Under the Curve*, AUC) de los experimentos que utilizan los datos procedentes del conjunto $[L_H, R_H, B]$ alcanzan unos valores de AUC igual

a 0.9844 (máximo de todos los experimentos) para ν SVC e igual a 0.9128 para la red MLP.

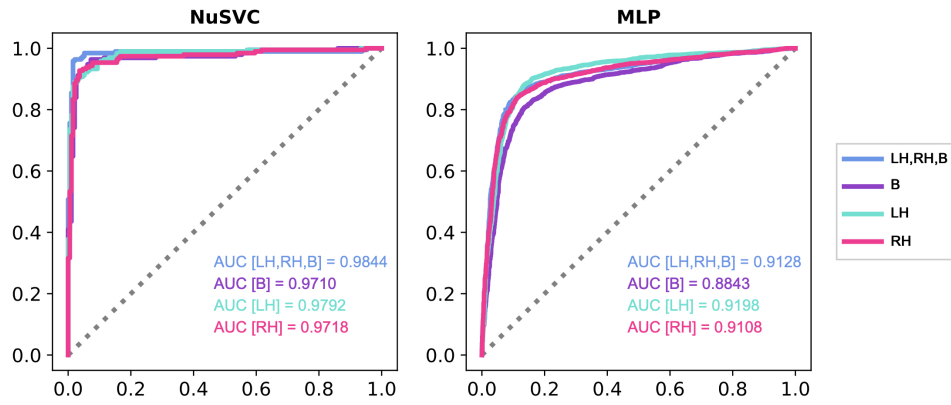


Figura 5.10: Curvas ROC para los experimentos que utilizan ν SVC y MLP.

Capítulo 6

Análisis multivariante de biomarcadores heterogéneos para el estudio de la base de datos DIAN

En este bloque se desarrolla un análisis de la base de datos DIAN para el estudio de la variante hereditaria de la EA. El modelo trata de determinar las diferencias entre sujetos de control (sanos) y pacientes con alguna mutación en los genes PSEN1, PSEN2 o APP; con especial atención a las etapas presintomáticas de su enfermedad.

6.1. Evaluación de la base de datos DIAN

6.1.1. Justificación del análisis

Si tomamos como referencia los modelos multivariantes, y en concreto los modelos que hacen uso de técnicas de ML, se pueden encontrar cientos de propuestas de tipo CAD que han tratado de modelar y predecir multitud de patrones procedentes de trastornos neurológicos como la demencia [2, 3, 124, 324, 347]. Estos modelos han contribuido de manera significativa al entendimiento que tenemos hoy en día sobre este tipo de enfermedades y constituyen una herramienta muy eficiente a la hora de validar hipótesis sobre sus desencadenantes, su pronóstico e incluso a la hora de medir la efectividad de nuevas líneas de tratamiento.

A nivel clínico, uno de los trabajos más importantes en el estudio de la EA es el artículo publicado por **Jack et al.** en 2010 [451] que trata sobre la relación que tienen diferentes tipos de marcadores en pacientes con EA incluyendo la acumulación del péptido β -amiloide ($A\beta$) [499], el deterioro cognitivo y la atrofia cerebral a lo largo del tiempo. En concreto, y tal y

como se ha intentado replicar en la Figura 6.1 sobre la actualización del trabajo presentada en 2013 [500], la evidencia del trabajo sugiere que la deposición de placas de $A\beta$ empieza a producirse entre 20 – 30 años antes de la aparición de los primeros síntomas cognitivos [82, 108].

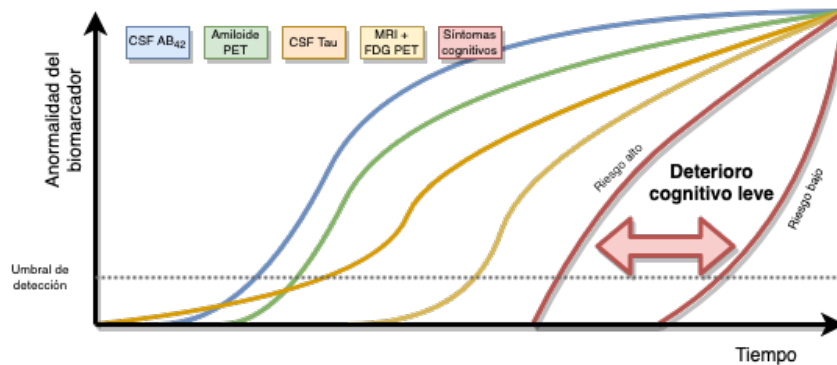


Figura 6.1: Modelo teórico propuesto por **Jack et al.**.

Aunque desde hace tiempo la enfermedad ha sido catalogada como una entidad heterogénea [405, 501, 502], muchos de los estudios sobre la EA han tratado de determinar factores claves en el desarrollo de la demencia comparando sujetos sanos frente a pacientes que desarrollan la enfermedad independientemente de su variante histológica. Afortunadamente la tendencia en los últimos años ha ido cambiando sobre todo a raíz de las conclusiones de trabajos como el publicado por **Ryan et al.** en [503], en el que los autores inciden por un lado en la necesidad de hacer un seguimiento de los sujetos que desarrollan la EA a una edad temprana incluyendo un análisis de las 3 posibles mutaciones genéticas y analizando cada tipo de mutación por separado. Aunque este cambio ha sido positivo a nivel de explicar algunos de los mecanismos que activan el desarrollo de la enfermedad, todos los análisis llevados a cabo hasta la fecha han tratado de caracterizar la evolución de la enfermedad mediante modelos de regresión como LME (*Linear-Mixed-Effects*) o mediante análisis univariantes utilizando tests estadísticos como el t-Test o similares. En este contexto, resultaría interesante por un lado validar el modelo propuesto por **Jack et al.** haciendo uso de técnicas más avanzadas como ML sobre repositorios como DIAN, y por otro determinar si existen diferencias relevantes a nivel de imagen médica o de cualquiera de los biomarcadores recogidos en la base de datos que pueda ayudar a predecir qué tipo de mutación presenta cada sujeto y cómo se espera que puede evolucionar este. Si, como algunos trabajos sugieren, cada mutación desencadena la enfermedad de forma diferente, entonces crear un modelo pronóstico a partir de los marcadores que sean más relevantes puede ayudar a definir nuevas líneas de tratamiento concentrando el esfuerzo de la investigación clínica en el desarrollo de dichas terapias [503–507].

Con base en el marco presentado, el trabajo con título *Autosomal Dominantly Inherited Alzheimer Disease: Analysis of genetic subgroups by Machine Learning* publicado en [58], propone un análisis de la base de datos DIAN utilizando técnicas de ML, que permite identificar patrones con los que diferenciar a los portadores de alguna mutación en los genes PSEN1, PSEN2 y/o APP (*Mutation-Carriers*, MC); frente a sujetos sanos, asintomáticos y sin mutación (*Non-Carriers*, NC), y que además evalúa la utilidad de ciertos tipos de marcadores en diferentes etapas presintomáticas de la enfermedad.

6.1.2. Propuesta para el análisis de fuentes heterogéneas y definición de experimentos

El análisis que se pretende hacer con los datos de DIAN está sujeto a una restricción fundamental que tiene que ver con la heterogeneidad de los datos de entrada y con la forma de agrupar a los sujetos de entrada. Con el fin de dar respuesta a algunas cuestiones abiertas hasta la fecha sobre DIAD, para este trabajo se han definido varios experimentos de clasificación desarrollados como sigue a continuación:

- **Experimento 1.** Trata de comparar sujetos NC frente a participantes MC para establecer un modelo base que sirva como referencia para las comparaciones que se propongan más tarde. Aunque ya en algún trabajo previo se han señalado la importancia de analizar a los MC por separado según el gen responsable de su mutación [503], estos tipos de comparaciones son útiles de cara a la práctica clínica porque ayudan a identificar qué pruebas clínicas son realmente relevantes (al menos desde el punto de vista estadístico) y porque permiten generar un modelo simplificado de la enfermedad.
- **Experimento 2.** Pretende determinar si subagrupar a los MC en función de su gen mutado permite obtener mejores resultados de clasificación. Hasta la fecha, todos los trabajos que habían propuesto comparar algún subgrupo de MC lo habían hecho comparando NC frente a MC_{PSEN1} (MC con una mutación en el gen PSEN1) pero siempre desde el punto de vista de la estadística clásica con tests estadísticos como el t-Test o modelos de regresión como LME [405, 406, 503, 508]. En este caso lo que se propone es comparar NC frente a MC_{PSEN1} pero utilizando un modelo de ML y, posteriormente, ampliar dicho modelo a una comparación multiclase comparando los grupos NC, MC_{PSEN1} , MC_{PSEN2} y MC_{APP} .

En una fase preliminar del análisis de los datos de DIAN se observó que el subconjunto MC_{PSEN1} estaba conformado por sujetos con una variabilidad muy alta. Para este subconjunto se planteó la hipótesis

adicional de que existieran dos subgrupos (*clusters*) de sujetos. Un primer subgrupo compuesto por pacientes con características muy similares a los NC (MC_{PSEN1}^{NC+}) y un segundo grupo totalmente diferente (MC_{PSEN1}^{NC-}). De acuerdo con esto, también se propuso la necesidad de validar esta hipótesis e identificar qué sujetos dentro del subconjunto MC_{PSEN1} podría estar comportándose como sujetos totalmente asintomáticos y, por tanto, limitando la capacidad del sistema de aprendizaje automático a la hora de distinguir *NC* frente a *MC*.

- **Experimento 3.** Una vez ajustado el modelo para DIAD, sería interesante realizar una comparación entre sujetos NC y MC en diferentes etapas presintomáticas de la enfermedad. Para ello, y tomando como referencia el valor de la variable EYO (*Estimated Years to Onset*), se han propuesto varios experimentos de clasificación comparando MC con un tiempo estimado hasta la aparición de sus primeros síntomas mayor o igual que ciertos umbrales de referencia (más de 5 años, más de 10 años y más de 15 años) y NC.

6.1.3. Esquema de clasificación

El esquema de clasificación que se ha seguido para analizar las muestras de la base de datos DIAN tiene la misma estructura que el diagrama descrito en la Figura 2.3. En este esquema se incluyen los siguientes bloques:

- Una etapa inicial para descartar muestras incompletas (*missing data*) seguida de un balanceo de clases para los datos de entrenamiento que evita el posible sesgo (*bias*) asociado a que alguna de las clases tenga más muestras que el resto [509].
- Un bloque para la estandarización del conjunto de datos de entrenamiento.
- Una selección de características basada en el uso del test de ANOVA, [151, 152], para descartar aquellas variables que aportaban poca información. Aunque como ya se comentó en el Capítulo 4, para aquellos escenarios en los que los histogramas de los datos de entrada puedan tener una respuesta impulsiva y/o con colas pesadas es más conveniente recurrir a tests no paramétricos como el test de Mann-Whitney-Wilcoxon, [149]; en este trabajo se ha recurrido al test de ANOVA con base en dos criterios:
 1. Que se puede asumir que los datos siguen una distribución normal según el test de Kolmogórov-Smirnov (con intervalo de confianza del 95 %) para la mayor parte de las características de entrada al sistema de clasificación [510].

2. Que un test paramétrico como ANOVA tiene más poder estadístico que un test no-paramétrico en cuanto a que si realmente existe alguna diferencia entre clases, es más fácil que un test paramétrico detecte esta diferencia.
- Ya que la selección de características puede no ser suficiente en algunos experimentos, en cuanto a que el número de características de entrada sea menor que el número de sujetos a la entrada del clasificador, a la salida del bloque de selección se ha incluido una etapa de reducción de características mediante PCA [144] que ayuda a prevenir el posible sobreajuste de los modelos de clasificación a los datos de entrenamiento (*overfitting*).
 - Respecto al bloque de clasificación, para este trabajo se ha hecho uso de clasificadores SVM con *kernel* lineal [49, 485].
 - Para validar los resultados de clasificación se ha empleado un esquema de validación cruzada de tipo *k-fold* con $k = 10$ siguiendo las recomendaciones publicadas en [141]. Para asegurar la validez de los resultados, la validación cruzada ha sido repetida hasta en 10 ocasiones probando en cada una de ellas una inicialización diferente (aleatoria). Además de este modelo, se ha realizado un análisis de los límites superiores del error empírico basado en la teoría de la convergencia uniforme descrita en [305], y la particularización propuesta por Górriz et al. en [511] para clasificadores lineales, que ayuda a justificar los resultados de clasificación que se obtienen para cada experimento.

6.1.4. Análisis de resultados

Siguiendo el esquema de clasificación propuesto, la primera parte del análisis de sujetos con DIAD ha consistido en el ajuste del número de componentes de PCA (N_{Comp}). Para ello, y en función del tipo de dato de entrada que se quiere analizar (NIB,FDG,MRI,PIB), se han definido cuatro modelos de clasificación con la estructura descrita en la Sección 6.1.3.

En la práctica, se ha limitado el número máximo de variables seleccionadas por ANOVA (ordenadas de menor a mayor p_{valor}) a 50 características por cada tipo de dato. Luego, en la etapa de reducción con PCA, este número se reduce hasta 20 componentes. Esto permite evaluar mejor qué tipo de características son más útiles de cara a la clasificación de NC frente a MC a la vez que se simplifica la interpretabilidad de los resultados. Si se diera el caso de que algún conjunto de datos de entrada no tenga 50 o más muestras, o que incluso este número sea menor que el número máximo de componentes con PCA, entonces se seleccionarían tantas características como variables de entrada tenga ese conjunto.

Experimento 1 Respecto primer experimento, inicialmente se propuso hacer una comparación directa entre sujetos NC y MC independientemente de su sintomatología. Sin embargo, y debido a la escasa diferencia entre ambos grupos (con tasas de clasificación máximas, en términos del *balanced accuracy*, equivalentes al 60.39 % para FDG y al 60.08 % para PIB), se fijó que esta comparación debería realizarse utilizando solo participantes asintomáticos de forma que se redujese la heterogeneidad de la muestra. Pese a esta idea, como se puede observar en la Figura 6.2, los resultados de clasificación continuaron siendo bajos.

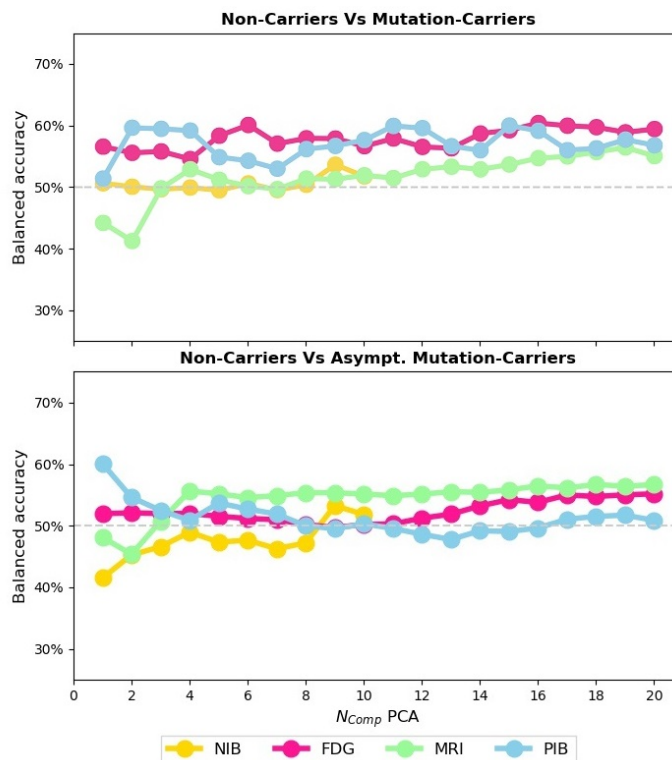


Figura 6.2: Resultados de clasificación (en términos del parámetro *balanced accuracy*) obtenidos para las comparaciones entre NC y MC considerando todos los MC (arriba) o solo los MC asintomáticos (abajo).

Si en vez de comparar directamente los NC frente a los MC, se realiza una clasificación multiclase de forma que se comparen todos los tipos de sujetos a la vez (NC, MC_{PSEN1} , MC_{PSEN2} y MC_{APP}), la cantidad de individuos por cada clase se ve muy reducida debido a la falta de sujetos con mutaciones en los genes PSEN2 (20 sujetos de los cuales 19 son asintomáticos) y APP (49 sujetos de los cuales 29 son asintomáticos), y la condición de que los experimentos sean balanceados. Este hecho reduce considerablemente la capacidad de generalización del modelo de clasificación como puede comprobarse en la

Figura 6.3 donde se muestran los resultados de esta clasificación en función del número de componentes de PCA. Tomando como referencia estos resultados, se ha decidido centrar el análisis del conjunto MC en la evaluación de los individuos con alguna mutación en el gen PSEN1 (MC_{PSEN1}) por ser el subconjunto con mayor cantidad de muestras.

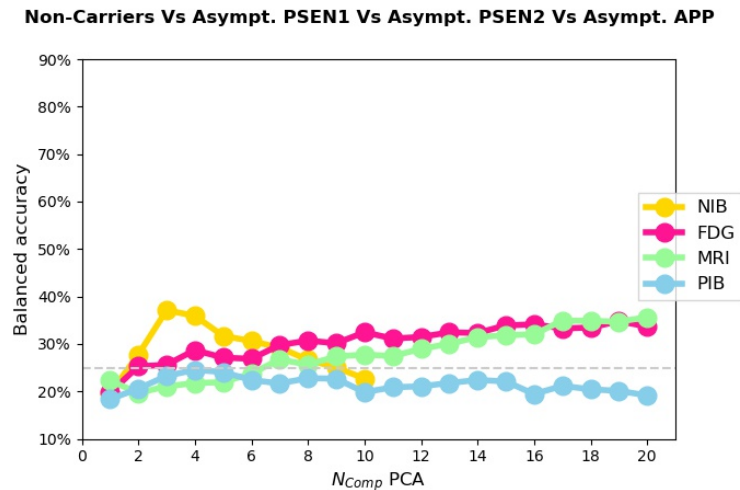


Figura 6.3: Resultados de clasificación (en términos del parámetro *accuracy*) obtenidos para las comparaciones entre NC, PSEN1, PSEN2 y APP.

Experimento 2 Como se puede comprobar en la Figura 6.4, la comparación entre NC y MC_{PSEN1} asintomáticos consigue una mejora en la tasa de clasificación equivalente al 11.31% para $N_{Comp} = 15$ y usando como características de entrada los marcadores de imagen PIB.

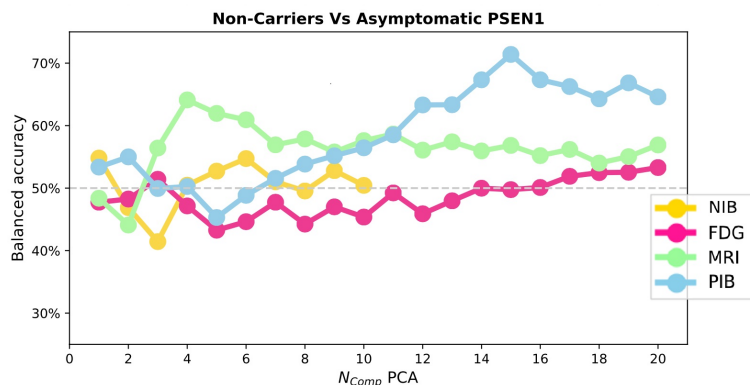


Figura 6.4: Resultados de clasificación (en términos del parámetro *balanced accuracy*) para las comparaciones entre NC y MC_{PSEN1} asintomáticos.

El resultado alcanzado con los marcadores de imagen PIB coincide con el conjunto de características de entrada que mejor significancia estadística consiguen según el test de ANOVA independientemente de si la comparación aglutina a todos los MC en un solo conjunto o si evalúa a los MC_{PSEN1} por separado como puede observarse en la Figura 6.5. Nótese que en esta representación se ha incluido el valor del logaritmo del p_{valor} asociado a cada región del cerebro, en vez del propio p_{valor} , debido a que los resultados más bajos de este parámetro se corresponden con una mayor diferencia entre clases.

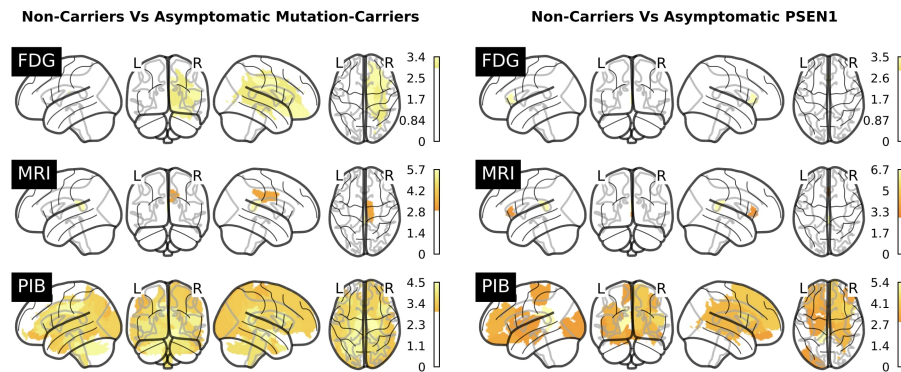


Figura 6.5: Representación del logaritmo del p_{valor} asociado a cada región para las comparaciones de NC frente a MC asintomáticos (izquierda), y de NC frente a MC_{PSEN1} asintomáticos (derecha).

Como por lo general, las tasas de clasificación siguen siendo bajas incluso cuando se comparan NC frente a MC_{PSEN1} asintomáticos, para este experimento se propuso incluir un análisis de los límites superiores del error empírico basados en la teoría de la convergencia uniforme para distintos valores del parámetro N_{Comp} de PCA. Con esto lo que se pretende es justificar si el clasificador está realmente intentando explicar algún patrón subyacente en los datos de entrada o si los resultados de clasificación se deben a la naturaleza aleatoria de la muestra. Los resultados de este análisis, para los dos tipos de datos con mejores tasas de clasificación (en este caso, PIB y MRI) y usando un umbral de significancia del 95% para el rechazo de la hipótesis nula, han sido incluidos en la Tabla 6.1.

| Tipo | N_{Comp} | CV | Resustitución | Límite | H_0 |
|------|------------|---------|---------------|---------|----------|
| MRI | 3 | 56.41 % | 61.21 % | 16.34 % | Rechazar |
| | 10 | 57.60 % | 65.82 % | 28.21 % | Rechazar |
| | 15 | 56.83 % | 68.01 % | 29.31 % | Rechazar |
| | 20 | 56.90 % | 69.83 % | 30.35 % | Rechazar |
| PIB | 3 | 50.00 % | 62.70 % | 16.34 % | Rechazar |
| | 10 | 57.46 % | 85.38 % | 28.21 % | Aceptar |
| | 15 | 71.40 % | 87.50 % | 29.31 % | Aceptar |
| | 20 | 64.60 % | 89.48 % | 30.35 % | Aceptar |

Tabla 6.1: Resumen de resultados de clasificación (en términos del parámetro *balanced accuracy*) entre NC y MC_{PSEN1} asintomáticos incluyendo el resultado de la resustitución y el límite superior del error empírico según **Górriz et al.**

Clustering Si nos centramos en el análisis del conjunto MC_{PSEN1} a nivel de sintomatología, se puede observar que mientras un grupo de sujetos empiezan a mostrar síntomas asociados a cuadros de depresión o a dificultades a la hora de resolver ciertos tipos de pruebas en los tests neuropsicológicos; a una edad muy temprana, otros retrasan este tipo de problemas en torno a 5 – 10 años después. Este hecho, unido al análisis de algunos de los histogramas de las variables con mejor p_{valor} como los representados en la Figura 6.6 para los marcadores FDG, llevó a la idea de el conjunto de sujetos MC_{PSEN1} podría en realidad estar formado por dos subconjuntos de individuos.

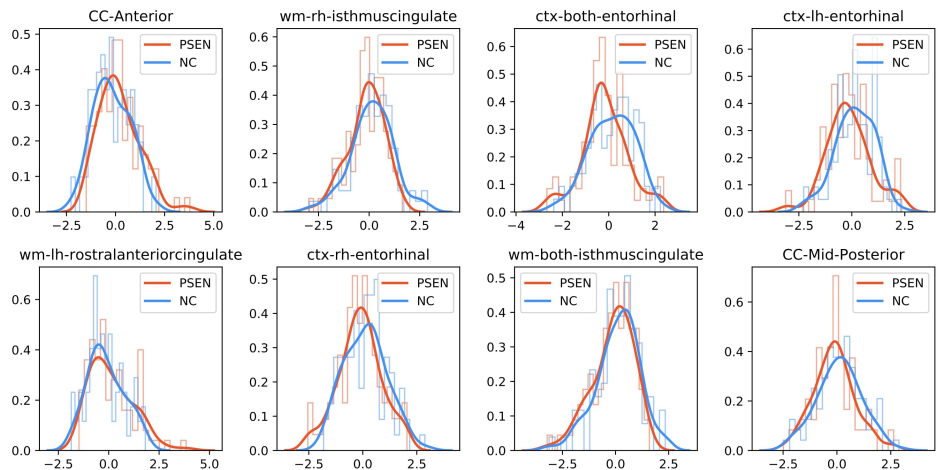


Figura 6.6: Histogramas de los 8 marcadores FDG con mejor p_{valor} según el test de ANOVA cuando se comparan los subconjuntos NC y MC_{PSEN1} asintomáticos.

Para demostrar esta afirmación, se ha propuesto el uso del Criterio de Información de Akaike, [474], aplicado al algoritmo de *clustering* de *k*-medias (*k-means*, [512, 513]) como método para determinar el número *clusters* dentro del grupo de los MC_{PSEN1} e identificar los sujetos que pertenecerían a cada *clusters*. El resultado, representado en la Figura 6.7, muestra dos subconjuntos de sujetos: uno (verde, MC_{PSEN1}^{NC+}) con características muy similares al conjunto de los NC (amarillo) y otro (violeta, MC_{PSEN1}^{NC-}) desplazado espacialmente respecto al centro de la nube de los NC. Las demográficas asociadas a estos subconjuntos se incluyen en la Tabla 6.2.

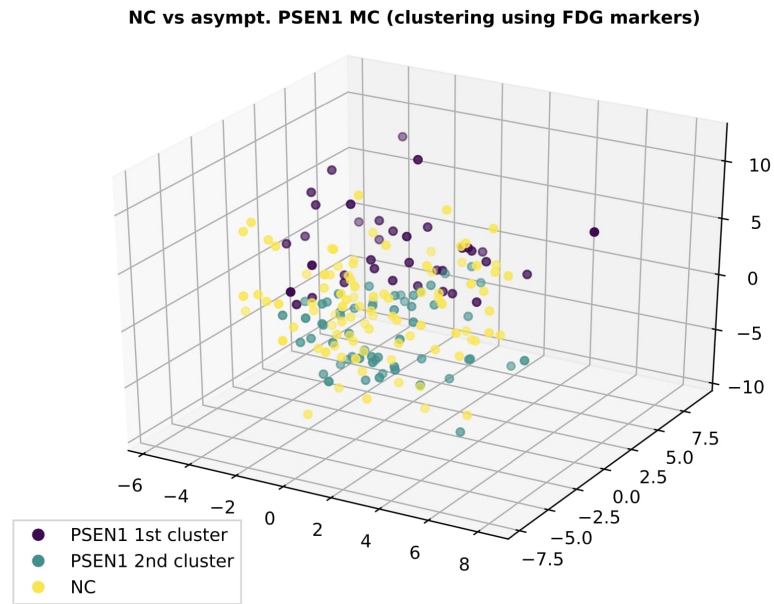


Figura 6.7: *Clustering* generado a partir de los marcadores FDG incluyendo NC y MC_{PSEN1} asintomáticos. Representación gráfica incluyendo las 3 primeras componentes de PCA.

En la Figura 6.8 se han recogido los resultados de la clasificación entre los NC y los subconjuntos de MC_{PSEN1} poniendo de manifiesto una vez más la escasa diferencia que existe (a nivel de marcadores de imagen) entre los sujetos del subconjunto MC_{PSEN1}^{NC+} y los NC. Como se puede observar, el subgrupo de los MC_{PSEN1}^{NC-} alcanza tasas de clasificación cercanas al 80 % mientras que el otro grupo apenas supera la barrera del 65 %.

Experimento 3 Una de las variables más interesantes del repositorio DIAN es la que trata de estimar cuántos años faltan para que aparezcan los primeros síntomas en un sujeto (*Estimated Years to Onset*, EYO) en relación con la edad a la que aparecieron los primeros síntomas en sus progenitores (uno o ambos) [105, 107, 514]. Con el fin de medir cambios en etapas presintomáticas de la enfermedad, se ha propuesto un experimento

| Conjunto | Cluster | Sexo | N | Edad | MMSE |
|-----------------------|------------------------------------|---------|----|-------------|------------|
| Todos | MC _{PSEN1} ^{NC-} | Hombres | 18 | 40.2 ± 11.4 | 26.7 ± 5.0 |
| | | Mujeres | 28 | 38.6 ± 10.9 | 26.7 ± 6.0 |
| | | Ambos | 46 | 39.2 ± 11.1 | 26.7 ± 5.7 |
| | MC _{PSEN1} ^{NC+} | Hombres | 23 | 42.6 ± 10.6 | 26.6 ± 5.0 |
| | | Mujeres | 51 | 39.4 ± 9.8 | 26.8 ± 4.7 |
| | | Ambos | 74 | 40.4 ± 10.1 | 26.8 ± 4.8 |
| Solo asintomáticos | MC _{PSEN1} ^{NC-} | Hombres | 16 | 41.7 ± 10.7 | 26.9 ± 5.3 |
| | | Mujeres | 21 | 37.4 ± 9.7 | 26.7 ± 5.6 |
| | | Ambos | 37 | 39.2 ± 10.4 | 26.8 ± 5.5 |
| | MC _{PSEN1} ^{NC+} | Hombres | 20 | 44.2 ± 10.4 | 26.6 ± 5.0 |
| | | Mujeres | 37 | 39.3 ± 9.1 | 26.8 ± 4.7 |
| | | Ambos | 57 | 41.0 ± 9.9 | 27.8 ± 4.7 |

Tabla 6.2: Demográficas de los subconjuntos MC_{PSEN1}^{NC-} y MC_{PSEN1}^{NC+} cuando se utilizan todos los sujetos MC_{PSEN1} o solo los asintomáticos.

en el que se compara el grupo de los MC_{PSEN1} asintomáticos¹ frente a los NC en diferentes instantes de tiempo anteriores a la manifestación (estimada) de sus síntomas. Los resultados de estas comparaciones se recogen en la Figura 6.9. Como puede observarse, los mejores resultados de clasificación (en términos del parámetro *balanced accuracy*) se obtienen para MRI cuando EYO ≥ 15 años; PIB cuando EYO ≥ 10 años; y PIB cuando EYO ≥ 5 años.

Justificación del número de componentes de PCA Un detalle que no se ha comentado para cada experimento por separado es la selección del número de componentes de PCA. Las condiciones que se han marcado para este trabajo es que la elección de N_{Comp} debe ser suficiente como para explicar al menos el 90% de la variabilidad de los datos de entrada. A la vista de los resultados presentados en la Figura 6.10 y a los resultados de clasificación obtenidos en todos los experimentos, se justifica que el número de componentes sea $N_{\text{Comp}} = 15$.

Análisis de curvas ROC Con respecto al análisis de las curvas ROC asociadas a cada experimento, y con particular atención a los marcadores de imagen, lo primero que cabe destacar es que los bajos resultados de clasificación asociados a las comparaciones entre NC y MC también se ven reflejados en la forma de sus curvas (Figura 6.11).

¹Puesto que es el experimento que mejores resultados consigue independientemente del subtipo de MC_{PSEN1}.

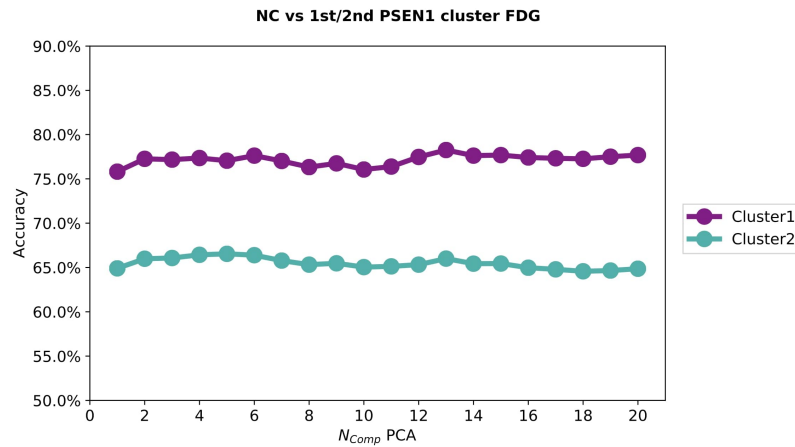


Figura 6.8: Tasas de clasificación, en términos del parámetro *balanced accuracy*, a la hora de comparar los subconjuntos MC_{PSEN1}^{NC-} (Cluster1) y MC_{PSEN1}^{NC+} (Cluster2) frente a los sujetos NC.

Nótese que, a pesar de que reducir la heterogeneidad de la muestra de entrada mediante comparaciones entre NC y MC_{PSEN1} consigue una mejora en la clasificación que se traduce también en una mejora notable en la forma de sus curvas ROC (Figura 6.12).

Por último, y con el fin de no sobrecargar esta sección con todas las representaciones de las curvas ROC para cada experimento de clasificación, se ha incluido un resumen del área bajo la curva (*Area Under the Curve*, AUC) en la Figura 6.13.

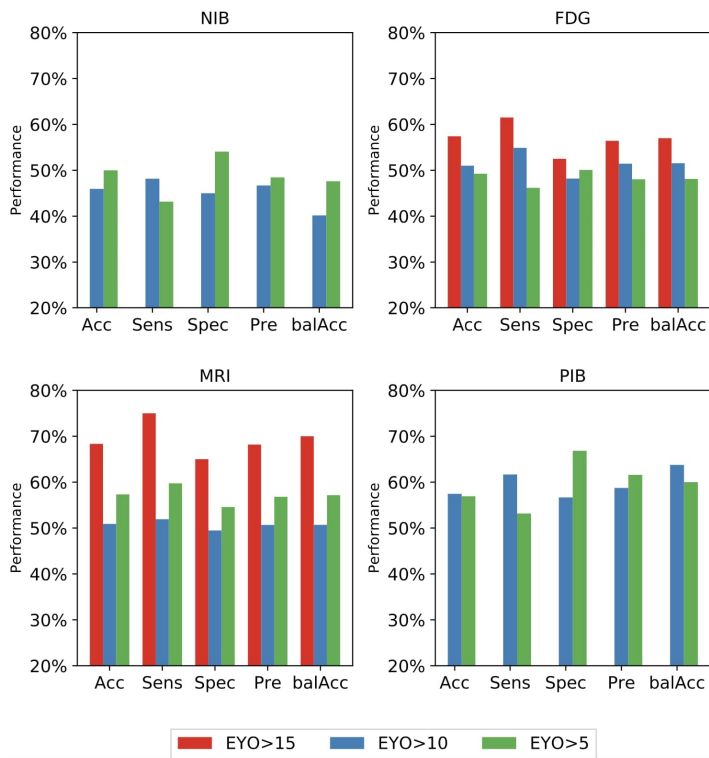


Figura 6.9: Resultados de clasificación cuando se comparan los grupos NC y MC_{PSEN1} asintomáticos con diferentes umbrales de la variable EYO.

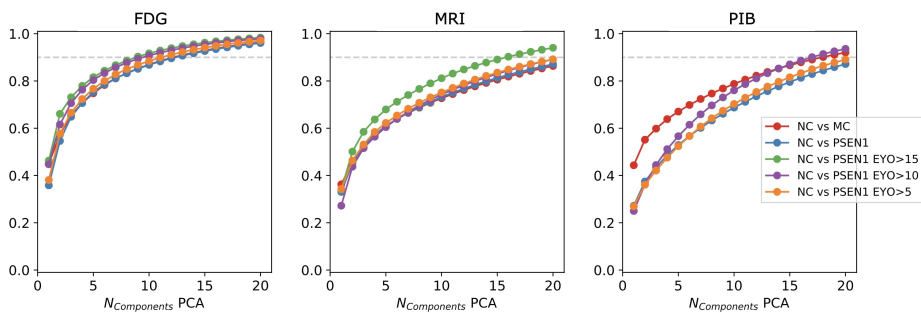


Figura 6.10: Porcentaje de variabilidad explicada por PCA en función del número de componentes N_{Comp} .

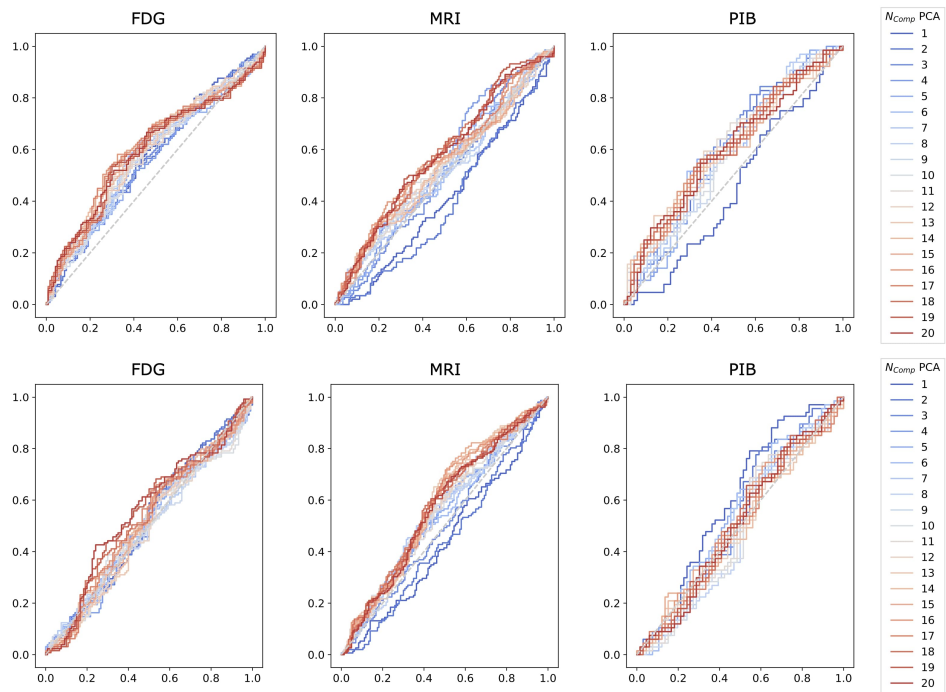


Figura 6.11: Curvas ROC asociadas a la comparación NC frente a MC incluyendo todos los casos (fila superior) y NC frente a MC asintomáticos (fila inferior).

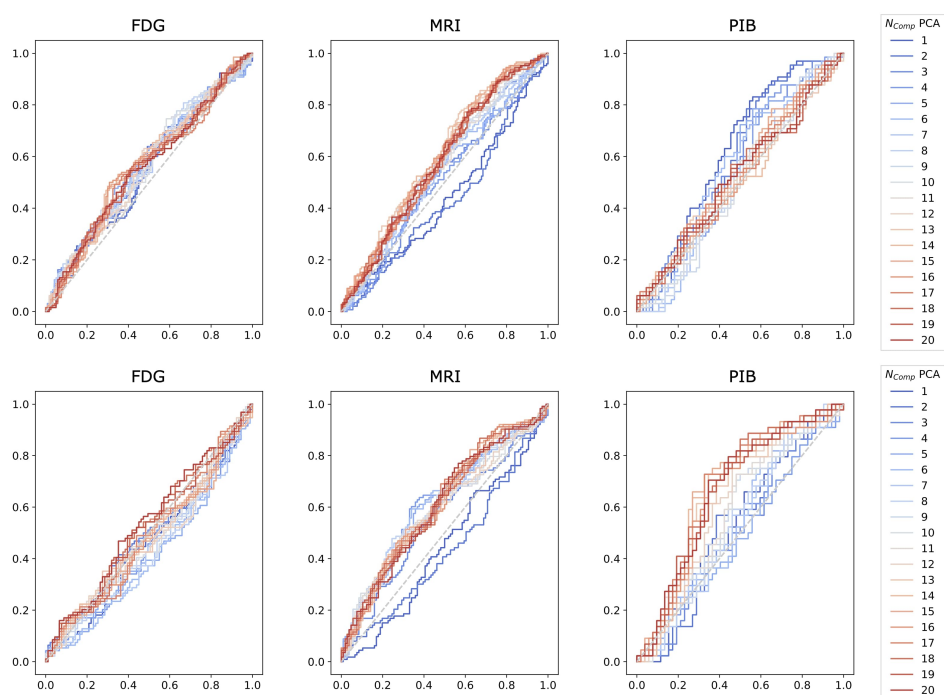


Figura 6.12: Curvas ROC asociadas a la comparación NC frente a MC_{PSEN1} incluyendo todos los casos (fila superior) y NC frente a MC_{PSEN1} asintomáticos (fila inferior).

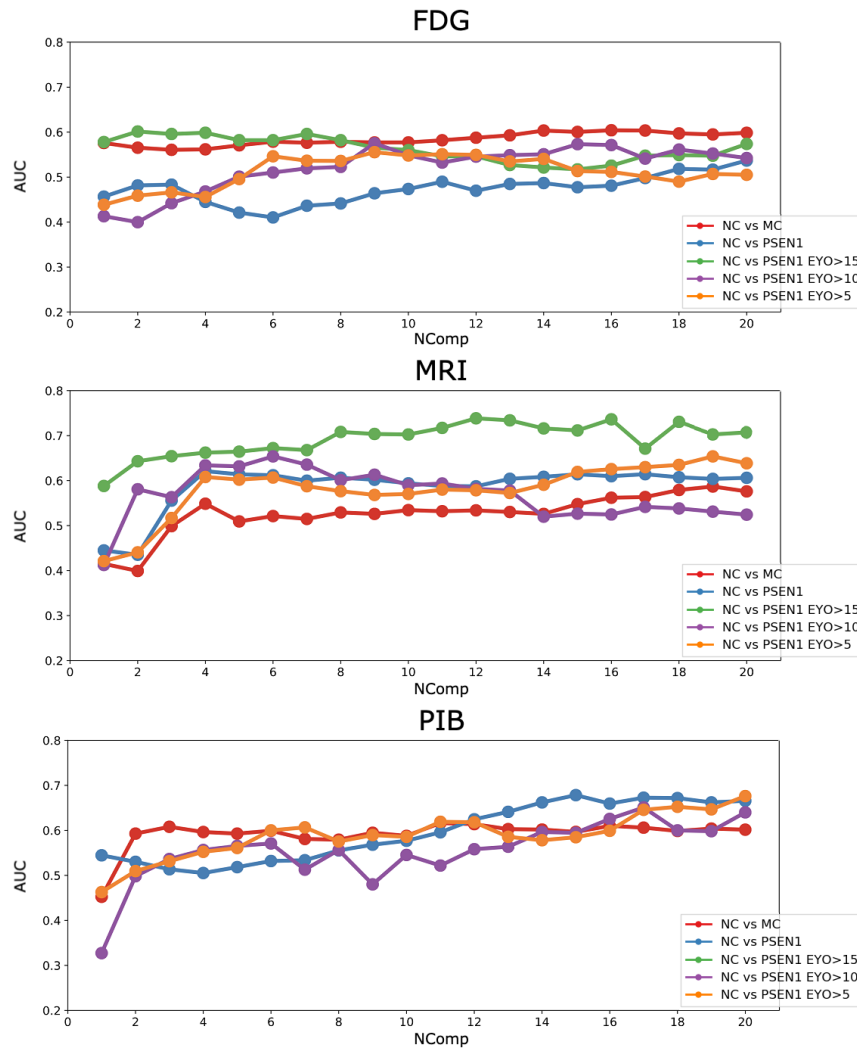


Figura 6.13: Resultados del área bajo la curva calculada para cada experimento de clasificación.

Capítulo 7

Análisis multivariante mediante aprendizaje de conjuntos

En este capítulo se presenta una nueva metodología para el aprendizaje de conjuntos (*ensemble*) que trata de mejorar el WMV tradicional (Sección 7.1.1). La propuesta, desarrollada en la Sección 7.1.2, está basada en el uso de un doble bucle de validación cruzada a partir del cual se obtienen los pesos iniciales de cada tipo de marcador de entrada y un bloque de inventariado que ayuda a descartar las fuentes que menos contribuyen a la clasificación final (Sección 7.1.3). Los resultados incluidos en la Sección 7.1.4 demuestran que la metodología supera con creces la alternativa de WMV incluso pese a incluir fuentes con tasas de clasificación/regresión bajas.

7.1. Metodología para la mejora del voto por mayoría ponderado

7.1.1. Justificación

Como ya se ha comentado en la Sección 2.8, incluso en la actualidad existen pocos trabajos que hayan propuesto nuevas metodologías para el análisis de datos heterogéneos haciendo uso de técnicas de *ensemble* [232, 515–519].

Supongamos un escenario de clasificación/regresión con un conjunto de datos de entrenamiento estadísticamente significativos según un test estadístico como t-Test, el test de Welch, ANOVA o el test de Mann-Whitney-Wilcoxon entre otros [148–151]. Puede ocurrir que el modelo generado a partir de estos datos no sea realmente generalizable debido, por ejemplo, a una mala selección de los tests estadísticos¹, o simplemente porque los datos

¹Por ejemplo si se utiliza algún tipo de test paramétrico como el t-Test, que asume

de entrenamiento no eran realmente representativos del problema subyacente que se quiere resolver. Una forma de atenuar este problema consiste en seleccionar de forma iterativa y aleatorizada varios subconjuntos de datos a partir de los datos de entrenamiento y prueba para generar un único modelo promediando los anteriores [45]. Este nuevo modelo de conjunto, que podría catalogarse dentro del conjunto de modelos de MV, es más generalizable e incluso más robusto frente a una mala selección de los parámetros de cada bloque o al ruido de los datos de entrada. No obstante, y aunque la propuesta parece solucionar muchas de las limitaciones del modelo de clasificación/regresión original, si la mayor parte de las características de entrada no son capaces de explicar el problema que se quiere resolver (bajas tasas de clasificación/regresión), puede ocurrir que el voto por mayoría simple resulte ineficaz ya que la contribución de varios bloques con escaso rendimiento puede eclipsar la que proporcionan los bloques con mejor comportamiento. La solución en este caso consistiría en ponderar los resultados de los bloques con mejores tasas de clasificación/regresión siguiendo un esquema de doble validación equivalente al que tendría un sistema de voto por mayoría ponderado (*Weighted Majority Voting*, WMV) [45].

Entre las propuestas que podemos encontrar en el campo de la neuroimagen, la mayoría de modelos de clasificación/regresión con *ensemble* que hacen uso de una metodología con WMV suelen realizar el pesado de las características de entrada basándose en el estudio que se quiere hacer sobre un conjunto reducido de las mismas ya sea porque se quiere analizar algún biomarcador clínico relacionado con una patología, o bien porque se haya determinado que la significancia estadística de esas variables era relevante² [412]. Entre las limitaciones que presentan este tipo de soluciones podemos señalar: 1) que no son automáticas; 2) que dependen de una interpretación del modelo que puede no corresponderse con la realidad; y 3) que tienden a despreciar variables de entrada que pueden ayudar al rendimiento del sistema completo.

7.1.2. Propuesta

En un escenario de clasificación/regresión con *ensemble* haciendo uso de WMV, el pesado de las características de entrada puede hacerse según algún tipo de métrica calculada de manera automática utilizando únicamente los datos de entrenamiento como variables de entrada [520]. Siguiendo el esquema representado en la Figura 7.1, el trabajo con título ***Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Par-***

una normalidad de los datos de entrada, en escenarios con datos que presentan una fuerte impulsividad, asimetría y/o colas pesadas.

²Ya sea utilizando los datos de entrada de ese mismo estudio o por experiencia de estudios anteriores.

kinson's Disease [54], propone el uso de un bucle anidado encargado de determinar una métrica ³ a partir de la cual se pueden determinar los pesos de cada tipo de variable de entrada; y un segundo bucle (externo) que se encarga de generar el modelo real de los datos de entrada haciendo uso de los pesos calculados en el bucle interno. Aunque los pesos de entrada pueden utilizarse directamente para determinar el modelo de *ensemble* ponderado, este añade una etapa adicional que penaliza las variables con peor rendimiento.

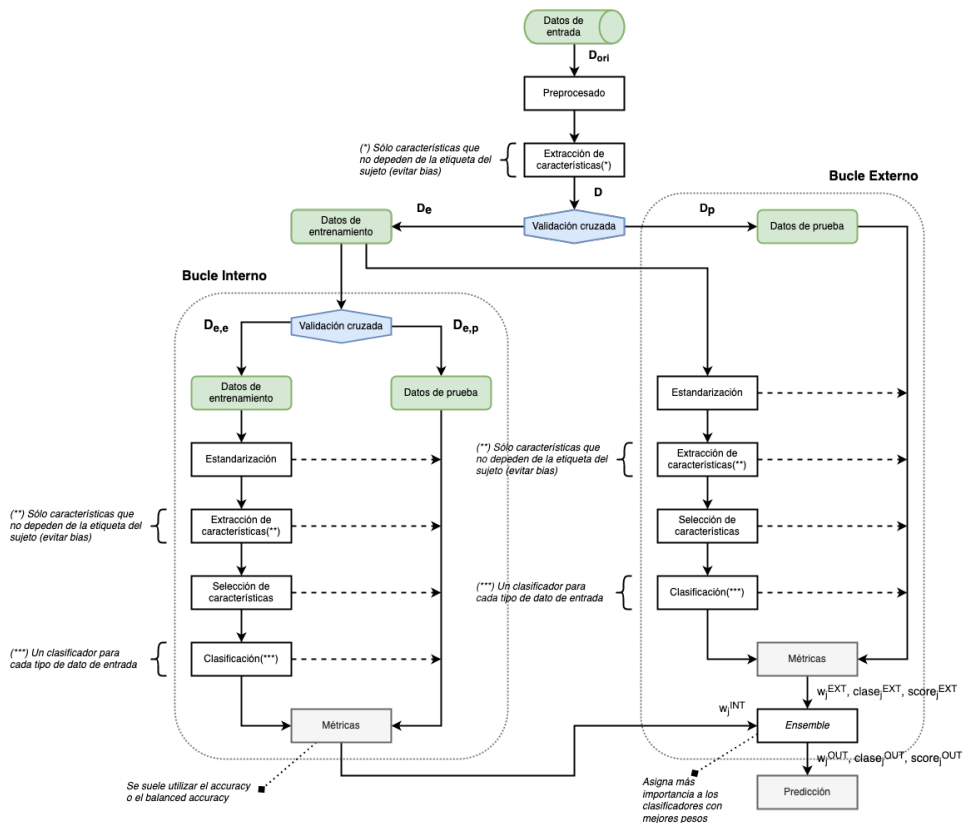


Figura 7.1: Diagrama del modelo de clasificación con *ensemble* propuesto.

El funcionamiento del modelo (explicado usando una métrica como la tasa de clasificación de los datos de entrada⁴) queda descrito como sigue a continuación:

³Significancia estadística, tasa de clasificación/regresión, o simplemente una selección de características relevantes según un algoritmo predeterminado.

⁴Aunque podría haberse utilizado cualquier otro tipo de métrica como el *score* que devuelven algunos tipos de clasificadores, las etiquetas predichas por los clasificadores e incluso valores relacionados con la significancia estadística de los datos de entrada.

1. Sea \mathbf{D}_{ori} la matriz de datos de entrada que contiene la información de n variables de N tipos de datos diferentes procedentes de S sujetos, esta información se preprocesa teniendo en cuenta que ninguna de las transformaciones que se apliquen a los datos de entrada pueda depender de las etiquetas de los sujetos (evitar el sesgo). Al final de este paso se habrá generado una matriz \mathbf{D} que servirá como entrada del modelo de *ensemble*.
2. La matriz de datos de entrada \mathbf{D} se divide en dos conjuntos de datos en el bucle externo: $\mathbf{D}_{\mathbf{e}}$ (entrenamiento) y $\mathbf{D}_{\mathbf{p}}$ (prueba). Por tanto, $\mathbf{D} = \mathbf{D}_{\mathbf{e}} \cup \mathbf{D}_{\mathbf{p}}$.
3. El conjunto de datos de entrenamiento se divide a su vez en dos nuevos subconjuntos de datos en el bucle anidado: $\mathbf{D}_{\mathbf{e},\mathbf{e}}$ (entrenamiento) y $\mathbf{D}_{\mathbf{e},\mathbf{p}}$ (prueba).
4. Sea \mathbf{c}_j el conjunto de vectores columna con las realizaciones de las características de entrada de tipo j para el conjunto de datos de entrenamiento $\mathbf{D}_{\mathbf{e},\mathbf{e}}$ con $j = 1, 2, \dots, N$, entonces el bucle anidado se entrena para todas las características en \mathbf{c}_j . Este paso se lleva a cabo para cada conjunto \mathbf{c}_j en $\mathbf{D}_{\mathbf{e},\mathbf{e}}$ y se evalúan sus resultados con $\mathbf{D}_{\mathbf{e},\mathbf{p}}$.
5. A la salida del bucle anidado se habrán generado j resultados de clasificación (uno por cada tipo de dato de entrada). Llamaremos w_j^{INT} a cada uno de estos resultados.
6. En el bucle externo, los datos \mathbf{C}_j de cada conjunto de datos de entrenamiento $\mathbf{D}_{\mathbf{e}}$ se utilizan para generar un nuevo modelo que se evalúa con los datos de prueba $\mathbf{D}_{\mathbf{p}}$ y resulta en las tasas de clasificación w_j^{EXT} .
7. Las tasas de clasificación del bucle anidado w_j^{INT} y las del bucle externo w_j^{EXT} , se utilizan como argumentos de entrada en el bloque de *ensemble*. Tras combinar los parámetros w_j^{INT} y w_j^{EXT} se habrán generado unos nuevos parámetros w_j^{OUT} a partir de los cuales se pueden predecir las etiquetas de los datos de prueba $\mathbf{D}_{\mathbf{p}}$.
8. En el bloque de *ensemble* se puede añadir una etapa adicional llamada “enventanado” que modifica los pesos de los parámetros de entrada penalizando aquellos conjuntos de datos de entrada con bajas tasas de clasificación [521].

7.1.3. Implementación

Para la evaluación de la metodología propuesta se ha desarrollado un sistema CAD para el diagnóstico de la EP combinando la información procedente de diferentes tipos de marcadores disponibles en el repositorio PPMI incluyendo:

- **Valores de intensidad [VAF]** - Tomando como referencia las imágenes FP-CIT SPECT correspondientes a la primera visita de cada sujeto en el estudio PPMI, y una vez se han preprocesado estas imágenes mediante la aplicación de una normalización espacial (usando como referencia el atlas *Automated Anatomical Labeling*, [183]) así como una normalización en intensidad con distribuciones α -estables ([374]); se han parcelado las regiones de interés para el estudio de la EP (putamen + núcleo caudado) y se han utilizado sus valores de intensidad como variables de entrada.
- **Características morfológicas [Morp]** - Para cada región de interés, se han extraído una serie de características morfológicas incluyendo el área la región, su volumen, su centro de masas (medido respecto a los valores de intensidad), su esfericidad y diferentes valores relacionados con la proyección de los valores de intensidad máximos de cada región de interés tal y como se describe en [393].
- **Marcadores del líquido cefalorraquídeo [CSF]** - Incluye diferentes marcadores como proteínas, lípidos y residuos metabólicos obtenidos a partir de una muestra de líquido cefalorraquídeo extraída mediante una punción lumbar [386–388].
- **Marcadores sanguíneos [Plasma y Serum]** - Incluye valores de marcadores presentes en el plasma y en el suero sanguíneo⁵ una vez se han extraído todas las células en el volumen de la muestra.
- **Marcadores genéticos [RNA]** - Compuesto principalmente de recuentos de proteínas asociadas a alteraciones genéticas [522].

Sobre la base de estos tipos de marcadores se han definido un total de 7 experimentos de clasificación probando diferentes combinaciones de marcadores⁶ como se indica en la Tabla 7.1. Este modelo de ayuda al diagnóstico de la EP ha sido diseñado teniendo en cuenta las siguientes dos consideraciones: 1) que haya balanceo de clases durante la fase de entrenamiento; y 2) que la influencia de cada tipo de marcador en la predicción final se determine comparando diferentes propuestas para el inventanado de los pesos de forma que se penalicen aún más aquellos marcadores con bajas tasas de clasificación. Respecto a este el último punto, en la Tabla 7.2 se han incluido varias alternativas para la mejora del voto por mayoría ponderado. Como se

⁵Las diferencias entre ambos residen en que el plasma se obtiene a partir del líquido que queda cuando se evita la coagulación de la muestra mientras que el suero sanguíneo se obtiene a partir del líquido remanente una vez la muestra se ha coagulado.

⁶Aunque la idea original era proponer un único experimento que combinara todos los tipos de marcadores, ha sido necesario probar con diferentes combinaciones de marcadores debido a la falta de resultados de los tests clínicos para muchos de los sujetos del repositorio.

puede observar en la representación gráfica de cada una de estas propuestas incluidas en la Figura 7.2, aquellos marcadores con tasas de clasificación inferiores o iguales al 50% tendrán una contribución nula en la predicción de los datos de prueba de forma que no afectan a la tasa de clasificación final del sistema.

| Experimento | Nº sujetos | VAF | Morp | CSF | RNA | Serum |
|-------------|------------|-----|------|-----|-----|-------|
| 1 | 334 | ✓ | ✓ | ✓ | | |
| 2 | 150 | ✓ | ✓ | ✓ | ✓ | |
| 3 | 306 | ✓ | ✓ | ✓ | | ✓ |
| 4 | 148 | ✓ | ✓ | ✓ | ✓ | ✓ |
| 5 | 150 | ✓ | ✓ | | ✓ | |
| 6 | 148 | ✓ | ✓ | | ✓ | ✓ |
| 7 | 310 | ✓ | ✓ | | | ✓ |

Tabla 7.1: Relación de experimentos de clasificación propuestos. Al final no se han incluido los marcadores de tipo Plasma debido a su baja significancia estadística según el test de Welch como se verá más adelante.

| Método | Criterio |
|-----------------------------|--|
| MV simple | $\sum_j \text{Sign}(clase_j^{\text{EXT}})$ |
| WMV simple | $\sum_j \text{Sign}(clase_j^{\text{EXT}}) w_j^{\text{INT}}$ |
| WMV inventanado lineal | $\sum_j \text{Sign}(clase_j^{\text{EXT}}) (2w_j^{\text{INT}} - 0.5)$ |
| WMV inventanado cuadrático | $\sum_j \text{Sign}(clase_j^{\text{EXT}}) (w_j^{\text{INT}2} + 0.5w_j^{\text{INT}} - 0.5)$ |
| WMV inventanado exponencial | $\sum_j \text{Sign}(clase_j^{\text{EXT}}) (0.1e^{2.6175w_j^{\text{INT}}} - 0.3792)$ |
| Distancia a hiperplano | $\sum_j \text{Sign}(score_j^{\text{EXT}})$ |

Tabla 7.2: Propuestas para el inventanado de los pesos w_j^{INT} usando la predicción de las clases del bucle externo (escenario binario con etiquetas 1 y -1).

7.1.4. Análisis de resultados

La metodología propuesta en el apartado 7.1.2 ha sido evaluada para el conjunto de datos VAF, Morp, CSF, Plasma, Serum y RNA del repositorio PPMI siguiendo el esquema de la Figura 7.1. Entre los bloques seleccionados para este esquema se incluyen: 1) algoritmos de selección de características basados en la aplicación de un test estadístico como el test de Welch, [148]; 2) clasificadores de tipo SVM con *kernel* lineal [49, 485]; y 3) uso de esquemas

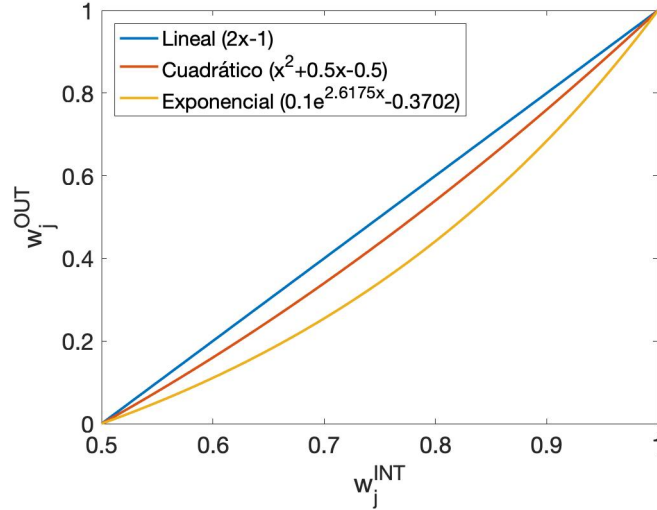


Figura 7.2: Representación gráfica de tres de las propuestas de inventanado para la propuesta de clasificación con *ensemble*.

de validación cruzada de tipo LOO, con el fin de que todos los experimentos pudieran ser comparables entre sí pese al reducido número de sujetos por clase en algunos casos [141]. Respecto a los datos procedentes de imagen funcional FP-CIT SPECT, se ha incluido una etapa de preprocesado de las imágenes que incluye:

- Una primera fase en la que todos los volúmenes FP-CIT SPECT son normalizados al espacio de una plantilla previamente generada siguiendo el procedimiento descrito en [483]. El registro de cada imagen se lleva a cabo mediante una transformación afín (si lo que se quieren obtener son marcadores de tipo morfológico, *Morp*) o no-lineal (para las comparaciones con marcadores *VAF*), utilizando *SPM12*.
- Una segunda fase en la que cada imagen se normaliza en intensidad siguiendo el método de normalización con distribuciones α -estables descrito en la Sección 4.4. Respecto a esta transformación cabe mencionar que, para el conjunto de imágenes incluidas en este trabajo, los histogramas con la distribución de intensidades de cada sujeto quedaron centrados en torno al valor $\delta = 28.42$ y dispersión $\gamma = 5.41$.
- Una tercera fase de extracción de características morfológicas a partir de las cuales se genera el conjunto de datos *Morp* según el procedimiento descrito en [393].

Antes de comenzar a desarrollar los resultados de clasificación, cabe mencionar que en un análisis preliminar de los datos de entrada realizado utilizando el test de Welch, se observó que los marcadores del tipo *Plasma* no

contribuían en absoluto a la separación entre sujetos sanos y pacientes con EP ($p_{\text{valor}} = 0.489$). De acuerdo con este análisis previo y junto a la falta de resultados de los marcadores asociados a este tipo de prueba clínica, se decidió descartar directamente este tipo de marcador y centrar el análisis en el resto de biomarcadores. Ahora sí, para cada experimento de clasificación utilizando los tipos de marcadores definidos en la Tabla 7.1, los resultados de clasificación en el bucle anidado para cada uno de estos conjuntos quedan como se indica a continuación en la Tabla 7.3.

Con el objetivo de asegurar la fiabilidad de estos resultados, se ha llevado a cabo un test de permutación no paramétrico sobre los conjuntos de biomarcadores con peores resultados de clasificación (CSF, RNA y Serum). Para la realización de cada uno de estos tests, se han generado un total de 1000 conjuntos de etiquetas diagnósticas aleatorias (con la misma longitud que la de los datos de entrenamiento/prueba) y se ha realizado la clasificación de los datos de entrada utilizando estas etiquetas aleatorias. A partir de los histogramas con los resultados de clasificación como los que se adjuntan en la Figura 7.3, y en combinación con una prueba de t-Test realizada a posteriori sobre estos resultados ($p_{\text{valor}} \leq 0.05$), se determina si se rechaza la hipótesis nula del test estadístico [523]. Como la condición para este rechazo se cumple en todos los casos (Tabla 7.4), se puede afirmar que los resultados de cada test de permutación no proceden de una distribución normal con media igual a la tasa de acierto calculada por la clasificación original (la que se obtiene utilizando las etiquetas originales de los datos de entrada).

Volviendo al modelo del diagrama general, y una vez iterado el bucle anidado, ya se puede proceder con la metodología de clasificación con *ensemble* probando uno a uno los diferentes métodos de enventanado incluidos en la Tabla 7.2. Los resultados de la clasificación final (en términos de su parámetro tasa de acierto), quedan resumidos como sigue a continuación en la Tabla 7.5. Entre los métodos incluidos, aquel que mejores resultados de clasificación consigue en todos los experimentos es el método de *ensemble* mediante voto por mayoría ponderada con enventanado cuadrático con una tasa máxima del 96.08 % de clasificación, lo que supone una mejora del 7.46 % respecto al método estándar del voto por mayoría. Si se utiliza el método cuadrático como referencia, se puede comprobar en la Figura 7.4 cómo los resultados de clasificación finales dependen casi en su totalidad de la contribución de los marcadores de imagen con un promedio de tasa de acierto igual a 94.38 % para VAF y a 90.64 % para Morp. De hecho si se evalúan los pesos w_j^{EXT} para cada tipo de marcador en el segundo experimento (por poner un ejemplo), se puede observar que mientras los marcadores de imagen tienen un peso cercano a 1 ($w_{\text{VAF}} = 0.90$ y $w_{\text{Morp}} = 0.77$), los marcadores procedentes de las pruebas CSF y RNA apenas superan el límite 0.1 ($w_{\text{CSF}} = 0.10$ y $w_{\text{RNA}} = 0.14$). Siguiendo con los resultados del método cuadrático, los pesos para cada tipo de marcador y experimento quedan como se refleja en la Tabla 7.6.

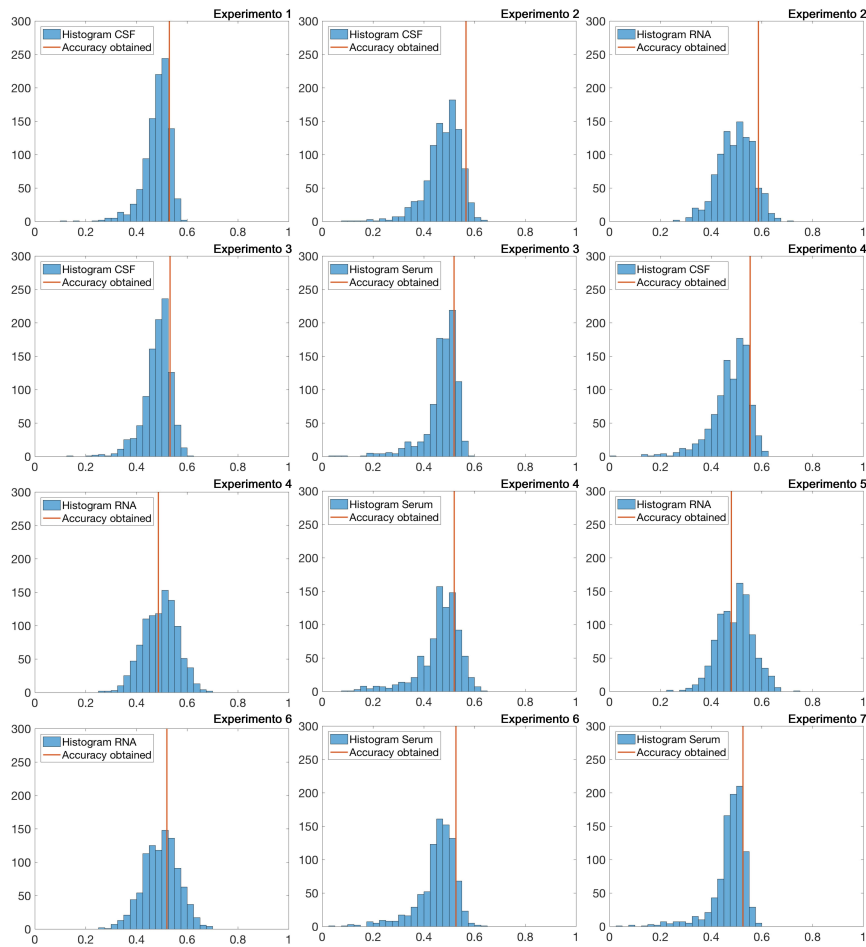


Figura 7.3: Resultados de los tests de permutación.

Por último, con el objetivo de subrayar la diferencia entre los conjuntos de biomarcadores CSF, RNA y Serum; y tomando como referencia el método de *ensemble* con enventanado cuadrático⁷, se ha incluido una representación de las curvas ROC asociadas a cada experimento en la Figura 7.5. Nótese que para cada curva también se han detallado los resultados del área bajo la curva (AUC) de cada marcador.

⁷Propuesta con mejores resultados de clasificación como se indica en la Tabla 7.5.

| Exp. | Parámetro | VAF | Morp | CSF | RBF | Serum |
|-------------|------------------|------------|-------------|------------|------------|--------------|
| 1 | Accuracy | 82.93 % | 88.32 % | 52.99 % | - | - |
| | Sensitivity | 84.43 % | 87.43 % | 73.05 % | - | - |
| | Specificity | 81.44 % | 89.22 % | 32.93 % | - | - |
| | Precision | 81.98 % | 89.02 % | 52.14 % | - | - |
| 2 | Accuracy | 96.00 % | 90.67 % | 56.67 % | 58.67 % | - |
| | Sensitivity | 96.00 % | 90.67 % | 74.67 % | 58.67 % | - |
| | Specificity | 96.00 % | 90.67 % | 38.67 % | 58.67 % | - |
| | Precision | 96.00 % | 90.67 % | 54.90 % | 58.67 % | - |
| 3 | Accuracy | 96.73 % | 91.50 % | 53.27 % | - | 51.96 % |
| | Sensitivity | 96.08 % | 91.50 % | 70.59 % | - | 24.18 % |
| | Specificity | 97.39 % | 91.50 % | 35.95 % | - | 79.74 % |
| | Precision | 97.35 % | 91.50 % | 52.43 % | - | 54.41 % |
| 4 | Accuracy | 96.62 % | 89.86 % | 55.41 % | 48.65 % | 52.03 % |
| | Sensitivity | 95.95 % | 89.19 % | 74.32 % | 47.30 % | 24.32 % |
| | Specificity | 97.30 % | 90.54 % | 36.49 % | 50.00 % | 79.73 % |
| | Precision | 97.26 % | 90.41 % | 53.92 % | 48.61 % | 54.55 % |
| 5 | Accuracy | 96.00 % | 91.33 % | - | 48.00 % | - |
| | Sensitivity | 96.00 % | 92.00 % | - | 49.33 % | - |
| | Specificity | 96.00 % | 90.67 % | - | 46.67 % | - |
| | Precision | 96.00 % | 90.79 % | - | 48.05 % | - |
| 6 | Accuracy | 95.95 % | 90.54 % | - | 52.03 % | 52.70 % |
| | Sensitivity | 94.59 % | 89.19 % | - | 50.00 % | 18.92 % |
| | Specificity | 97.30 % | 91.89 % | - | 54.05 % | 86.49 % |
| | Precision | 97.22 % | 91.67 % | - | 52.11 % | 58.33 % |
| 7 | Accuracy | 96.45 % | 92.26 % | - | - | 52.58 % |
| | Sensitivity | 95.48 % | 92.90 % | - | - | 23.87 % |
| | Specificity | 97.42 % | 91.61 % | - | - | 81.29 % |
| | Precision | 97.37 % | 91.72 % | - | - | 56.06 % |

Tabla 7.3: Promedio de resultados de la clasificación individual en el bucle anidado.

| Exp. | Tipo | Hipótesis nula | p_{valor} | Intervalo confianza | Estadísticos t-Test |
|------|-------|----------------|--------------------|---------------------|--|
| 1 | CSF | Rechazada | ≈ 0 | [0.478,0.485] | $t_{\text{stat}} = -29.218$ $\sigma = 0.052$ |
| 2 | CSF | Rechazada | ≈ 0 | [0.473,0.482] | $t_{\text{stat}} = -40.242$ $\sigma = 0.070$ |
| | RNA | Rechazada | ≈ 0 | [0.492,0.500] | $t_{\text{stat}} = -42.142$ $\sigma = 0.068$ |
| 3 | CSF | Rechazada | ≈ 0 | [0.479,0.485] | $t_{\text{stat}} = -29.630$ $\sigma = 0.054$ |
| | Serum | Rechazada | ≈ 0 | [0.424,0.442] | $t_{\text{stat}} = -18.967$ $\sigma = 0.145$ |
| 4 | CSF | Rechazada | ≈ 0 | [0.472,0.482] | $t_{\text{stat}} = -32.791$ $\sigma = 0.074$ |
| | RNA | Rechazada | ≈ 0 | [0.489,0.498] | $t_{\text{stat}} = 3.322$ $\sigma = 0.068$ |
| | Serum | Rechazada | ≈ 0 | [0.392,0.414] | $t_{\text{stat}} = -21.221$ $\sigma = 0.174$ |
| 5 | RNA | Rechazada | ≈ 0 | [0.490,0.498] | $t_{\text{stat}} = 6.512$ $\sigma = 0.068$ |
| 6 | RNA | Rechazada | ≈ 0 | [0.492,0.501] | $t_{\text{stat}} = -10.894$ $\sigma = 0.070$ |
| | Serum | Rechazada | ≈ 0 | [0.384,0.405] | $t_{\text{stat}} = -25.135$ $\sigma = 0.1668$ |
| 7 | Serum | Rechazada | ≈ 0 | [0.425,0.443] | $t_{\text{stat}} = -19.952$ $\sigma = 0.144$ |

Tabla 7.4: Resultados del t-Test aplicado a los resultados de clasificación del test de permutación. El estadístico σ hace referencia a la desviación típica del t-Test.

| Exp. | MV simple | WMV simple | WMV lineal | WMV cuadrát. | WMV exponen. | Distancia hiperplano |
|----------|-----------|------------|------------|----------------|--------------|----------------------|
| 1 | 85.63 % | 85.63 % | 88.02 % | 88.32 % | 85.63 % | 86.83 % |
| 2 | 88.67 % | 94.67 % | 94.67 % | 94.67 % | 94.67 % | 91.33 % |
| 3 | 85.62 % | 95.75 % | 95.42 % | 96.08 % | 95.75 % | 93.79 % |
| 4 | 83.11 % | 93.92 % | 94.59 % | 95.27 % | 89.86 % | 88.51 % |
| 5 | 93.33 % | 93.33 % | 96.00 % | 96.00 % | 94.00 % | 92.67 % |
| 6 | 77.70 % | 93.24 % | 94.59 % | 95.27 % | 95.27 % | 90.54 % |
| 7 | 85.68 % | 92.76 % | 93.88 % | 94.27 % | 92.53 % | 90.61 % |

Tabla 7.5: Resultados de clasificación final (en términos del parámetro tasa de acierto) obtenidos al aplicar diferentes métodos de *ensemble*.

| Exp. | w_{VAF} | w_{Morp} | w_{CSF} | w_{RNA} | w_{Serum} |
|----------|-----------|------------|-----------|-----------|-------------|
| 1 | 0.60 | 0.72 | 0.05 | - | - |
| 2 | 0.90 | 0.77 | 0.10 | 0.14 | - |
| 3 | 0.92 | 0.79 | 0.05 | - | 0.03 |
| 4 | 0.92 | 0.76 | 0.08 | 0.00 | 0.03 |
| 5 | 0.90 | 0.79 | - | 0.00 | - |
| 6 | 0.90 | 0.77 | - | 0.03 | 0.04 |
| 7 | 0.91 | 0.81 | - | - | 0.04 |

Tabla 7.6: Pesos obtenidos para cada tipo de dato cuando se usa el método de enventanado cuadrático.

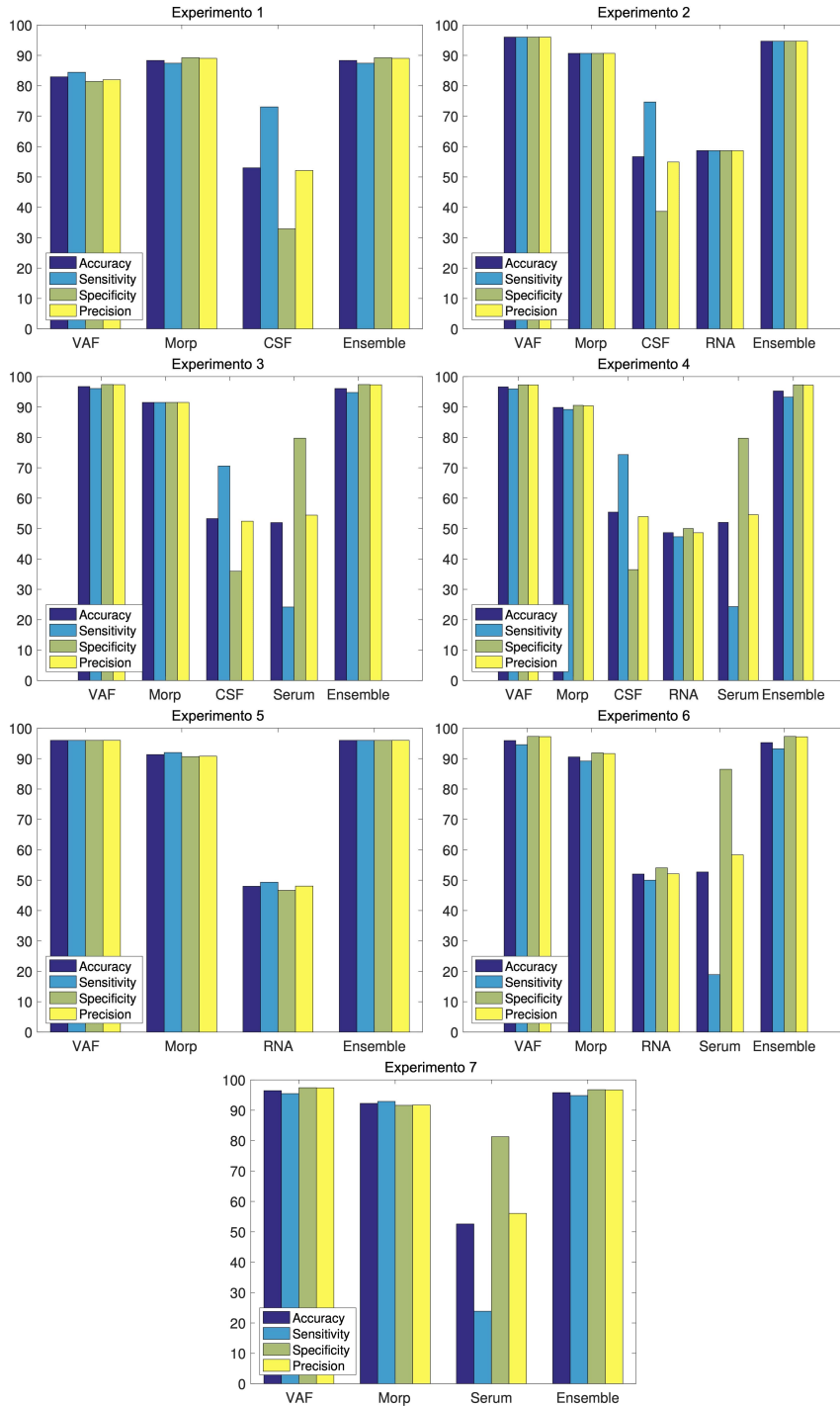


Figura 7.4: Resultados de clasificación por tipo de marcador y cuando se aplica el método de *ensemble* con enventanado cuadrático.

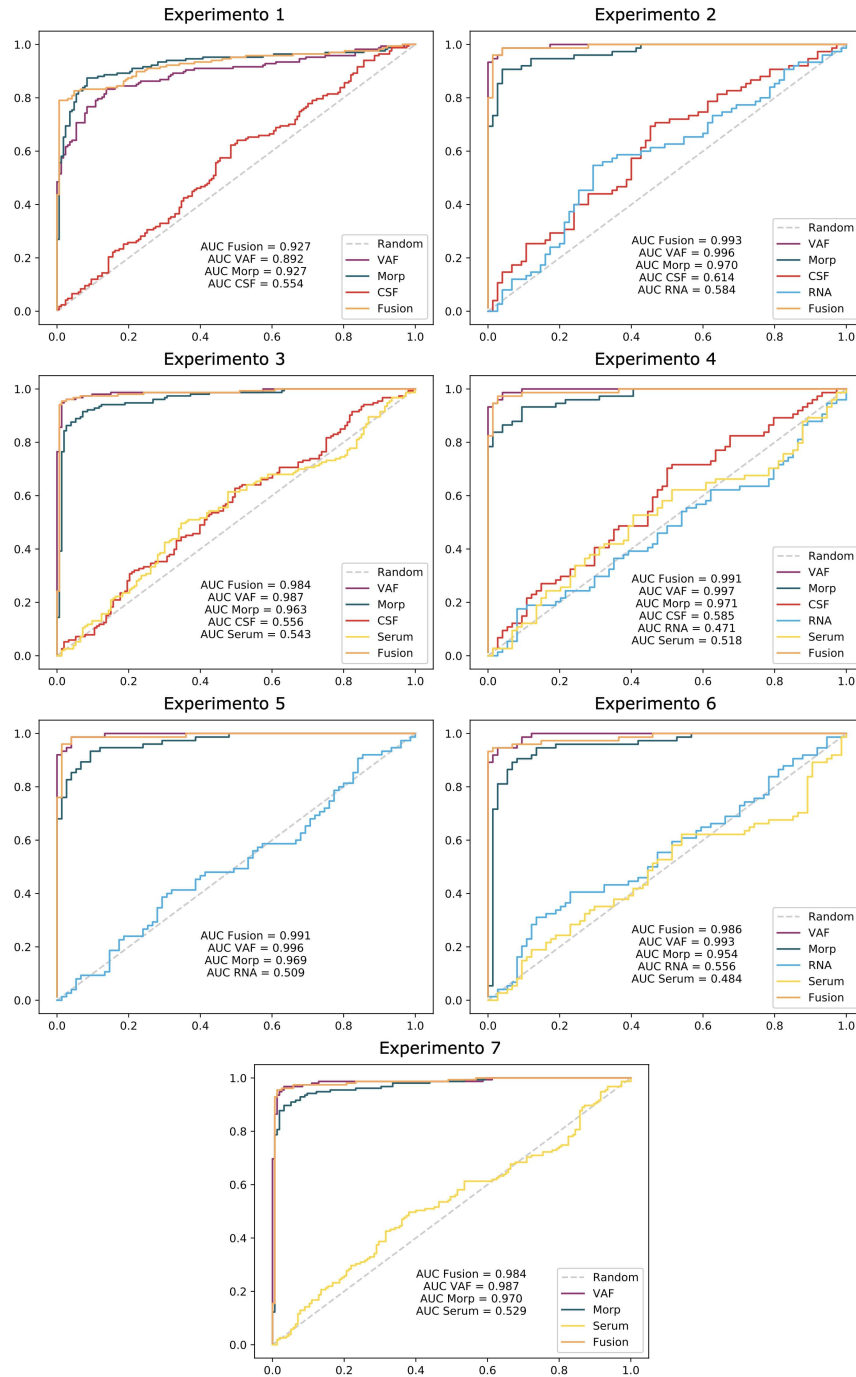


Figura 7.5: Curvas ROC asociadas a cada experimento de clasificación para cada tipo de marcador junto a la curva ROC que se obtiene aplicando el método de *ensemble* con inventariado cuadrático.

Bloque E

Discusión, conclusiones y futuras líneas de trabajo

Capítulo 8

Discusión y Conclusiones

Las aportaciones presentadas para la realización de esta tesis doctoral ya han sido comentadas en detalle en los capítulos anteriores. Este último capítulo de discusión resume las aportaciones de cada trabajo al campo de la neuroimagen y cómo cada una de estas contribuciones ayuda, o puede ayudar, a mejorar nuestra comprensión sobre la demencia.

8.1. Discusión

En líneas generales, y tal y como se indicó en la Sección 1.2 sobre los objetivos generales de la tesis doctoral, este trabajo constituye una mejora de los métodos de análisis en neuroimagen en tres de sus puntos más importantes:

1. Introducir, adaptar y/o aplicar nuevos métodos matemáticos con un mejor ajuste de los datos de entrada para los modelos de ayuda al diagnóstico (Capítulo 4).
2. Proponer nuevos biomarcadores para evaluar trastornos neurodegenerativos como la EA o la EP (Capítulo 5).
3. Evaluar fuentes de datos heterogéneas y desarrollar un método que permita combinar toda la información disponible en algunas de las bases de datos más relevantes a nivel mundial para el estudio de la demencia (Capítulos 6 y 7).

8.1.1. Uso de distribuciones alfa-estables en el preprocesado de imágenes neuronales

El trabajo publicado en [51] (Sección 4.3) supone una mejora sustancial en la parametrización de los tejidos cerebrales bajo la suposición de mezclas de distribuciones con colas pesadas. Si se compara el modelo propuesto

durante esta tesis doctoral con el trabajo publicado por **Zhang et al.** en [206] donde los autores combinan los métodos EM+HMRF con el uso de distribuciones gaussianas para la segmentación de los tejidos GM y WM, las ventajas de combinar los métodos EM+HMRF con distribuciones α -estables son varias. Por un lado, el uso de distribuciones α -estables permite modelar cada histograma de intensidad utilizando solo una distribución. En cambio, bajo un modelo con mezclas de distribuciones gaussianas casi siempre la primera de las componentes intentará ajustarse a los datos más extremos de cada distribución de intensidad. Esto se observa en los resultados de los experimentos de mezclas de distribuciones de la Sección 4.2 cuando se utilizan distribuciones gaussianas (Figuras 4.2 y 4.3). La consecuencia directa de este comportamiento, es que cualquier modelo de mezclas de distribuciones α -estables casi siempre requerirá un menor número de componentes que un modelo de distribuciones gaussianas para explicar la misma cantidad de información. Además, este mismo razonamiento se puede aplicar también al ajuste de la simetría de cada distribución que se quiere modelar gracias al parámetro β , por lo que se enfatiza más si cabe el efecto positivo del uso de este tipo de distribuciones. Por otro lado, y como puede comprobarse a partir de los resultados obtenidos en las Secciones 4.2 y 4.3 para los modelos de mezclas de distribuciones y de segmentación respectivamente, como toda distribución gaussiana constituye un caso límite de una distribución α -estable para $\alpha = 2$ y $\beta = 0$, se puede afirmar que cualquier resultado que se consiga ajustando una distribución α -estable siempre será igual o mejor que el que se conseguiría utilizando una distribución gaussiana [30]. Según este desarrollo, los resultados del modelo de segmentación publicado en [51] no son más que una aplicación directa del modelo de mezclas de distribuciones α -estables presentado en [60]. Y aunque ya existían modelos anteriores que habían tratado de modelar una mezcla de distribuciones α -estables [470, 471, 524], todos estos dependían de unas condiciones de contorno tan restrictivas que (en algunos casos) era incluso necesario fijar algunos de los parámetros de cada distribución para conseguir un ajuste en muchos casos peor del que se consigue utilizando el modelo presentado aquí.

Por último, y comparando la propuesta del modelo de mezclas con el método EM frente al caso extremo que sería ajustar cada componente mediante un modelo bayesiano con un muestreo de Monte-Carlo (como se hace en el trabajo publicado en [471]), la diferencia entre ambos trabajos en cuanto a tiempo de ejecución es considerable. Para justificar esta afirmación, se ha evaluado el tiempo de ejecución de un modelo de mezclas de distribuciones, en este caso relacionadas con un proceso relacionado con la actividad enzimática en la sangre [525], utilizando ambas propuestas y se han comparado sus tiempos de ejecución. Los resultados de esta comparación han demostrado que el tiempo de ejecución con el algoritmo EM fue hasta 7.65 veces más rápida que la ejecución utilizando el método bayesiano. En este sentido, y aunque el algoritmo EM está más limitado de cara a la necesidad

de ajustar apropiadamente sus condiciones iniciales, se entiende que la solución con el método EM compensa su limitación con tiempos de ejecución mucho más reducidos.

Por último, y como una de las aplicaciones más importantes sobre el uso de distribuciones α -estables en el campo de la neuroimagen, cabe destacar los resultados de la normalización en intensidad aplicada en todos los trabajos con imagen funcional presentados para la tesis doctoral. Todas las contribuciones incluidas en este trabajo han hecho uso del método de normalización en intensidad con distribuciones α -estables propuesto originalmente por **Salas-Gonzalez et al.** en [374] (Sección 4.4) a excepción de los marcadores FDG y PIB de la base de datos de DIAN¹. Esta decisión se justifica a partir de los resultados obtenidos en trabajos anteriores, como el publicado en [179], donde los autores demuestran que la normalización con distribuciones α -estables constituye uno de los métodos que mejor ajuste consigue a los datos de entrada y que, además, cuenta con la ventaja adicional de conseguir una mayor separación entre clases (medida utilizando parámetros como la divergencia de Kullback-Leibler, la distancia euclidiana, la distancia de Hellinger o el resultado del test de Mann-Whitney Wilcoxon), confirmando los resultados del trabajo presentado en [52] (Sección 4.4).

Futuras líneas de trabajo

Aunque las aplicaciones prácticas de las distribuciones α -estables son numerosas, a corto plazo en el campo de la neuroimagen se pueden señalar varios puntos donde su implementación podría mejorar los modelos actuales o como herramientas de ayuda para el modelado de datos:

1. Aunque la mezcla de distribuciones constituye uno de los pasos fundamentales en la segmentación de imágenes estructurales o durante el proceso de registro espacial, herramientas como *SPM* todavía ejecutan una gran cantidad de procesos intermedios que requieren de modelos estadísticos basados en distribuciones gaussianas [171]. Con base en esto, se propone la adaptación de todos los procedimientos que hagan uso de mezclas de distribuciones gaussianas por modelos de mezclas de distribuciones α -estables.
2. Igual que existen pruebas como el test de Anderson-Darling [28] o el test de Shapiro-Wilk [29] que nos permiten medir la “normalidad” de un conjunto de datos de entrada, sería interesante desarrollar un test que sea capaz de medir la impulsividad o el peso de las colas de una distribución. Este test ayudaría a justificar, aún más si cabe, el uso

¹Estas características aparecen originalmente en la base de datos como valores SUVR finales tras aplicar el mismo proceso de normalización y segmentación siguiendo los procedimientos descritos en [110, 111].

de distribuciones α -estables sobre todo en escenarios de clasificación/-regresión con modelos matemáticos que asumen la normalidad de los datos de entrada (como por ejemplo ocurre cuando se utiliza ANOVA como método de selección de características).

3. Por último, también se explora la posibilidad de utilizar distribuciones α -estables como distribuciones a priori para el modelado de pesos en redes profundas de tipo bayesiano y sus implementaciones variacionales.

8.1.2. Marcadores morfológicos para el análisis de la demencia

La búsqueda de nuevos biomarcadores para el modelado de enfermedades como la EA o la EP constituye una de las líneas de investigación más importantes en la actualidad dentro del campo de la neurociencia. Aunque en muchos trabajos casi siempre se suele hablar de marcadores biológicos como ocurre en [379], lo cierto es que los marcadores de imagen también constituyen una potente herramienta que ha permitido profundizar en el conocimiento que tenemos hoy en día sobre el funcionamiento del cerebro humano [526–529].

Comparando el uso de isosuperficies con otros trabajos relacionados

En lo que respecta al trabajo realizado para esta tesis doctoral, la propuesta original presentada en [53] llevó a la conclusión de que el uso de la magnitud del gradiente u otra herramienta equivalente, podría llevar a una caracterización morfológica de la EP más fácil de interpretar y sin la necesidad de renunciar al ajuste estadístico que se consigue con otras propuestas basadas en un análisis a nivel de vóxeles de intensidad. Con la aparición de los primeros modelos de análisis morfológico en imagen funcional para evaluar la EP, como los trabajos publicados en [41, 395, 530], surgió la idea de utilizar isosuperficies como una medida más discriminativa y precisa capaz de explicar visualmente la distribución de intensidades en imágenes FP-CIT SPECT de pacientes con EP. Esta idea supone un avance importante respecto a los citados modelos anteriores por varios motivos:

1. Las isosuperficies representan volúmenes cerrados en 3D cuyo contorno modela la distribución de intensidad de la imagen funcional a diferencia de los modelos presentados por **Molina-Molina et al.** y **Segovia et al.**, donde los autores hacen uso de contornos o proyecciones en 2D; y a diferencia de [530] donde los autores pretenden modelar un plano tridimensional generado a partir de un conjunto de capas (*slices*) seleccionadas manualmente y que en ningún caso son capaces de

- modelar por completo la distribución de intensidad de la ROI a la que hacen referencia.
2. Las isosuperficies no requieren de la definición ni del uso de ninguna ROI para delimitar el estriado ni ninguna otra estructura cerebral de interés como en [530].
 3. Tampoco se requiere de una normalización espacial no-lineal. Es más, si se aplica una normalización afín de las imágenes es por hacer comparables las magnitudes morfológicas de cada sujeto entre sí pero incluso esta puede no ser estrictamente necesaria.
 4. Según el método empleado para definir los umbrales de intensidad, puede que tampoco sea necesario utilizar ningún tipo de normalización en intensidad. Aunque en la publicación original presentada en [59] se habla de varias propuestas para definir los umbrales de intensidad, la forma desarrollada en la Sección 5.1 no requiere de ningún tipo de normalización en intensidad.
 5. Mejora los resultados de clasificación en comparación con el resto de modelos (Tabla 8.1).

| Ref. | Muestra | Tipo de característica | Resultado |
|-------|---------|--|-----------|
| [530] | 715 | Slices seleccionados manualmente + Ajuste de superficie de 3 ^o grado | 97.2 % |
| [54] | 388 | Como en [41] pero usando I_{\max} | 92.3 % |
| [395] | 147 | Líneas de contorno + 4 slices del estriado | 92.5 % |
| [41] | 189 | Proyec. en 2D del estriado ($0.8 \cdot I_{\max}$) | 94.2 % |

Tabla 8.1: Comparativa de los resultados de clasificación reportados en trabajos en trabajos anteriores sobre el uso de características morfológicas para el estudio de la EP.

Nótse que en esta comparativa no se han incluido los resultados del trabajo publicado por **Martínez-Ibáñez et al.** en [531], donde los autores también hace referencia al uso de un volumen de intensidad para caracterizar sujetos con EP utilizando una red CNN, por varios motivos: 1) los autores no han justificado el umbral de intensidad seleccionado; 2) no se proporciona información alguna sobre si la red está incurriendo en un sobreajuste a los datos de entrenamiento (más de 12.6 millones de características en total para cada sujeto de entrada); y 3) la red no proporciona una interpretación física de los datos de entrada o de qué regiones del volumen son las que más contribuyen a la clasificación de los sujetos de entrada.

Características más discriminativas y diferencias entre hemisferios

Si nos fijamos en las variables morfológicas con mejor significancia estadística según el test de Mann-Whitney-Wilcoxon incluidas en la Tabla 5.4 para $N = 20$, la mayoría de estas pertenecen a isosuperficies situadas entre las capas 12 – 14. Este hecho coincide con los resultados del trabajo anterior sobre el uso de la magnitud del gradiente de imágenes FP-CIT SPECT para el estudio de la EP presentado en [53] y valida la afirmación de que es el propio contorno del estriado el que realmente contribuye a diferenciar los sujetos sanos de los pacientes con EP.

Otro de los aspectos más interesantes de los resultados presentados en la Sección 5.1.4, es que se observa una ligera diferencia (en términos de clasificación) entre los hemisferios izquierdo y derecho cuando se evalúan sus características por separado. Ya sea utilizando un clasificador de tipo SVM o el propio ν SVC, estas diferencias (en torno al 0.5 – 0.6 %) no hacen más que evidenciar una afectación a veces no bilateral (pero bien conocida) en este tipo de demencias [118, 532].

Comparando los resultados obtenidos con VAF y PCA

Los resultados de clasificación obtenidos utilizando isosuperficies han sido comparados frente a métodos de clasificación más tradicionales como VAF o PCA siguiendo el mismo esquema general (Figura 2.4.2) pero teniendo en cuenta algunas consideraciones adicionales:

- **VAF.** Requiere aplicar una normalización espacial (no-lineal) de las imágenes de entrada a un espacio común. En este caso, una plantilla generada siguiendo el procedimiento descrito en [483]. También requiere de una normalización en intensidad como en [7]. Con el fin de evitar el sobreajuste del clasificador a los datos de entrenamiento, se ha definido una ROI a partir de la plantilla generada anteriormente seleccionando aquellas posiciones cuya intensidad superaba el 50 % de la intensidad máxima de la plantilla. Este límite ha sido fijado experimentalmente como se indica en la Figura 8.1. Empleando un esquema de validación cruzada de tipo 10-*fold* y un clasificador SVM con *kernel* lineal², se consigue un resultado máximo de clasificación en términos del parámetro *balanced accuracy* equivalente al 85.67 %.
- **PCA.** Siguiendo el mismo esquema de procesado que el definido para VAF, se ha aplicado un análisis con PCA probando diferentes valores

²Las diferencias entre utilizar SVM y ν SVC radican en la selección del hiperparámetro $\nu = 0.1$ que penaliza las clasificaciones incorrectas. Como este término se puede reescribir en términos del parámetro C de SVM, se puede afirmar que ambos esquemas son equivalentes.

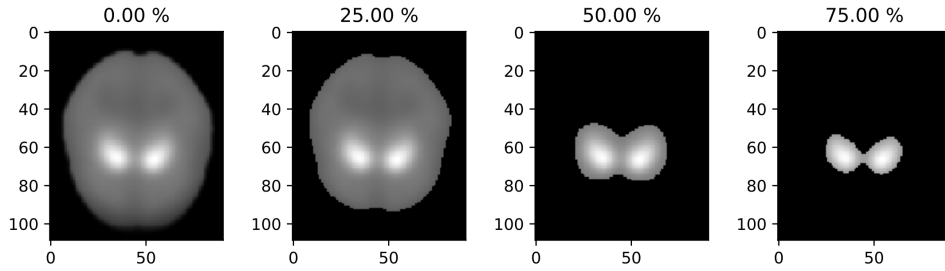


Figura 8.1: Representación gráfica de una ROI definida utilizando diferentes umbrales de intensidad para el análisis VAF.

del parámetro N_{Comp} . En este escenario, y haciendo uso también de un esquema de validación cruzada de tipo *10-fold* y un clasificador SVM con *kernel* lineal, se alcanza un *balanced accuracy* máximo de 88.89% cuando $N_{\text{Comp}} = 24$. En la Figura 8.2 se han incluido los resultados de clasificación para diferentes valores del parámetro N_{Comp} junto a una justificación del porcentaje total de la varianza explicada por el sistema.

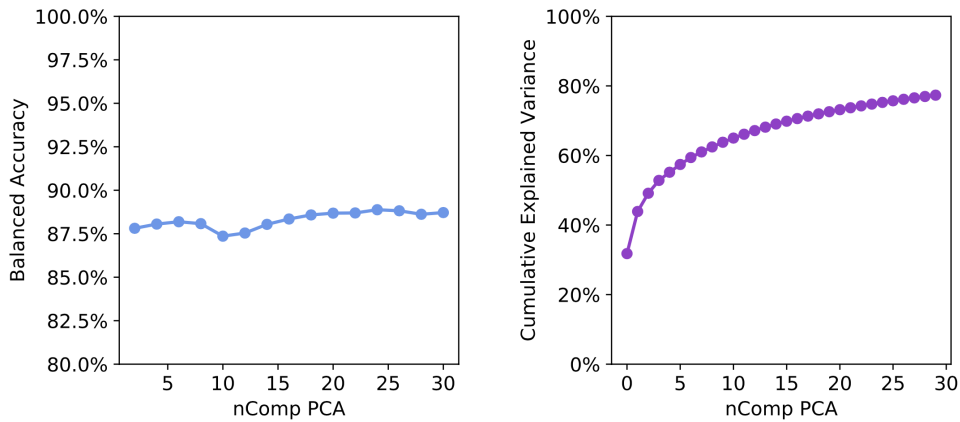


Figura 8.2: Resultados de clasificación (en términos del parámetro *balanced accuracy*) para diferentes valores del parámetro N_{Comp} utilizando el modelo con PCA (izquierda). Varianza acumulada en función del parámetro N_{Comp} (derecha).

A la vista de estos resultados, se demuestra la fiabilidad de los marcadores morfológicos propuestos en [59] y se justifica su capacidad para discernir entre sujetos sanos y pacientes con EP mediante un enfoque que consigue reducir drásticamente el número de variables de entrada (en comparación con VAF) y que mantiene la interpretación física de los resultados (a diferencia de PCA).

Futuras líneas de trabajo

La evaluación de sujetos con un diagnóstico probable de EP mediante isosuperficies presenta la ventaja biológica de que los focos de intensidad en imagen FP-CIT SPECT siempre se encuentran concentrados en la zona del estriado. Si se quiere extender este método a otros escenarios como la segmentación de lesiones en pacientes con esclerosis múltiple o el análisis de la morfología de algún proceso oncogénico con metástasis, entonces será necesario realizar un correcto etiquetado de cada foco de intensidad para poder analizar su comportamiento y/o evolución. Sobre la base de esto se definen dos propuestas:

1. Aunque existen métodos, como los que proporcionan librerías como `Qhull` que permiten etiquetar cada región aislada por separado, existen casos en los que la proximidad de dos focos de intensidad pueden traducirse en una única isosuperficie cuyos patrones morfológicos (incluyendo proyecciones, coeficientes de esfericidad y demás) puede afectar de forma negativa al resultado final de una clasificación/regresión. Así, se propone analizar la topología geométrica de las lesiones más frecuentes en escenarios de neuroimagen y desarrollar un modelo de aprendizaje que permita detectar este tipo de hallazgos radiológicos anómalos.
2. A la hora de evaluar a pacientes con algún tipo de demencia mediante imagen médica, es común encontrar patologías para las que la atrofia de una o varias regiones del cerebro es relevante [84, 85]. En este tipo de escenarios, la aplicación de isosuperficies puede jugar un papel relevante para cuantificar cómo evolucionan este tipo de regiones con el tiempo (análisis longitudinal) sin la necesidad de imponer la aplicación de una normalización no-lineal a un espacio definido por una plantilla.

8.1.3. Análisis de marcadores procedentes de fuentes heterogéneas

Evaluación de la base de datos DIAN

Muchos estudios señalan que la caracterización de las etapas presintomáticas de trastornos neurodegenerativos como la EA o la EP constituye un factor clave para el desarrollo de nuevas líneas de tratamiento [82, 101, 111, 404, 533]. Particularmente, la histología de la EA no debería considerarse como una entidad con una etiología única [534]. Algunas variantes de esta demencia (entre las que se debe incluir su forma hereditaria) evolucionan de forma diferente y esto ha llevado a pensar a algunos investigadores que cada una de estas formas podría tener su origen en uno o varios mecanismos biológicos diferentes que conducen a las neuronas de la corteza

cerebral a un mismo fin: la cascada de procesos biológicos que llevan a la muerte celular. La consecuencia más inmediata de esta afirmación es que si se analizan de manera adecuada cada uno de estos mecanismos se podrían proponer nuevas líneas de tratamiento capaces de frenar e incluso detener el curso de la enfermedad. Esta idea contrasta con la mayoría de ensayos clínicos llevados a cabo en los últimos años para los cuales no se ha intentado diferenciar entre grupos de sujetos con características más homogéneas. Esto es, discriminar las clases de entrada de pacientes con EA en función de su histopatología exacta, del grado de afectación cognitiva, o de cualquier otro tipo de característica relevante que pudiera facilitar el análisis de cada variante por separado.

En el marco descrito hasta ahora, la base de datos de DIAN ha aportado a la comunidad científica un conjunto de datos procedentes de individuos bien diferenciados en función del gen responsable de su enfermedad y/o del estado general del sujeto en el momento de cada visita al estudio (lo que permite homogeneizar la muestra de entrada). El análisis de cada subconjunto por separado teniendo en cuenta diferentes tipos de marcadores de entrada (pruebas clínicas principalmente de imagen pero también incluyendo otros tipos de marcadores como la presencia de ciertas proteínas, lípidos, productos metabólicos y demás) puede ayudar no solo a justificar el modelo propuesto por **Jack et al.** [500] (Figura 6.1), sino que también contribuye a determinar la utilidad de cada tipo de prueba clínica. Esto último tiene un impacto directo sobre la calidad de vida del paciente al reducir la cantidad de pruebas y tiempo necesario para su diagnóstico y seguimiento, pero también sobre el ahorro económico (y de tiempo) que supondría poder prescindir de algunas de estas pruebas por ser poco relevantes o por tener una relación de coste/beneficio diagnóstico baja.

Es por todo ello que en la siguiente lista se ha intentado incidir sobre los aspectos más relevantes de cada uno de los experimentos desarrollados para el trabajo presentado en [58] incluyendo, para cada escenario, una comparativa entre los resultados obtenidos aquí y los publicados en trabajos anteriores:

- *Comparación entre NC y MC (Experimento 1)*. Primero considerando todos los estados de la enfermedad (MC sintomáticos y asintomáticos como un único grupo) y luego homogeneizando la muestra de forma que se incluyeran, dentro del grupo de los MC, solo a sujetos asintomáticos. En ambos casos las tasas de clasificación han resultado ser muy pobres (Figura 6.2), pero ligeramente favorables cuando se analizan los marcadores de tipo PIB.

A pesar de que la lista de regiones significativas según el test de ANOVA es elevada, como se puede comprobar a partir del mapa de significancia representado en la Figura 6.5, la superposición y la heterogeneidad de los datos de entrada limitan la eficacia del modelo de

clasificación con ML. Aún así, la relación de características de entrada que más han contribuido a la clasificación coincide con los resultados publicados en algunos trabajos previos como los presentados en [102, 402–404], donde se mencionan diferencias significativas entre NC y MC en regiones como el lóbulo temporal (a nivel medial en la zona hipocampo y lateralmente en el neocórtex), el cíneo, el precíneo, el giro cingulado, el putamen y el tálamo.

- *Subgrupos de MC (Experimento 2 - Parte I)*. Inicialmente se propuso el diseño de un experimento de clasificación multiclase que comparara cada subgrupo de MC (en función del gen responsable de su mutación) entre sí y frente al conjunto de los NC. Aunque se esperaba que la diferenciación de los MC proporcionara unos márgenes de clasificación lo suficientemente buenos como para poder relacionar cada tipo de dato de entrada con la patogénesis de cada variante de DIAD, lo cierto es que la clasificación entre NC, MC_{PSEN1} , MC_{PSEN2} y MC_{APP} , ha resultado ser insuficiente como se observa en la Figura 6.3.

Aunque hay varias hipótesis que podrían justificar los malos resultados del experimento multiclase, se entiende que los resultados de este se están más relacionados con la superposición de las realizaciones de casi todos los tipos de marcadores de entrada que con otro tipo de limitantes como la cantidad de sujetos que se han podido comparar (escenario de clasificación balanceado con 22 sujetos por clase limitados por la escasez de participantes en los subconjuntos MC_{PSEN2} y MC_{APP} como se indica en la Tabla 3.5). Es por esto que, con el objetivo de reducir la heterogeneidad de la muestra de entrada comparando solo uno de los subconjuntos de MC y eligiendo solo los casos de personas asintomáticas, se decidió finalmente comparar NC frente a MC_{PSEN1} asintomáticos obteniendo tasas de clasificación (en términos del parámetro *balanced accuracy*) equivalentes al 71.4%. Estos resultados ayudarían a justificar la capacidad de generalización del modelo y permiten proponer el uso de la imagen funcional como uno de los principales candidatos a la hora de modelar la evolución de los pacientes con DIAD en etapas todavía presintomáticas de su enfermedad.

Cuando se analizan los resultados del test de ANOVA sobre la significancia estadística de cada área cerebral al comparar NC frente a MC_{PSEN1} asintomáticos (Figura 6.5), se pueden destacar las siguientes regiones: el cerebelo, el córtex entorrinal, el córtex cingulado (incluyendo las zonas del córtex rostral anterior/posterior, el caudal anterior y el istmo), la región medial del córtex orbitofrontal, el polo frontal, el giro medio temporal, la circunvolución parahipocámpica, el cuerpo calloso, el *pars triangularis* (en el área de Broca), el giro precentral, el diencéfalo ventral, el tronco encefálico y la ínsula. Como hay muy pocos

trabajos que hayan intentado comparar los NC frente a los MC_{PSEN1} , no se han podido confirmar la mayoría de estos resultados. No obstante, se pueden mencionar los trabajos de **Lemére et al.**, [508], donde se hace referencia a cambios significativos en la región del cerebelo para un estudio llevado a cabo con un grupo reducido de sujetos con la mutación E280A en el gen PSEN1; el artículo de **Ringman et al.**, [405], que menciona una deposición elevada de β -amiloide en la WM en este subconjunto de MC (aunque sin especificar en qué regiones); y por último el trabajo de **Quiroz et al.** publicado en [406], en donde se mencionan diferencias tanto a nivel de β -amiloide $A\beta_{42}$ como del ratio $A\beta_{42}/A\beta_{40}$ entre NC y MC_{PSEN1} , así como cambios a nivel de imagen en regiones asociadas al precúneo, el córtex cingulado posterior, el lóbulo temporal (a nivel medial) y el hipocampo.

- *Subgrupos de MC (Experimento 2 - Parte II).* A partir de un análisis a posteriori sobre las distribuciones de los datos de entrada del subconjunto MC_{PSEN1} asintomáticos, se observa (sobre todo en imagen funcional FDG y PIB) la existencia de al menos dos componentes bien diferenciadas dentro de la distribución global (Figura 6.6). Según esto, se estableció que el análisis de cada una de estas subclases por separado podría ayudar a explicar por qué algunos MC_{PSEN1} presentan unas características tan similares a los NC frente a otros claramente diferenciados independientemente de la edad del sujeto (algunos incluso en etapas muy presintomáticas).

Tras la aplicación del Criterio de Información de Akaike junto al algoritmo de k-medias para determinar cada una de estas subclases (Figura 6.7), se analizaron las tasas de clasificación de la comparación entre NC y cada uno de estos clústers con resultados muy interesantes. Por un lado, tenemos un grupo de sujetos cuyas características son muy parecidas a las de los NC (MC_{PSEN1}^{NC+}) y para los que el solapamiento entre variables de entrada es muy alto, lo que nos lleva a resultados de clasificación que apenas superan el 65% de *balanced accuracy* en el mejor de los casos. Por el contrario, el conjunto de los MC_{PSEN1}^{NC-} sí que consigue ampliar este margen hasta el 80% de clasificación por todo lo contrario: variables con resultados muy diferentes a los que se obtienen cuando se evalúa el conjunto de los NC. Aunque en la Figura 6.8 solo se han incluido los resultados de la clasificación utilizando los marcadores FDG, las diferencias entre cada clúster y el conjunto de los NC se mantienen aproximadamente iguales (como mínimo un 6.1% de diferencia entre clasificaciones en el escenario más desfavorable para MC_{PSEN1}^{NC-}).

Otro punto importante de estos resultados en relación con las demográficas asociadas a cada clúster (Tabla 6.2), es que se justifica

que la separación entre MC_{PSEN1} es independiente de factores como el sexo, la edad e incluso el estado general de la enfermedad.

- *Análisis temporal (Experimento 3)*. Si comparamos el conjunto de los MC_{PSEN1} en diferentes etapas presintomáticas de la enfermedad, fijando diferentes umbrales para la variable EYO, se espera a priori que las diferencias entre NC y MC_{PSEN1} sean mayores cuanto más cerca se encuentran los sujetos de entrada la edad a la que empiezan a desarrollar sus primeros síntomas cognitivos. Sin embargo, y en relación con los últimos resultados publicados por **Jack et al.** y **Luckett et al.** en [502] y [535] respectivamente, existen etapas intermedias en las que esta diferencia puede incluso llegar a decrecer.

Para el trabajo presentado aquí, y siguiendo con lo incluido en la Figura 6.9, se observan las siguientes diferencias entre NC y MC_{PSEN1} asintomáticos a lo largo del tiempo:

- **MC_{PSEN1} con EYO ≥ 15 años.** Todavía no se detectan diferencias significativas entre NC y MC_{PSEN1} ni en PIB ni en NIB. Esto se debe a una falta de deposiciones de β -amiloide en el cerebro lo que implica a su vez una falta de subproductos derivados de la muerte celular que experimentan las neuronas del córtex y que podrían llegar a estar presentes en las pruebas de NIB. Por el contrario, tanto para MRI como para FDG sí que se alcanzan tasas que pueden llegar a superar la barrera del 70% de *balanced accuracy* confirmando la existencia de patrones estructurales y/o metabólicos muy sutiles en sujetos todavía muy jóvenes.

Aunque solo en [103] se menciona la existencia de algunas diferencias a nivel de imagen PiB, la mayoría de publicaciones relacionadas con el análisis temporal de los datos de DIAN hasta la fecha ([403, 449, 506, 536]) hablan más bien de patrones de atrofia cerebral y/o metabólica en regiones como el lóbulo paracentral y la circunvolución postcentral, la ínsula, el putamen, el córtex cingulado (caudal anterior y córtex rostral anterior), el giro fusiforme, la región medial del córtex orbitofrontal, la circunvolución parahipocámpica, el cuerpo calloso (anterior, medio y posterior), los plexos coroideos, el núcleo *accumbens*, el cuerno inferior de los ventrículos laterales, el córtex temporal inferior, el polo frontal, el cúneo, el *pars orbitalis* (en el área de Broca), el giro lingual y el giro supramarginal.

- **MC_{PSEN1} con EYO ≥ 10 años.** Aparecen los primeros cambios tanto en PIB como en NIB. A nivel estructural, la reducción del volumen del hipocampo es considerable coincidiendo con lo publicado en [103, 404]. También se han encontrado diferencias en el

núcleo *accumbens* y en la amígdala cerebral. Respecto a los marcadores de imagen funcional destacan regiones como el precúneo, el córtex entorrinal, la circunvolución precentral, el lóbulo parietal (tanto a nivel lateral como a nivel de su giro postcentral) y el córtex cingulado posterior. Todos estos resultados confirman los hallazgos publicados en [403, 514]. Existen más regiones estadísticamente significativas como el córtex poscentral, el polo frontal, el lóbulo paracentral, el cuerpo calloso, la corteza pericalcarina, el cerebelo, el cuerno inferior de los ventrículos laterales, los plexos coroideos, el *pars opecularis*, la parte lateral del córtex orbitofrontal, la ínsula, el giro supramarginal y el giro medio temporal; pero ninguno de ellos han sido citados en trabajos anteriores.

Por último, respecto a NIB, ninguno de los biomarcadores de entrada ha resultado ser estadísticamente significativo a diferencia de lo descrito por **Fagan et al.** en [449].

- **MC_{PSEN1} con EYO \geq 5 años.** Conforme la variable EYO decrece, los volúmenes de algunas estructuras cerebrales como el tálamo; el núcleo caudado; o los lóbulos temporal, parietal y occipital se van haciendo cada vez más pequeños en sujetos MC_{PSEN1} [404, 505, 536]. Se sabe que pocos años antes de aparecer los primeros síntomas se producen cambios en el metabolismo de áreas como el precúneo o el córtex entorrinal; y que las diferencias entre NC y MC se acentúan también debido al hipometabolismo y atrofia que experimenta la región del hipocampo. El análisis realizado en este trabajo confirma la práctica totalidad de estos resultados y también incluye diferencias en regiones como el cerebelo, el diencefalo ventral, el cuerpo calloso, el núcleo *accumbens*, los plexos coroideos, el *pars opecularis*, la ínsula, la amígdala cerebral, el giro frontal medial, la circunvolución precentral, el lóbulo paracentral, el córtex poscentral y el córtex cingulado posterior.

Aunque en la bibliografía actual se pueden encontrar diversos trabajos que han tratado de proponer un análisis de los datos de DIAN, hasta la fecha de publicación del trabajo presentado en [58] nunca se había planteado hacer este análisis con ML. Es más, a fecha de 1 de abril de 2021, y hasta donde se ha profundizado en la búsqueda de artículos relacionados, solo hay publicado un trabajo por parte de **Lockett et al.** donde los autores han planteado el uso de una red neuronal con 4 capas y 10 neuronas por capa para analizar longitudinalmente los marcadores PIB, FDG y MRI de una muestra de 131 MC y 74 NC sujetos del repositorio de DIAN [535]. Respecto a la lista de regiones estadísticamente significativas en esta comparación (representada en el diagrama incluido en la Figura), hay que señalar que aunque esta lista coincide en términos generales con los resultados presentados aquí los resultados de ambos trabajos no pueden ser comparados directamente por

varios motivos: 1) porque no se propone el mismo criterio para la subagrupar a los MC; 2) porque los autores de [535] no han justificado la selección de los individuos incluidos en su estudio; y 3) porque el uso de la red neuronal puede estar incurriendo en un sobreajuste de los datos de entrenamiento que no ha sido defendida mediante la inclusión de curvas de aprendizaje o cualquier otro mecanismo que ayude a determinar si la red neuronal se ha diseñado correctamente.



Figura 8.3: Regiones más significativas para el análisis longitudinal de la base de datos de DIAN propuesto por **Luckett et al.**

La primera conclusión general que se puede derivar del estudio desarrollado en el Capítulo 6 es que se puede asumir que las pruebas de imagen funcional basadas en el uso del radiotrazador ^{11}C -PiB proporcionan un modelo más ajustado de la enfermedad sobre todo a partir de etapas con un EYO < 15 años. Como era de esperar, la deposición de β -amiloide desempeña el papel más importante en el proceso neurodegenerativo de la demencia, al menos entre los marcadores disponibles en la base de datos de DIAN, con el interés adicional de que puede ayudar a predecir la evolución de un sujeto incluso en etapas presintomáticas de su enfermedad. Este hecho puede traducirse en la práctica clínica en un seguimiento más exhaustivo y eficiente de los pacientes con DIAD pero también podría extenderse a la práctica totalidad de casos de EA.

Segundo. A nivel de clasificación, este estudio ha demostrado que la subagrupación de los MC según el gen responsable de la enfermedad de cada sujeto no solo consigue que las clases de entrada sean más homogéneas, sino que también mejora la interpretación de los resultados obtenidos e incrementa las tasas de clasificación. Respecto a este último punto, los resultados más relevantes tienen que ver con la mejora (en términos del parámetro *balanced accuracy*) del 11.3% cuando se comparan NC frente a MC_{PSEN1}

asintomáticos, y del 18.3% en el caso de comparar NC frente a MC_{PSEN1}^{NC} asintomáticos. Aunque este comportamiento se mantiene para cualquier tipo de marcador de entrada, esta diferencia es mucho más pronunciada cuando se utilizan marcadores procedentes de imagen funcional (FDG o PIB). Si en vez de utilizar valores aislados de los ratios de intensidad para cada región del cerebro (SUVR) se utilizaran las imágenes completas de cada individuo³, es muy posible que la mejora en cuanto a separación entre clases y clasificación sea más pronunciada. Esto abre la puerta a la incorporación de técnicas de preprocesado más eficientes como las propuestas de normalización en intensidad presentada en la Sección 4.4, la segmentación de los tejidos descrita en la Sección 4.3, e incluso una caracterización morfológica aplicando los métodos descritos en la Sección 5.1.

Por último, y presentada como la conclusión más importante de este trabajo, el hecho de que aparezcan dos tipos de sujetos con características tan diferentes entre sí dentro del subconjunto de los MC_{PSEN1} es algo que nunca antes se había llegado a demostrar. Aunque en el trabajo publicado por **Ryan et al.** en [503], los autores sugirieron la existencia de al menos dos tipos de sujetos dentro del conjunto de los MC_{PSEN1} (diferenciados en función de su sintomatología en fases más avanzadas de la enfermedad), el hecho de que aparezcan estos dos subconjuntos en etapas presintomáticas refuerza la idea de que la EA puede presentar varios mecanismos de desarrollo diferentes que deberán evaluarse de manera más exhaustiva en el futuro.

Modelo de aprendizaje de conjuntos

Durante la realización de esta tesis se han analizado un gran número de variables procedentes de diferentes fuentes de datos heterogéneos entre los que cabe mencionar:

- Pruebas de imagen médica. Incluyendo marcadores procedentes tanto de imagen estructural (MRI) como de imagen funcional (18F-FDG, FP-CIT o 11C-PIB).
- Marcadores morfológicos extraídos a partir de imágenes funcionales.
- Biomarcadores que no están relacionados con la imagen médica como, por ejemplo, marcadores genéticos; proteínas y/o lípidos presentes en pruebas de plasma y/o en el suero sanguíneo; e incluso derivados en el líquido cefalorraquídeo.

Todos los análisis multivariantes con ML llevados a cabo durante este trabajo pretenden modelar el comportamiento de enfermedades como la EA o la EP a partir de uno o varios conjuntos de estos marcadores. En el caso

³Convenientemente seleccionadas por un equipo de radiólogos expertos.

particular del análisis presentado en el Capítulo 7, resultaba interesante intentar combinar el aprendizaje de varias fuentes de datos con el objetivo de desarrollar un modelo lo más ajustado posible a la realidad a costa de perder cierto grado de interpretación [537]. Para explicar este hecho, pensemos en la combinación de las salidas del aprendizaje de cada tipo de dato en el bucle anidado. Un resultado de clasificación que pondera el aprendizaje de varias fuentes de datos de entrada no tiene por qué venir siempre definido en origen por un mismo conjunto de marcadores de entrada aunque siempre habrá marcadores cuya influencia pueda mejorar notablemente el comportamiento del sistema global. Aun así, la metodología que se describe en el Capítulo 7, sobre el trabajo publicado en [54], presenta algunas ventajas significativas respecto a otros modelos más tradicionales de aprendizaje de conjuntos por varios motivos:

1. Consigue un modelo de clasificación robusto que evita el sesgo entre datos de entrenamiento/prueba gracias a su doble bucle de validación cruzada siguiendo el esquema descrito en la Figura 7.1 [153].
2. Evita el sobreajuste a los datos de entrenamiento mediante la inclusión de etapas de selección de características bien diseñadas evitando que el número de características de entrada en cada bloque clasificador sea superior al número de muestras de entrada.
3. Permite combinar cualquier número de fuentes de datos heterogéneas.
4. Permite utilizar cualquier tipo de métrica tanto para escenarios de clasificación como para escenarios de regresión. En el caso de estos últimos, habrá que modificar la forma en la que el sistema de enventanado pondera unas fuentes respecto a otras para intentar minimizar el efecto negativo que un error elevado en el bucle anidado puede tener sobre el sistema global.
5. Constituye una metodología totalmente flexible en cuanto a elección de bloques de extracción/selección de características, algoritmos de clasificación/regresión e incluso cualquier otro tipo de etapa intermedia que normalmente podemos encontrar dentro de un esquema con validación cruzada. Como ejemplo de esta flexibilidad se puede mencionar que la metodología publicada en [54] ha sido también evaluada haciendo uso de clasificadores de tipo KNN siguiendo las recomendaciones de los revisores de la revista científica *Frontiers in Neuroinformatics* (www.frontiersin.org) durante el proceso de revisión del trabajo.
6. Permite ajustar el peso de los datos de entrada gracias a su técnica de enventanado. Aunque la propuesta presentada aquí es totalmen-

te automática, el pesado puede incluso llegar a forzarse a los valores definidos por el usuario⁴.

A nivel de tasas de clasificación, si se comparan los resultados conseguidos con la metodología de *ensemble* propuesta respecto a los que se obtendrían en un escenario con MV tradicional, la mejora en términos de la tasa de clasificación es del 7.46%. Se entiende fácilmente que cuanto menos discriminativos son las fuentes de datos de entrada y mayor número de estas fuentes tengamos, mayor será la penalización que sufre el sistema de MV. En cambio, como nuestra propuesta incluye el método de inventanado, por muchas fuentes poco eficientes que se quieran incluir, estas no afectarán (o lo harán con un impacto mínimo) a la clasificación final. Si las tasas de clasificación individuales son muy altas, entonces los pesos de esos tipos de marcadores tendrán un valor cercano a la unidad. De hecho, en el escenario de clasificación propuesto utilizando los datos procedentes de la base de datos del PPMI, se observa que solo las fuentes de datos VAF y Morp (con tasas de acierto promedio de 94.4% y 90.6% respectivamente como se deriva de la Tabla 7.3) son las que realmente contribuyen a la clasificación final del sistema con pesos como los que se incluyen en la Tabla 7.6.

Aunque muchos trabajos anteriores han señalado la relevancia de ciertos tipos de marcadores NIB como la α -sinucleína o las proteínas $\tau/p\text{-}\tau$ como se indica en [384, 386, 538–540]; y pese a que algunos de los biomarcadores de este conjunto de datos de entrada conseguían excelentes resultados en cuanto a significancia estadística según el test de Welch; lo cierto es que de cara a la clasificación final esta fuente de datos no ha contribuido apenas a la clasificación final del sistema. Esto se ha visto reflejado en la clasificación final del sistema completo y del bucle anidado en particular donde también ha quedado demostrado, según los tests de permutación, que conforme aumenta la varianza de los resultados de clasificación también lo hace su p_{valor} asociado (Figura 7.3). Este hecho parece estar relacionado con el pequeño tamaño muestral de los grupos a comparar por lo que aumentar el tamaño de la muestra de entrada e incluso medir la distribución de cada variable se convierten en objetivos clave para terminar de descartar, o no, el conjunto de los biomarcadores NIB.

Como ya se comentó anteriormente en la Sección 2.8, la mayoría de los modelos de *ensemble* propuestos en los últimos años para el análisis de datos de neuroimagen están basados en técnicas de conjuntos tradicionales (como el voto por mayoría simple) o en el uso de redes neuronales. En la bibliografía actual aparecen pocas referencias a trabajos que hayan intentado proponer alguna metodología de clasificación nueva en el campo de la neuroimagen pero se pueden citar brevemente algunos de estos trabajos:

⁴No se recomienda esta opción porque la metodología en sí está preparada para intentar sacar el máximo rendimiento de cada fuente de datos de entrada de manera independiente.

- **Suk et al.**, [361] - En su forma más simplificada, propone combinar el aprendizaje de varios bloques de aprendizaje pero ajustados utilizando diferentes hiperparámetros para luego combinar el resultado del aprendizaje de todo el conjunto de bloques mediante una red CNN.
- **Grassi et al.**, [541] - Describe una combinación de *scores* de clasificación de varios clasificadores diferentes. Pero no queda claro cómo combina estos *scores* ni si son realmente comparables entre sí.
- **Lella et al.**, [542] - En realidad se trata de un modelo de voto por mayoría en el que se combinan las salidas de manera similar a como lo hace un perceptrón. De ahí el nombre de *softvoting ensemble*.

Como idea general, los modelos predictivos que definen y validan los patrones derivados del análisis de imágenes cerebrales se encuentran en muchos casos limitados por la falta de generalidad de los sistemas de clasificación o regresión asociados a un dominio de datos de entrenamiento limitado [543]. Al margen de la necesidad de ampliar el tamaño muestral, algunos trabajos han señalado la necesidad de extender cualquier tipo de análisis sobre un tipo de dato concreto a la búsqueda de relaciones entre varios tipos de biomarcadores o que al menos sean capaces de combinar la información procedente de varias fuentes de interés [544, 545]. En este contexto, y tal y como se señala en trabajos como [46–48], el incremento de la complejidad de sistemas de análisis como los basados en redes neuronales profundas y su aumento en las tasas de clasificación/regresión, no está teniendo en la práctica clínica el impacto que se esperaba para este tipo de soluciones debido, principalmente, a su falta de interpretabilidad y gran coste computacional. El desarrollo de metodologías como la de *ensemble* propuesta en esta tesis doctoral no solo ayudaría a incrementar las tasas de clasificación/regresión de problemas como los que tratan de resolver las redes neuronales profundas sino que además, nos permiten controlar qué tipos de marcadores (y en qué proporción) contribuyen a una mejor identificación de sujetos con demencia, predecir su evolución y tratar de caracterizar la patogénesis de la enfermedad.

Futuras líneas de trabajo

Tanto de cara al análisis individual de fuentes de datos heterogéneas como a su posible combinación mediante un modelo de aprendizaje de conjuntos, se pueden señalar varias líneas de actuación:

1. Sobre el análisis multivariante con fuentes de datos heterogéneas aplicado a la base de datos de DIAN:
 - Debido al reducido número de sujetos diagnosticados con LOAD en la base de datos de DIAN (solo 13 casos) y a la falta de datos

para aquellos pacientes que sí tenían la etiqueta diagnóstica, no se han podido realizar comparaciones entre ambos tipos de sujetos. Aunque algunos equipos de investigación han sugerido la idea de comparar los resultados de las bases de datos DIAN y ADNI (<http://adni.loni.usc.edu>), esta comparación debe hacerse con cuidado ya que en la actualidad el repositorio ADNI no incluye ninguna variable con la que se pueda definir si un sujeto con EA en el repositorio realmente pertenece al grupo de los LOAD o al de los DIAD.

- Proponer un análisis de regresión de los sujetos asintomáticos como el que se propone en [535] pero subagrupando el conjunto de los MC según el gen responsable de su enfermedad.
 - A raíz del comentario del trabajo publicado por **Ryan et al.** en [503], sería interesante determinar si realmente los sujetos de los subconjuntos MC_{PSEN1}^{NC+} y MC_{PSEN1}^{NC-} se corresponden con los que realmente expresan una sintomatología diferente en las fases más avanzadas de su enfermedad. Desde el punto de vista clínico este hecho supondría una revolución porque podría ayudar al personal especializado a intentar predecir la patogénesis que experimenta cada sujeto.
2. Sobre el análisis multivariante con fuentes de datos heterogéneas mediante aprendizaje de conjuntos:
- Incluir más fuentes de marcadores estadísticamente significativos que potencialmente puedan incrementar las tasas de clasificación finales.
 - Aplicar el modelo a un análisis de regresión con los sujetos del repositorio del PPML.
 - Probar más alternativas para el inventariado en el bloque de evaluación del bucle exterior.

8.2. Conclusiones (Versión en castellano)

El avance de la neurociencia ha impulsado la comprensión que hoy en día tenemos sobre el funcionamiento del cerebro humano y sus dolencias. La neuroimagen, desarrollada en principio como una herramienta de apoyo visual para confirmar el diagnóstico de patologías como las lesiones traumáticas, la demencia e incluso la identificación de algunos procesos oncogénicos; ha dado paso al desarrollo de una gran variedad de propuestas que, en combinación con su análisis mediante métodos estadísticos, no solo sirven como herramientas de apoyo al diagnóstico clínico tradicional, sino que también permiten medir la evolución de temporal de muchos de estos trastornos neurológicos.

En los últimos veinte años se ha pasado de análisis estadísticos tradicionales, que solo mencionaban las diferencias entre dos o más grupos de sujetos, a técnicas de análisis más avanzadas como el aprendizaje máquina. Estos modelos son capaces de predecir la condición de los pacientes con demencia e incluso el posible curso de su enfermedad. Pero para que esta capacidad de predicción sea lo más acertada posible, hace falta que se cumplan varias condiciones. La primera de ellas, tiene que ver con el modelado de las distribuciones de los datos de entrada. Existen sistemas, como los de segmentación de imágenes estructurales, que permiten explicar la distribución espacial de los tejidos cerebrales como una suma de distribuciones de probabilidad. Aunque en principio se puede asumir que la naturaleza de estas componentes es gaussiana, con base en el teorema central del límite, lo cierto es que su ajuste no consigue modelar los valores de mayor intensidad de la imagen ni justifica la impulsividad o la asimetría de algunas de sus componentes. Por el contrario, al utilizar distribuciones α -estables, no solo se consigue solventar este problema sino que además se ha conseguido desarrollar nuevos métodos de segmentación más eficientes y con aplicaciones prácticas diferentes a la propia segmentación. Uno de estos ejemplos es el método de normalización en intensidad de imágenes funcionales mediante distribuciones α -estables. Este sistema iguala las distribuciones de intensidad de imágenes de pacientes con Parkinson y sujetos sanos en imagen FP-CIT SPECT. Además, incrementa la distancia entre clases lo que facilita su diferenciación en problemas de clasificación (diagnóstico), y la cuantificación de su evolución en problemas de regresión (pronóstico).

El segundo de los puntos tiene que ver con la disponibilidad de datos que son realmente capaces de explicar la patogénesis del proceso neurológico que se quiere estudiar. Aunque en la actualidad la mayoría de ensayos clínicos a nivel mundial están centrados en el descubrimiento de algunas alteraciones genéticas, así como en el desarrollo de mecanismos que bloqueen la producción de ciertas moléculas y proteínas; el análisis de imágenes cerebrales con inteligencia artificial también puede contribuir, en este sentido, a la identificación de nuevos patrones estructurales y/o funcionales capaces

de cuantificar el proceso biológico de interés. Por poner un ejemplo, si estas nuevas metodologías determinan que existe una relación directa entre la atrofia de una determinada zona del cerebro y el grado de deterioro cognitivo de un paciente con la enfermedad de Alzheimer, este descubrimiento podría conducir a un estudio más exhaustivo de esa región para tratar de determinar la patogénesis de la enfermedad. El desarrollo de nuevos marcadores de imagen, como la propuesta de aplicación de isosuperficies para el estudio de la enfermedad de Parkinson, permite cuantificar el grado de neurodegeneración de un individuo e incluso predecir su sintomatología a corto y medio plazo.

En paralelo a la explicación anterior, y ya que uno de los limitantes más importantes de la medicina en la actualidad, y de la neurología en particular, es la falta de profesionales especializados y los costes asociados tanto a la formación de estas personas como a las pruebas clínicas necesarias para diagnosticar a un paciente; el análisis de estudios como DIAN resultan esenciales para aportar datos sobre la demencia incluso en fases presintomáticas de la enfermedad. Y esto se debe a dos motivos fundamentales. Por un lado, porque se justifica la necesidad de ciertos tipos de pruebas clínicas (en este caso imagen funcional 11C-PiB principalmente) frente a otro tipo de pruebas, dolorosas para el paciente en algunos casos como pueden ser las punciones lumbares, que realmente no son capaces de explicar el curso de la enfermedad. Y por otro lado, porque gracias al análisis de DIAN se confirma que el Alzheimer no debe tratarse como una enfermedad con una etiología única, y esto puede afirmarse incluso entre sujetos que comparten el mismo gen mutado responsable de su enfermedad.

Con la aparición de las primeras bases de datos que incluyen tanto resultados de imagen médica como de otros tipos de pruebas clínicas, se ha dado paso a la necesidad de combinar toda esa información para desarrollar un modelo de la demencia más preciso. En este contexto el aprendizaje de conjuntos, como herramienta capaz de combinar todas estas fuentes heterogéneas, ha estado tradicionalmente limitado por el peso que los conjuntos de datos poco significativos a la entrada han tenido sobre las salidas de estos sistemas. Con el fin de vencer esta limitación, la propuesta del método de conjuntos con pesado de voto por mayoría y técnica de inventariado se sitúa como una referencia a la hora de descartar fuentes de datos que, aunque puedan llegar a ser significativas desde el punto de vista de la estadística tradicional, en realidad no aportan información relevante al modelo de clasificación/regresión global.

En definitiva, la evolución del aprendizaje máquina en neurociencia requiere del desarrollo de nuevos modelos matemáticos más precisos; de una mejor caracterización de los procesos biológicos del sistema nervioso central a partir de nuevos marcadores (incluyendo características procedentes tanto de imagen cerebral como de cualquier otro tipo de prueba clínica); y de la combinación, en términos de aprendizaje, de todas las fuentes de informa-

ción relevantes que se le pueda proporcionar a un sistema que siempre está inmerso en un proceso de optimización continua.

8.3. Conclusions (English version)

The advance of neuroscience has boosted our understanding about the human brain and its disorders. Neuroimaging, initially developed as a visual support tool to confirm the diagnosis of pathologies such as traumatic injuries, dementia and even the identification of some oncogenic processes, has led to the development of a wide variety of approaches. These approaches, in combination with statistical methods, not only serve as support tools for traditional clinical diagnosis, but also allow us to measure the temporal evolution of many neurological disorders.

In the last twenty years, we have moved from traditional statistical analyses highlighting the differences between two or more groups of subjects to more advanced analysis techniques. These techniques, some of them applying machine learning, can predict the condition of patients with dementia and even the possible course of the disease. But for this predictive capability to be as accurate as possible, several conditions must be met. The first of these conditions concerns the modeling of the distributions of the input data. There are systems, such as structural image segmentation, that allow the spatial distribution of brain tissues to be explained as a mixture of probability distributions. Although the central limit theorem provides a theoretical basis for assuming that these components are gaussian, the truth is that these distributions fail at modelling high intensities, nor do they justify the impulsivity or asymmetry of some components. On the contrary, α -stable distributions can account for these features, helping to solve this problem. Furthermore, they have also led to the development of new, more efficient methods with practical applications other than segmentation. One of these examples is the method of intensity normalization of functional images via α -stable distributions, which helps to equalize intensities in FP-CIT SPECT imaging at the same time that increases the distance between classes. This last feature is of fundamental importance to enhance separation in classification (diagnosis) and regression (prognosis) problems.

The second point is related to the availability of data that can explain the pathogenesis of the neurological process to be studied. Nowadays most clinical trials are focused on the discovery of certain genetic alterations as well as on the development of mechanisms that block the production of certain molecules and proteins. In this sense, the analysis of brain images with artificial intelligence can also contribute to the identification of new structural and/or functional patterns capable of quantifying the biological process of interest. For example, if these new methodologies determine that there is a direct relationship between the atrophy of a certain brain area and the degree of cognitive impairment of a patient with Alzheimer's disease, it could lead to a more exhaustive study of that region in order to determine the pathogenesis of the disease. Based on this, the development of new imaging markers, such as isosurfaces in Parkinson's disease, makes it possible to

quantify the degree of neurodegeneration of an individual and even predict its symptomatology in the short and medium term.

In parallel, Artificial Intelligence techniques could help to overcome two of the most important limitations in medicine today: the lack of specialized professional and the costs associated with both training and clinical tests. In this regard, the analysis of studies such as DIAN are essential in providing data about dementia even in pre-symptomatic stages of the disease. There are two main reasons for this. On the one hand, because it justifies the need for certain types of clinical tests (in this case functional imaging ^{11}C -PiB mainly) as opposed to other types of tests such as lumbar punctures that are very painful for the patient and are not really able to explain the course of the disease. On the other hand, because the application of artificial intelligence to DIAN data has confirmed in this work that Alzheimer's disease should not be treated as a disease with a single etiology, even among subjects who share the same mutated gene responsible for their disease.

With the release of the first databases that include both medical imaging and other types of clinical tests, the need to combine all this information to build a more accurate model of dementia has arisen. In this context ensemble learning, as a tool capable of combining all these heterogeneous sources, has traditionally been limited by the weight that insignificant input information sources have had on the output of these systems. In order to overcome this limitation, we propose an ensemble learning methodology with majority voting weighting and windowing technique. This technique is positioned as a reference that can discard input data sources that do not contribute to the global classification/regression model although they may be significant from the point of view of traditional statistics.

In short, the evolution of machine learning in neuroscience requires the development of new and more accurate mathematical models; a better characterization of the biological processes of the central nervous system based on new markers; and the combination, in terms of learning, of all the relevant sources of information that can be provided to a system that is always immersed in a process of continuous optimization.

Bibliografía

- [1] M. D. Rosenberg, B. J. Casey, and A. J. Holmes, “Prediction complements explanation in understanding the developing brain,” *Nature Communications*, vol. 9, feb 2018.
- [2] K. Sakai and K. Yamada, “Machine learning studies on major brain diseases: 5-year trends of 2014-2018,” *Japanese Journal of Radiology*, vol. 37, pp. 34–72, nov 2018.
- [3] U. Raghavendra, U. R. Acharya, and H. Adeli, “Artificial intelligence techniques for automated diagnosis of neurological disorders,” *European Neurology*, vol. 82, no. 1-3, pp. 41–64, 2019.
- [4] B. Khagi, G.-R. Kwon, and R. Lama, “Comparative analysis of alzheimer’s disease classification by CDR level using CNN, feature selection, and machine-learning techniques,” *International Journal of Imaging Systems and Technology*, vol. 29, pp. 297–310, mar 2019.
- [5] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. Gilardi, and A. Quattrone, “Machine learning on brain MRI data for differential diagnosis of parkinson’s disease and progressive supranuclear palsy,” *Journal of Neuroscience Methods*, vol. 222, pp. 230–237, jan 2014.
- [6] F. Segovia, I. Illan, D. Salas-Gonzalez, F. Martinez-Murcia, A. Ortiz, J. Gorriz, and J. Ramirez, “PETRA: A web-based system supporting computer aided diagnosis of alzheimer’s disease,” in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, IEEE, oct 2016.
- [7] I. Illán, J. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, F. Segovia, R. Chaves, M. Gómez-Río, and C. Puntonet, “18f-fdg pet imaging analysis for computer aided alzheimer’s diagnosis,” *Information Sciences*, vol. 181, pp. 903–916, feb 2011.
- [8] J. Ramírez, J. Górriz, D. Salas-Gonzalez, A. Romero, M. López, I. Álvarez, and M. Gómez-Río, “Computer-aided diagnosis of alzheimer’s disease,” *Journal of Medical Imaging and Biomedical Engineering*, vol. 1, pp. 1–10, 2017.

- mer's type dementia combining support vector machines and discriminant set of features," *Information Sciences*, vol. 237, pp. 59–72, jul 2013.
- [9] L. Khedher, J. Ramírez, J. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, mar 2015.
- [10] F. Segovia, I. A. Illán, J. M. Górriz, J. Ramírez, A. Rominger, and J. Levin, "Distinguishing parkinson's disease from atypical parkinsonian syndromes using PET data and a computer system based on support vector machines and bayesian networks," *Frontiers in Computational Neuroscience*, vol. 9, nov 2015.
- [11] S. Badoud, D. V. D. Ville, N. Nicastró, V. Garibotto, P. R. Burkhard, and S. Haller, "Discriminating among degenerative parkinsonisms using advanced 123i-ioflupane spect analyses," *NeuroImage: Clinical*, vol. 12, no. Supplement C, pp. 234–240, 2016.
- [12] A. Augimeri, A. Cherubini, G. L. Cascini, D. Galea, M. E. Caligiuri, G. Barbagallo, G. Arabia, and A. Quattrone, "Cada-computer-aided datscan analysis," *EJNMMI Physics*, vol. 3, no. 1, p. 4, 2016.
- [13] R. Wang, B. Xu, Z. Guo, T. Chen, J. Zhang, Y. Chen, and H. Zhu, "Suite PET/CT neuroimaging for the diagnosis of parkinson's disease," *Nuclear Medicine Communications*, vol. 38, pp. 164–169, feb 2017.
- [14] T. Tong, K. Gray, Q. Gao, L. Chen, and D. Rueckert, "Multi-modal classification of alzheimer's disease using nonlinear graph fusion," *Pattern Recognition*, vol. 63, pp. 171–181, 2017.
- [15] N. Nicastró, J. Wegrzyk, M. G. Preti, V. Fleury, D. Van de Ville, V. Garibotto, and P. R. Burkhard, "Classification of degenerative parkinsonism subtypes by support-vector-machine analysis and striatal 123i-fp-cit indices," *Journal of Neurology*, vol. 266, pp. 1771–1781, Apr 2019.
- [16] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, and M. Filippi, "Automated classification of alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks," *NeuroImage: Clinical*, vol. 21, p. 101645, 2019.
- [17] D. Jarrett, E. Stride, K. Vallis, and M. J. Gooding, "Applications and limitations of machine learning in radiation oncology," *The British Journal of Radiology*, vol. 92, p. 20190001, aug 2019.

- [18] M. Alber, A. B. Tepole, W. R. Cannon, S. De, S. Dura-Bernal, K. Garikipati, G. Karniadakis, W. W. Lytton, P. Perdikaris, L. Petzold, and E. Kuhl, “Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences,” *npj Digital Medicine*, vol. 2, nov 2019.
- [19] L.-A. Claude, J. Houenou, E. Duchesnay, and P. Favre, “Will machine learning applied to neuroimaging in bipolar disorder help the clinician? a critical review and methodological suggestions,” *Bipolar Disorders*, vol. 22, pp. 334–355, mar 2020.
- [20] Y. Hase, K. Horsburgh, M. Ihara, and R. N. Kalaria, “White matter degeneration in vascular and other ageing-related dementias,” *Journal of Neurochemistry*, vol. 144, pp. 617–633, jan 2018.
- [21] S. E. Nasrabady, B. Rizvi, J. E. Goldman, and A. M. Brickman, “White matter changes in alzheimer’s disease: a focus on myelin and oligodendrocytes,” *Acta Neuropathologica Communications*, vol. 6, mar 2018.
- [22] G. Donatelli, R. Ceravolo, D. Frosini, M. Tosetti, U. Bonuccelli, and M. Cosottini, “Present and future of ultra-high field mri in neurodegenerative disorders,” *Current Neurology and Neuroscience Reports*, vol. 18, p. 31, apr 2018.
- [23] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T. L. Jones, T. R. Barrick, F. A. Howe, and X. Ye, “Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels,” *Computer Methods and Programs in Biomedicine*, vol. 157, pp. 69–84, apr 2018.
- [24] F. Raschke, T. R. Barrick, T. L. Jones, G. Yang, X. Ye, and F. A. Howe, “Tissue-type mapping of gliomas,” *NeuroImage: Clinical*, vol. 21, p. 101648, 2019.
- [25] J. Ashburner and K. J. Friston, “Unified segmentation,” *NeuroImage*, vol. 26, pp. 839–851, jul 2005.
- [26] H. Greenspan, A. Ruf, and J. Goldberger, “Constrained gaussian mixture model framework for automatic segmentation of MR brain images,” *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1233–1245, sep 2006.
- [27] H. Merisaari, R. Parkkola, E. Alhoniemi, M. Teräs, L. Lehtonen, L. Haataja, H. Lapinleimu, and O. S. Nevalainen, “Gaussian mixture model-based segmentation of MR images taken from premature infant brains,” *Journal of Neuroscience Methods*, vol. 182, pp. 110–122, aug 2009.

- [28] T. W. Anderson and D. A. Darling, “Asymptotic theory of certain ”goodness of fit criteria based on stochastic processes,” *The Annals of Mathematical Statistics*, vol. 23, pp. 193–212, jun 1952.
- [29] S. Shapiro and M. Wilk, “An analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, pp. 591–611, dec 1965.
- [30] J. P. Nolan, “Numerical calculation of stable densities and distribution functions,” *Communications in Statistics. Stochastic Models*, vol. 13, pp. 759–774, jan 1997.
- [31] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, and E. W. Lang, “Why using the alpha-stable distribution in neuroimage?,” in *Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications*, SCITEPRESS - Science and Technology Publications, 2014.
- [32] A. Öhrfelt, U. Andreasson, A. Simon, H. Zetterberg, Å. Edman, W. Potter, D. Holder, V. Devanarayan, J. Seeburger, A. D. Smith, K. Blennow, and A. Wallin, “Screening for new biomarkers for subcortical vascular dementia and alzheimer’s disease,” *Dementia and Geriatric Cognitive Disorders Extra*, vol. 1, pp. 31–42, jan 2011.
- [33] L. H. Meeter, L. D. Kaat, J. D. Rohrer, and J. C. van Swieten, “Imaging and fluid biomarkers in frontotemporal dementia,” *Nature Reviews Neurology*, vol. 13, pp. 406–419, jun 2017.
- [34] C. M. Gambino, B. L. Sasso, G. Bivona, L. Agnello, and M. Ciaccio, “Aging and neuroinflammatory disorders: New biomarkers and therapeutic targets,” *Current Pharmaceutical Design*, vol. 25, pp. 4168–4174, dec 2019.
- [35] P. N. E. Young, M. Estarellas, E. Coomans, M. Srikrishna, H. Beaumont, A. Maass, A. V. Venkataraman, R. Lissaman, D. Jiménez, M. J. Betts, E. McGlinchey, D. Berron, A. O’Connor, N. C. Fox, J. B. Pereira, W. Jagust, S. F. Carter, R. W. Paterson, and M. Schöll, “Imaging biomarkers in neurodegeneration: current and future practices,” *Alzheimer’s Research & Therapy*, vol. 12, apr 2020.
- [36] H. Zetterberg and B. B. Bendlin, “Biomarkers for alzheimer’s disease – preparing for a new era of disease-modifying therapies,” *Molecular Psychiatry*, vol. 26, pp. 296–308, apr 2020.
- [37] A. Meyer-Lindenberg, “From maps to mechanisms through neuroimaging of schizophrenia,” *Nature*, vol. 468, pp. 194–202, nov 2010.

- [38] R. Penadés, N. Franck, L. González-Vallespí, and M. Dekerle, “Neuroimaging studies of cognitive function in schizophrenia,” in *Advances in Experimental Medicine and Biology*, pp. 117–134, Springer International Publishing, 2019.
- [39] S. Tavakol, J. Royer, A. J. Lowe, L. Bonilha, J. I. Tracy, G. D. Jackson, J. S. Duncan, A. Bernasconi, N. Bernasconi, and B. C. Bernhardt, “Neuroimaging and connectomics of drug-resistant epilepsy at multiple scales: From focal lesions to macroscale networks,” *Epilepsia*, vol. 60, pp. 593–604, mar 2019.
- [40] A. Li, A. Zalesky, W. Yue, O. Howes, H. Yan, Y. Liu, L. Fan, K. J. Whitaker, K. Xu, G. Rao, J. Li, S. Liu, M. Wang, Y. Sun, M. Song, P. Li, J. Chen, Y. Chen, H. Wang, W. Liu, Z. Li, Y. Yang, H. Guo, P. Wan, L. Lv, L. Lu, J. Yan, Y. Song, H. Wang, H. Zhang, H. Wu, Y. Ning, Y. Du, Y. Cheng, J. Xu, X. Xu, D. Zhang, X. Wang, T. Jiang, and B. Liu, “A neuroimaging biomarker for striatal dysfunction in schizophrenia,” *Nature Medicine*, vol. 26, pp. 558–565, mar 2020.
- [41] F. Segovia, J. M. Gorriz, J. Ramirez, F. J. Martinez-Murcia, and D. Castillo-Barnes, “Assisted diagnosis of parkinsonism based on the striatal morphology,” *International Journal of Neural Systems*, vol. 29, p. 1950011, mar 2019.
- [42] K. L. Moulder, B. Snider, S. L. Mills, V. D. Buckles, A. M. Santacruz, R. J. Bateman, and J. C. Morris, “Dominantly inherited alzheimer network: facilitating research and clinical trials,” *Alzheimer’s Research & Therapy*, vol. 5, p. 48, oct 2013.
- [43] J. C. Latourelle, M. T. Beste, T. C. Hadzi, R. E. Miller, J. N. Oppenheim, M. P. Valko, D. M. Wuest, B. W. Church, I. G. Khalil, B. Hayete, and C. S. Venuto, “Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed parkinson’s disease: a longitudinal cohort study and validation,” *The Lancet Neurology*, vol. 16, pp. 908–916, nov 2017.
- [44] D. Opitz and R. Maclin, “Popular ensemble methods: An empirical study,” *Journal of Artificial Intelligence Research*, vol. 11, pp. 169–198, aug 1999.
- [45] L. Rokach, *Pattern Classification Using Ensemble Methods*. World Scientific Publishing Company, 2009.
- [46] A. M. Chekroud and N. Koutsouleris, “The perilous path from publication to practice,” *Molecular Psychiatry*, vol. 23, pp. 24–25, nov 2017.

- [47] H. Choi, “Deep learning in nuclear medicine and molecular imaging: Current perspectives and future directions,” *Nuclear Medicine and Molecular Imaging*, vol. 52, pp. 109–118, nov 2017.
- [48] Y. Li, J. Jiang, J. Lu, J. Jiang, H. Zhang, and C. Zuo, “Radiomics: a novel feature extraction method for brain neuron degeneration disease using 18f-FDG PET imaging and its implementation for alzheimer’s disease and mild cognitive impairment,” *Therapeutic Advances in Neurological Disorders*, vol. 12, p. 175628641983868, jan 2019.
- [49] Vapnik, *Statistical Learning Theory*. New York, USA: John Wiley & Sons, 1998.
- [50] F. J. Martínez-Murcia, J. M. Górriz, and J. Ramírez, “Computer-aided diagnosis in neuroimaging,” in *Computer-aided Technologies - Applications in Engineering and Medicine*, InTech, dec 2016.
- [51] D. Castillo-Barnes, I. Peis, F. J. Martínez-Murcia, F. Segovia, I. A. Illán, J. M. Górriz, J. Ramírez, and D. Salas-Gonzalez, “A heavy tailed expectation maximization hidden markov random field model with applications to segmentation of MRI,” *Frontiers in Neuroinformatics*, vol. 11, nov 2017.
- [52] D. Castillo-Barnes, C. Arenas, F. Segovia, F. J. Martínez-Murcia, I. A. Illán, J. M. Górriz, J. Ramírez, and D. Salas-Gonzalez, “On a heavy-tailed intensity normalization of the parkinson’s progression markers initiative brain database,” in *Natural and Artificial Computation for Biomedicine and Neuroscience: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2017, Corunna, Spain, June 19-23, 2017, Proceedings, Part I*, pp. 298–304, Springer International Publishing, 2017.
- [53] D. Castillo-Barnes, F. Segovia, F. J. Martínez-Murcia, D. Salas-Gonzalez, J. Ramírez, and J. M. Górriz, “Classification Improvement for Parkinson’s Disease Diagnosis Using the Gradient Magnitude in DaTSCAN SPECT Images,” in *Advances in Intelligent Systems and Computing*, vol. 771, pp. 100–109, Springer International Publishing, jun 2018.
- [54] D. Castillo-Barnes, J. Ramírez, F. Segovia, F. J. Martínez-Murcia, D. Salas-Gonzalez, and J. M. Górriz, “Robust Ensemble Classification Methodology for I123-Ioflupane SPECT Images and Multiple Heterogeneous Biomarkers in the Diagnosis of Parkinson’s Disease,” *Frontiers in Neuroinformatics*, vol. 12, p. 53, aug 2018.
- [55] D. Castillo-Barnes, D. Salas-Gonzalez, J. Ramirez, F. J. Martínez-Murcia, F. Segovia, and J. M. Gorriz, “Analysis of i[123]-ioflupane

- SPECT intensity iso-surfaces to assist the diagnosis of parkinsonism,” in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, IEEE, nov 2018.
- [56] D. Castillo-Barnes, J. Ramirez, F. J. Martinez-Murcia, F. Segovia, D. Salas-Gonzalez, and J. M. Gorriz, “Ensemble classification of heterogeneous biomarkers in the diagnosis of parkinsonism,” in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, IEEE, nov 2018.
- [57] D. Castillo-Barnes, F. J. Martinez-Murcia, F. Segovia, I. A. Illán, D. Salas-Gonzalez, J. M. Górriz, and J. Ramírez, “Comparison between affine and non-affine transformations applied to i[123]-fp-cit spect images used for parkinson’s disease diagnosis,” in *Understanding the Brain Function and Emotions* (J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo, and H. Adeli, eds.), pp. 379–388, Springer International Publishing, 2019.
- [58] D. Castillo-Barnes, L. Su, J. Ramírez, D. Salas-Gonzalez, F. J. Martinez-Murcia, I. A. Illan, F. Segovia, A. Ortiz, C. Cruchaga, M. R. Farlow, C. Xiong, N. R. Graff-Radford, P. R. Schofield, C. L. Masters, S. Salloway, M. Jucker, H. Mori, J. Levin, J. M. Gorriz, and D. I. A. N. (DIAN), “Autosomal dominantly inherited alzheimer disease: Analysis of genetic subgroups by machine learning,” *Information Fusion*, vol. 58, pp. 153–167, jan 2020.
- [59] D. Castillo-Barnes, F. J. Martinez-Murcia, A. Ortiz, D. Salas-Gonzalez, J. Ramirez, and J. M. Gorriz, “Morphological characterization of functional brain imaging by isosurface analysis in parkinson’s disease,” *International Journal of Neural Systems*, may 2020.
- [60] D. Castillo-Barnes, F. Martinez-Murcia, J. Ramírez, J. Górriz, and D. Salas-Gonzalez, “Expectation-maximization algorithm for finite mixture of α -stable distributions,” *Neurocomputing*, vol. 413, pp. 210–216, jul 2020.
- [61] A. Yendiki, A. Stevens, J. Augustinack, D. Salat, L. Zollei, and B. Fischl, “Anatomical priors for global probabilistic diffusion tractography,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, IEEE, jun 2009.
- [62] J. Wasserthal, P. F. Neher, D. Hirjak, and K. H. Maier-Hein, “Combined tract segmentation and orientation mapping for bundle-specific tractography,” *Medical Image Analysis*, vol. 58, p. 101559, dec 2019.
- [63] I. Lipp, G. D. Parker, E. C. Tallantyre, A. Goodall, S. Grama, E. Patitucci, P. Heveron, V. Tomassini, and D. K. Jones, “Tractography

- in the presence of multiple sclerosis lesions,” *NeuroImage*, vol. 209, p. 116471, apr 2020.
- [64] B. Marincek, D. Weishaupt, and V. D. Köchli, *How Does MRI Work?* Springer-Verlag GmbH, feb 2008.
- [65] D. o. E. National Research Council and C. o. P. S. M. Physical Sci, *Mathematics and Physics of Emerging Biomedical Imaging*. NATL ACADEMY PR, mar 1996.
- [66] J. Radon, “On the determination of functions from their integral values along certain manifolds,” *IEEE Transactions on Medical Imaging*, vol. 5, pp. 170–176, dec 1986.
- [67] A. Shukla and U. Kumar, “Positron emission tomography: An overview,” *Journal of Medical Physics*, vol. 31, no. 1, p. 13, 2006.
- [68] O. Maier, C. Schröder, N. D. Forkert, T. Martinetz, and H. Handels, “Classifiers for Ischemic Stroke Lesion Segmentation: A Comparison Study,” *PLOS ONE*, vol. 10, dec 2015.
- [69] C. Maggia, S. Doyle, F. Forbes, O. Heck, I. Troprès, C. Berthet, Y. Teyssier, L. Velly, J.-F. Payen, and M. Dojat, “Assessment of tissue injury in severe brain trauma,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 57–68, Springer International Publishing, 2016.
- [70] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. C. Mok, L. Shi, and P.-A. Heng, “Automatic detection of cerebral microbleeds from MR images via 3d convolutional neural networks,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1182–1195, may 2016.
- [71] L. Storelli, M. A. Rocca, E. Pagani, W. V. Hecke, M. A. Horsfield, N. D. Stefano, A. Rovira, J. Sastre-Garriga, J. Palace, D. Sima, D. Smeets, and M. F. and, “Measurement of whole-brain and gray matter atrophy in multiple sclerosis: Assessment with MR imaging,” *Radiology*, vol. 288, pp. 554–564, aug 2018.
- [72] V. S. Yedavalli, E. Tong, D. Martin, K. W. Yeom, and N. D. Forkert, “Artificial intelligence in stroke imaging: Current and future perspectives,” *Clinical Imaging*, vol. 69, pp. 246–254, jan 2021.
- [73] H. D. Rosas, D. H. Salat, S. Y. Lee, A. K. Zaleta, V. Pappu, B. Fischl, D. Greve, N. Hevelone, and S. M. Hersch, “Cerebral cortex and the clinical expression of huntington’s disease: complexity and heterogeneity,” *Brain*, vol. 131, pp. 1057–1068, feb 2008.

- [74] C. A. Ross, E. H. Aylward, E. J. Wild, D. R. Langbehn, J. D. Long, J. H. Warner, R. I. Scahill, B. R. Leavitt, J. C. Stout, J. S. Paulsen, R. Reilmann, P. G. Unschuld, A. Wexler, R. L. Margolis, and S. J. Tabrizi, “Huntington disease: natural history, biomarkers and prospects for therapeutics,” *Nature Reviews Neurology*, vol. 10, pp. 204–216, mar 2014.
- [75] M. Iv, M. Zhou, K. Shpanskaya, S. Perreault, Z. Wang, E. Tranvinh, B. Lanzman, S. Vajapeyam, N. Vitanza, P. Fisher, Y. Cho, S. Laughlin, V. Ramaswamy, M. Taylor, S. Cheshier, G. Grant, T. Y. Poussaint, O. Gevaert, and K. Yeom, “MR imaging-based radiomic signatures of distinct molecular subgroups of medulloblastoma,” *American Journal of Neuroradiology*, vol. 40, pp. 154–161, dec 2018.
- [76] S. Radelj, T. Zarković-Palijan, D. Kovacević, D. Marinović, and E. D. Hero, “Neuroimaging techniques in modern forensic psychiatry,” *Collegium antropologicum*, vol. 34 Suppl 2, pp. 287–290, apr 2010.
- [77] R. A. Dierckx, A. Otte, E. F. J. de Vries, A. van Waarde, and J. A. den Boer, eds., *PET and SPECT in Psychiatry*. Springer Berlin Heidelberg, 2014.
- [78] S. Aram, D. Kornev, R. Sadeghian, S. E. Sardari, S. K. Venu, H. Dasttestani, and A. Gandjbakhche, “Machine learning approaches and neuroimaging in cognitive functions of the human brain: A review,” in *Advances in Neuroergonomics and Cognitive Engineering*, pp. 23–29, Springer International Publishing, jun 2020.
- [79] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones, “Alzheimer’s disease,” *The Lancet*, vol. 377, pp. 1019–1031, mar 2011.
- [80] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub, and C. H. Phelps, “The diagnosis of dementia due to alzheimer’s disease: recommendations from the national institute on aging-alzheimer’s association workgroups on diagnostic guidelines for alzheimer’s disease,” *Alzheimer’s & dementia : the journal of the Alzheimer’s Association*, vol. 7, pp. 263–269, may 2011.
- [81] H. J. Markowitsch and A. Staniloiu, “Amnesic disorders,” *Lancet (London, England)*, vol. 380, pp. 1429–1440, oct 2012.
- [82] R. N. Martins, V. Villemagne, H. R. Sohrabi, P. Chatterjee, T. M. Shah, G. Verdile, P. Fraser, K. Taddei, V. B. Gupta, S. R. Rainey-Smith, E. Hone, S. Pedrini, W. L. Lim, I. Martins, S. Frost, S. Gupta,

- S. O'Bryant, A. Rembach, D. Ames, K. Ellis, S. J. Fuller, B. Brown, S. L. Gardener, B. Fernando, P. Bharadwaj, S. Burnham, S. M. Laws, A. M. Barron, K. Goozee, E. J. Wahjoepramono, P. R. Asih, J. D. Doecke, O. Salvado, A. I. Bush, C. C. Rowe, S. E. Gandy, and C. L. Masters, "Alzheimer's disease: a journey from amyloid peptides and oxidative stress, to biomarker technologies and disease prevention strategies-gains from aibl and dian cohort studies," *Journal of Alzheimer's Disease*, vol. 62, no. 3, pp. 965–992, 2018.
- [83] D. S. Knopman, S. T. DeKosky, J. L. Cummings, H. Chui, J. Corey-Bloom, N. Relkin, G. W. Small, B. Miller, and J. C. Stevens, "Practice parameter: diagnosis of dementia (an evidence-based review). report of the quality standards subcommittee of the american academy of neurology.," *Neurology*, vol. 56, pp. 1143–1153, may 2001.
- [84] L. A. van de Pol, A. Hensel, F. Barkhof, H. J. Gertz, P. Scheltens, and W. M. van der Flier, "Hippocampal atrophy in alzheimer disease: age matters.," *Neurology*, vol. 66, pp. 236–238, jan 2006.
- [85] K. Kantarci, R. Avula, M. L. Senjem, A. R. Samikoglu, B. Zhang, S. D. Weigand, S. A. Przybelski, H. A. Edmonson, P. Vemuri, D. S. Knopman, T. J. Ferman, B. F. Boeve, R. C. Petersen, and C. R. Jack, "Dementia with lewy bodies and alzheimer disease: neurodegenerative patterns characterized by dti.," *Neurology*, vol. 74, pp. 1814–1821, jun 2010.
- [86] J. L. Whitwell, D. W. Dickson, M. E. Murray, S. D. Weigand, N. Tosakulwong, M. L. Senjem, D. S. Knopman, B. F. Boeve, J. E. Parisi, R. C. Petersen, C. R. Jack, and K. A. Josephs, "Neuroimaging correlates of pathologically defined subtypes of alzheimer's disease: a case-control study.," *The Lancet. Neurology*, vol. 11, pp. 868–877, oct 2012.
- [87] Y. Shigemoto, D. Sone, E. Imabayashi, N. Maikusa, N. Okamura, S. Furumoto, Y. Kudo, M. Ogawa, H. Takano, Y. Yokoi, M. Sakata, T. Tsukamoto, K. Kato, N. Sato, and H. Matsuda, "Dissociation of tau deposits and brain atrophy in early alzheimer's disease: A combined positron emission tomography/magnetic resonance imaging study," *Frontiers in Aging Neuroscience*, vol. 10, jul 2018.
- [88] L. Mosconi, "Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 32, no. 4, pp. 486–510, 2005.
- [89] G. D. Rabinovici, H. J. Rosen, A. Alkalay, J. Kornak, A. J. Furst, N. Agarwal, E. C. Mormino, J. P. O'Neil, M. Janabi, A. Karydas, M. E.

- Growdon, J. Y. Jang, E. J. Huang, S. J. Dearmond, J. Q. Trojanowski, L. T. Grinberg, M. L. Gorno-Tempini, W. W. Seeley, B. L. Miller, and W. J. Jagust, “Amyloid vs fdg-pet in the differential diagnosis of ad and ftd.,” *Neurology*, vol. 77, pp. 2034–2042, dec 2011.
- [90] N. Smailagic, M. Vacante, C. Hyde, S. Martin, O. Ukoumunne, and C. Sachpekidis, “18f-fdg pet for the early diagnosis of alzheimer’s disease dementia and other dementias in people with mild cognitive impairment (mci),” *Cochrane Database of Systematic Reviews*, jan 2015.
- [91] V. J. Lowe, B. J. Kemp, C. R. Jack, M. Senjem, S. Weigand, M. Shiung, G. Smith, D. Knopman, B. Boeve, B. Mullan, and R. C. Petersen, “Comparison of 18f-FDG and PiB PET in cognitive impairment,” *Journal of Nuclear Medicine*, vol. 50, pp. 878–886, may 2009.
- [92] D. F. Wong, P. B. Rosenberg, Y. Zhou, A. Kumar, V. Raymont, H. T. Ravert, R. F. Dannals, A. Nandi, J. R. Brasic, W. Ye, J. Hilton, C. Lyketsos, H. F. Kung, A. D. Joshi, D. M. Skovronsky, and M. J. Pontecorvo, “In vivo imaging of amyloid deposition in alzheimer disease using the radioligand 18f-AV-45 (flobetapir f 18),” *Journal of Nuclear Medicine*, vol. 51, pp. 913–920, may 2010.
- [93] A. S. Fleisher, K. Chen, X. Liu, A. Roontiva, P. Thiyyagura, N. Ayutyanont, A. D. Joshi, C. M. Clark, M. A. Mintun, M. J. Pontecorvo, P. M. Doraiswamy, K. A. Johnson, D. M. Skovronsky, and E. M. Reiman, “Using positron emission tomography and florbetapir f18 to image cortical amyloid in patients with mild cognitive impairment or dementia due to alzheimer disease.,” *Archives of neurology*, vol. 68, pp. 1404–1411, nov 2011.
- [94] L. Yang, D. Rieves, and C. Ganley, “Brain amyloid imaging—fda approval of florbetapir f18 injection.,” *The New England journal of medicine*, vol. 367, pp. 885–887, sep 2012.
- [95] K. A. Johnson, S. Minoshima, N. I. Bohnen, K. J. Donohoe, N. L. Foster, P. Herscovitch, J. H. Karlawish, C. C. Rowe, M. C. Carrillo, D. M. Hartley, S. Hedrick, V. Pappas, and W. H. Thies, “Appropriate use criteria for amyloid pet: a report of the amyloid imaging task force, the society of nuclear medicine and molecular imaging, and the alzheimer’s association.,” *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 54, pp. 476–490, mar 2013.
- [96] V. L. Villemagne, N. Okamura, and C. C. Rowe, “Untangling tau imaging,” *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 4, pp. 39–42, jan 2016.

- [97] C. Xia, S. J. Makaretz, C. Caso, S. McGinnis, S. N. Gomperts, J. Sepulcre, T. Gomez-Isla, B. T. Hyman, A. Schultz, N. Vasdev, K. A. Johnson, and B. C. Dickerson, “Association of in vivo [18f]AV-1451 tau PET imaging results with cortical atrophy and symptoms in typical and atypical alzheimer disease,” *JAMA Neurology*, vol. 74, p. 427, apr 2017.
- [98] E. Levy, M. Carman, I. Fernandez-Madrid, M. Power, I. Lieberburg, S. van Duinen, G. Bots, W. Luyendijk, and B. Frangione, “Mutation of the alzheimer’s disease amyloid gene in hereditary cerebral hemorrhage, dutch type,” *Science*, vol. 248, no. 4959, pp. 1124–1126, 1990.
- [99] R. Sherrington, E. I. Rogaev, Y. Liang, E. A. Rogaeva, G. Levesque, M. Ikeda, H. Chi, C. Lin, G. Li, K. Holman, T. Tsuda, L. Mar, J. F. Foncin, A. C. Bruni, M. P. Montesi, S. Sorbi, I. Rainero, L. Pinessi, L. Nee, I. Chumakov, D. Pollen, A. Brookes, P. Sanseau, R. J. Polinsky, W. Wasco, H. A. R. Da Silva, J. L. Haines, M. A. Pericak-Vance, R. E. Tanzi, A. D. Roses, P. E. Fraser, J. M. Rommens, and P. H. St George-Hyslop, “Cloning of a gene bearing missense mutations in early-onset familial alzheimer’s disease,” *Nature*, vol. 375, p. 754, jun 1995.
- [100] E. Levy-Lahad, W. Wasco, P. Poorkaj, D. Romano, J. Oshima, W. Pettingell, C. Yu, P. Jondro, S. Schmidt, K. Wang, and e. al., “Candidate gene for the chromosome 1 familial alzheimer’s disease locus,” *Science*, vol. 269, no. 5226, pp. 973–977, 1995.
- [101] J. M. Ringman, L.-J. Liang, Y. Zhou, S. Vangala, E. Teng, S. Kremen, D. Wharton, A. Goate, D. S. Marcus, M. Farlow, B. Ghetti, E. McDade, C. L. Masters, R. P. Mayeux, M. Rossor, S. Salloway, P. R. Schofield, J. L. Cummings, V. Buckles, R. Bateman, J. C. Morris, and the Dominantly Inherited Alzheimer Network, “Early behavioural changes in familial alzheimer’s disease in the dominantly inherited alzheimer network,” *Brain*, vol. 138, no. 4, pp. 1036–1045, 2015.
- [102] N. P. Oxtoby, A. L. Young, D. M. Cash, T. L. S. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, and D. C. Alexander, “Data-driven models of dominantly-inherited alzheimer’s disease progression,” *Brain*, vol. 141, no. 5, pp. 1529–1544, 2018.
- [103] R. J. Bateman, C. Xiong, T. L. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, D. M. Holtzman, A. Santacruz, V. Buckles, A. Oliver, K. Moulder, P. S. Aisen, B. Ghetti, W. E. Klunk, E. McDade, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, P. R. Schofield, R. A. Sperling, S. Salloway, and J. C. Morris, “Clinical and biomarker

- changes in dominantly inherited alzheimer's disease," *New England Journal of Medicine*, vol. 367, no. 9, pp. 795–804, 2012.
- [104] E. M. Reiman, Y. T. Quiroz, A. S. Fleisher, K. Chen, C. Velez-Pardo, M. Jimenez-Del-Rio, A. M. Fagan, A. R. Shah, S. Alvarez, A. Arbelaez, M. Giraldo, N. Acosta-Baena, R. A. Sperling, B. Dickerson, C. E. Stern, V. Tirado, C. Munoz, R. A. Reiman, M. J. Huentelman, G. E. Alexander, J. B. Langbaum, K. S. Kosik, P. N. Tariot, and F. Lopera, "Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant alzheimer's disease in the presenilin 1 e280a kindred: a case-control study," *The Lancet Neurology*, vol. 11, no. 12, pp. 1048–1056, 2012.
- [105] T. JB, B. MR, B. RJ, and et al, "Functional connectivity in autosomal dominant and late-onset alzheimer disease," *JAMA Neurology*, vol. 71, no. 9, pp. 1111–1122, 2014.
- [106] M. Tang, D. C. Ryman, E. McDade, M. S. Jasielc, V. D. Buckles, N. J. Cairns, A. M. Fagan, A. Goate, D. S. Marcus, C. Xiong, R. F. Allegri, J. P. Chhatwal, A. Danek, M. R. Farlow, N. C. Fox, B. Ghetti, N. R. Graff-Radford, C. Laske, R. N. Martins, C. L. Masters, R. P. Mayeux, J. M. Ringman, M. N. Rossor, S. P. Salloway, P. R. Schofield, J. C. Morris, and R. J. Bateman, "Neurological manifestations of autosomal dominant familial alzheimer's disease: a comparison of the published literature with the dominantly inherited alzheimer network observational study (dian-obs)," *The Lancet Neurology*, vol. 15, no. 13, pp. 1317–1325, 2016.
- [107] D. C. Ryman, N. Acosta-Baena, P. S. Aisen, T. Bird, A. Danek, N. C. Fox, A. Goate, P. Frommelt, B. Ghetti, J. B. S. Langbaum, F. Lopera, R. Martins, C. L. Masters, R. P. Mayeux, E. McDade, S. Moreno, E. M. Reiman, J. M. Ringman, S. Salloway, P. R. Schofield, R. Sperling, P. N. Tariot, C. Xiong, J. C. Morris, R. J. Bateman, *et al.*, "Symptom onset in autosomal dominant alzheimer disease: A systematic review and meta-analysis," *Neurology*, vol. 83, pp. 253–260, jun 2014.
- [108] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack, J. Kaye, T. J. Montine, D. C. Park, E. M. Reiman, C. C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M. C. Carrillo, B. Thies, M. Morrison-Bogorad, M. V. Wagster, and C. H. Phelps, "Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease," *Alzheimer's & Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

- [109] S. Mills, J. Mallmann, A. Santacruz, A. Fuqua, M. Carril, P. Aisen, M. Althage, S. Belyew, T. Benzinger, W. Brooks, V. Buckles, N. Cairns, D. Clifford, A. Danek, A. Fagan, M. Farlow, N. Fox, B. Ghetti, A. Goate, D. Heinrichs, R. Hornbeck, C. Jack, M. Jucker, W. Klunk, D. Marcus, R. Martins, C. Masters, R. Mayeux, E. McDade, J. Morris, A. Oliver, J. Ringman, M. Rossor, S. Salloway, P. Schofield, J. Snider, P. Snyder, R. Sperling, C. Stewart, R. Thomas, C. Xiong, and R. Bateman, “Preclinical trials in autosomal dominant ad: Implementation of the dian-tu trial,” *Revue Neurologique*, vol. 169, pp. 737–743, oct 2013.
- [110] Y. Su, T. M. Blazey, C. J. Owen, J. J. Christensen, K. Friedrichsen, N. Joseph-Mathurin, Q. Wang, R. C. Hornbeck, B. M. Ances, A. Z. Snyder, L. A. Cash, R. A. Koeppe, W. E. Klunk, D. Galasko, A. M. Brickman, E. McDade, J. M. Ringman, P. M. Thompson, A. J. Saykin, B. Ghetti, R. A. Sperling, K. A. Johnson, S. P. Salloway, P. R. Schofield, C. L. Masters, V. L. Villemagne, N. C. Fox, S. Förster, K. Chen, E. M. Reiman, C. Xiong, D. S. Marcus, M. W. Weiner, J. C. Morris, R. J. Bateman, T. L. S. Benzinger, *et al.*, “Quantitative amyloid imaging in autosomal dominant alzheimer’s disease: Results from the DIAN study group,” *PLOS ONE*, vol. 11, pp. 1–14, mar 2016.
- [111] R. J. Bateman, T. L. Benzinger, S. Berry, D. B. Clifford, C. Duggan, A. M. Fagan, K. Fanning, M. R. Farlow, J. Hassenstab, E. M. McDade, S. Mills, K. Paumier, M. Quintana, S. P. Salloway, A. Santacruz, L. S. Schneider, G. Wang, C. Xiong, *et al.*, “The dian-tu next generation alzheimer’s prevention trial: Adaptive design and disease progression model,” *Alzheimer’s & Dementia*, vol. 13, pp. 8–19, aug 2016.
- [112] J. H. Kordower, C. W. Olanow, H. B. Dodiya, Y. Chu, T. G. Beach, C. H. Adler, G. M. Halliday, and R. T. Bartus, “Disease duration and the integrity of the nigrostriatal system in parkinson’s disease,” *Brain*, vol. 136, pp. 2419–2431, jul 2013.
- [113] S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr, and H. Adeli, “Parkinson’s disease: Cause factors, measurable indicators, and early diagnosis,” *Computers in Biology and Medicine*, vol. 102, pp. 234–241, nov 2018.
- [114] B. Cholerton, K. L. Poston, L. Tian, J. F. Quinn, K. A. Chung, A. L. Hiller, S.-C. Hu, K. Specketer, T. J. Montine, K. L. Edwards, and C. P. Zabetian, “Participant and study partner reported impact of cognition on functional activities in parkinson’s disease,” *Movement Disorders Clinical Practice*, vol. 7, pp. 61–69, dec 2019.

- [115] J. Jankovic, “Parkinson’s disease: clinical features and diagnosis,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 79, pp. 368–376, apr 2008.
- [116] J. L. Neumeyer, S. Wang, R. A. Milius, R. M. Baldwin, Y. Zea-Ponce, P. B. Hoffer, E. Sybirska, M. Al-Tikriti, and D. S. Charney, “[123i]-2 β -carbomethoxy-3 β -(4-iodophenyl)tropane: high-affinity SPECT (single photon emission computed tomography) radiotracer of monoamine reuptake sites in brain,” *Journal of Medicinal Chemistry*, vol. 34, pp. 3144–3146, oct 1991.
- [117] I. Navalpotro-Gomez, R. Dacosta-Aguayo, F. Molinet-Dronda, A. Martin-Bastida, A. Botas-Peñin, H. Jimenez-Urbietta, M. Delgado-Alvarado, B. Gago, A. Quiroga-Varela, and M. C. Rodriguez-Oroz, “Nigrostriatal dopamine transporter availability, and its metabolic and clinical correlates in parkinson’s disease patients with impulse control disorders,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, pp. 2065–2076, jul 2019.
- [118] J. Blesa, “Inter-hemispheric asymmetry of nigrostriatal dopaminergic lesion: a possible compensatory mechanism in parkinson’s disease,” *Frontiers in Systems Neuroscience*, vol. 5, 2011.
- [119] J. Booij, J. B. Habraken, P. Bergmans, G. Tissingh, A. Winogrodzka, E. C. Wolters, A. G. Janssen, J. C. Stoof, and E. A. van Royen, “Imaging of dopamine transporters with iodine-123-fp-cit spect in healthy controls and patients with parkinson’s disease,” *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 39, pp. 1879–1884, nov 1998.
- [120] K. Marek, R. Innis, C. van Dyck, B. Fussell, M. Early, S. Eberly, D. Oakes, and J. Seibyl, “[123i]-CIT SPECT imaging assessment of the rate of parkinson’s disease progression,” *Neurology*, vol. 57, pp. 2089–2094, dec 2001.
- [121] G. Rizzo, M. Copetti, S. Arcuti, D. Martino, A. Fontana, and G. Logroscino, “Accuracy of clinical diagnosis of parkinson disease,” *Neurology*, vol. 86, pp. 566–576, jan 2016.
- [122] E. Hustad, A. H. Skogholt, K. Hveem, and J. O. Aasly, “The accuracy of the clinical diagnosis of parkinson disease. the HUNT study,” *Journal of Neurology*, vol. 265, pp. 2120–2124, jul 2018.
- [123] T. Rus, P. Tomše, L. Jensterle, M. Grmek, Z. Pirtošek, D. Eidelberg, C. Tang, and M. Trošt, “Differential diagnosis of parkinsonian syndromes: a comparison of clinical and automated - metabolic brain pat-

- terns' based approach," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, pp. 2901–2910, apr 2020.
- [124] E. U. Haq, J. Huang, L. Kang, H. U. Haq, and T. Zhan, "Image-based state-of-the-art techniques for the identification and classification of brain diseases: a review," *Medical & Biological Engineering & Computing*, vol. 58, pp. 2603–2620, sep 2020.
- [125] A. Vitale, , R. Villa, L. Ugga, V. Romeo, A. Stanzione, and R. C. and, "Artificial intelligence applied to neuroimaging data in parkinsonian syndromes: Actuality and expectations," *Mathematical Biosciences and Engineering*, vol. 18, no. 2, pp. 1753–1773, 2021.
- [126] S. Cascianelli, M. Scialpi, S. Amici, N. Forini, M. Minestrini, M. Fravolini, H. Sinzinger, O. Schillaci, and B. Palumbo, "Role of artificial intelligence techniques (automatic classifiers) in molecular imaging modalities in neurodegenerative diseases," *Current Alzheimer Research*, vol. 14, pp. 198–207, jan 2017.
- [127] M. Kamber, R. Shinghal, D. Collins, G. Francis, and A. Evans, "Model-based 3-d segmentation of multiple sclerosis lesions in magnetic resonance brain images," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 442–453, 1995.
- [128] S. Valverde, A. Oliver, Y. Díez, M. Cabezas, J. Vilanova, L. Ramió-Torrentà, À. Rovira, and X. Lladó, "Evaluating the effects of white matter multiple sclerosis lesions on the volume estimation of 6 brain tissue segmentation methods," *American Journal of Neuroradiology*, vol. 36, pp. 1109–1115, feb 2015.
- [129] S. González-Vilà, A. Oliver, Y. Huo, X. Lladó, and B. A. Landman, "Brain structure segmentation in the presence of multiple sclerosis lesions," *NeuroImage: Clinical*, vol. 22, p. 101709, 2019.
- [130] R. E. Gabr, I. Coronado, M. Robinson, S. J. Sujit, S. Datta, X. Sun, W. J. Allen, F. D. Lublin, J. S. Wolinsky, and P. A. Narayana, "Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study," *Multiple Sclerosis Journal*, p. 135245851985684, jun 2019.
- [131] K. R. Laukamp, F. Thiele, G. Shakirin, D. Zopfs, A. Faymonville, M. Timmer, D. Maintz, M. Perkuhn, and J. Borggrefe, "Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI," *European Radiology*, vol. 29, pp. 124–132, jun 2018.

- [132] T. Kurc, S. Bakas, X. Ren, A. Bagari, A. Momeni, Y. Huang, L. Zhang, A. Kumar, M. Thibault, Q. Qi, Q. Wang, A. Kori, O. Gevaert, Y. Zhang, D. Shen, M. Khened, X. Ding, G. Krishnamurthi, J. Kalpathy-Cramer, J. Davis, T. Zhao, R. Gupta, J. Saltz, and K. Farahani, “Segmentation and classification in digital pathology for glioma research: Challenges and deep learning approaches,” *Frontiers in Neuroscience*, vol. 14, feb 2020.
- [133] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu, “Statistical normalization techniques for magnetic resonance imaging,” *NeuroImage: Clinical*, vol. 6, pp. 9–19, 2014.
- [134] A. Brahim, J. Ramírez, J. M. Górriz, L. Khedher, and D. Salas-Gonzalez, “Comparison between different intensity normalization methods in 123i-ioflupane imaging for the automatic detection of parkinsonism,” *PLOS ONE*, pp. 1–20, jun 2015.
- [135] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, M. López, F. Segovia, R. Chaves, P. Padilla, and C. G. P. and, “Feature selection using factor analysis for alzheimer’s diagnosis using f18-FDG PET images,” *Medical Physics*, vol. 37, pp. 6084–6095, nov 2010.
- [136] K. Fukuma, V. S. Prasath, H. Kawanaka, B. J. Aronow, and H. Takase, “A study on feature extraction and disease stage classification for glioma pathology images,” in *2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, jul 2016.
- [137] J. Ramirez, J. M. Gorriz, R. Chaves, M. Lopez, D. Salas-Gonzalez, I. Alvarez, and F. Segovia, “Spect image classification using random forests,” *Electronics Letters*, vol. 45, no. 12, pp. 604–605, 2009.
- [138] I. A. Illán, J. M. Górriz, J. Ramírez, F. Segovia, J. M. Jiménez-Hoyuela, and S. J. O. Lozano, “Automatic assistance to parkinson’s disease diagnosis in DaTSCAN SPECT imaging,” *Medical Physics*, vol. 39, pp. 5971–5980, sep 2012.
- [139] D. Mudali, L. K. Teune, R. J. Renken, K. L. Leenders, and J. B. T. M. Roerdink, “Classification of parkinsonian syndromes from FDG-PET brain data using decision trees with SSM/PCA features,” *Computational and Mathematical Methods in Medicine*, vol. 2015, pp. 1–10, 2015.
- [140] L.-Y. Ma, Y. Tian, C.-R. Pan, Z.-L. Chen, Y. Ling, K. Ren, J.-S. Li, and T. Feng, “Motor progression in early-stage parkinson’s disease: A clinical prediction model and the role of cerebrospinal fluid biomarkers,” *Frontiers in Aging Neuroscience*, vol. 12, jan 2021.

- [141] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection.,” in *IJCAI’95 Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2 of *IJCAI’95*, (Quebec, Canada), pp. 1137–1145, aug 1995.
- [142] A. P. Bradley, “The use of the area under the roc curve in the evaluation of machine learning algorithms,” *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [143] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer-Verlag GmbH, 2009.
- [144] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, aug 1987.
- [145] F. Song, Z. Guo, and D. Mei, “Feature selection using principal component analysis,” in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, IEEE, 2010.
- [146] S. Wold, A. Ruhe, H. Wold, and I. W. J. Dunn, “The collinearity problem in linear regression. the partial least squares (PLS) approach to generalized inverses,” *SIAM Journal on Scientific and Statistical Computing*, vol. 5, pp. 735–743, sep 1984.
- [147] G. Forman *et al.*, “An extensive empirical study of feature selection metrics for text classification.,” *J. Mach. Learn. Res.*, vol. 3, no. Mar, pp. 1289–1305, 2003.
- [148] B. Welch, “The generalization of student’s problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.
- [149] H. Mann and D. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.
- [150] W. Cochran and G. Cox, *Experimental designs (2nd ed.)*. Oxford, England: Wiley, 2 ed., 1950.
- [151] R. G. Shaw and T. Mitchell-Olds, “Anova for unbalanced data: An overview,” *Ecology*, vol. 74, pp. 1638–1645, sep 1993.
- [152] M. Kutner, *Applied linear statistical models*. Boston: McGraw-Hill Irwin, 2005.

- [153] D. Scheinost, S. Noble, C. Horien, A. S. Greene, E. M. Lake, M. Salehi, S. Gao, X. Shen, D. O'Connor, D. S. Barron, S. W. Yip, M. D. Rosenberg, and R. T. Constable, "Ten simple rules for predictive modeling of individual differences in neuroimaging," *NeuroImage*, vol. 193, pp. 35–45, jun 2019.
- [154] T. B. Smith and K. S. Nayak, "MRI artifacts and correction strategies," *Imaging in Medicine*, vol. 2, pp. 445–457, aug 2010.
- [155] M. Zaitsev, J. Maclaren, and M. Herbst, "Motion artifacts in MRI: A complex problem with many partial solutions," *Journal of Magnetic Resonance Imaging*, vol. 42, pp. 887–901, jan 2015.
- [156] H. Gudbjartsson and S. Patz, "The rician distribution of noisy mri data.," *Magnetic resonance in medicine*, vol. 34, pp. 910–914, dec 1995.
- [157] S. V.R., D. R. Edla, J. Joseph, and V. Kuppili, "Analysis of controversies in the formulation and evaluation of restoration algorithms for MR images," *Expert Systems with Applications*, vol. 135, pp. 39–59, nov 2019.
- [158] J. Yang, J. Fan, D. Ai, S. Zhou, S. Tang, and Y. Wang, "Brain MR image denoising for rician noise using pre-smooth non-local means filter," *BioMedical Engineering OnLine*, vol. 14, no. 1, p. 2, 2015.
- [159] S. A. Akar, "Determination of optimal parameters for bilateral filter in brain MR image denoising," *Applied Soft Computing*, vol. 43, pp. 87–96, jun 2016.
- [160] J.-M. Bonny, J.-P. Renou, and M. Zanca, "Optimal measurement of magnitude and phase from MR data," *Journal of Magnetic Resonance, Series B*, vol. 113, pp. 136–144, nov 1996.
- [161] J. Rajan, A. J. den Dekker, and J. Sijbers, "A new non-local maximum likelihood estimation method for rician noise reduction in magnetic resonance images using the kolmogorov-smirnov test," *Signal Processing*, vol. 103, pp. 16–23, oct 2014.
- [162] L. Tan and L. Shi, "Multiwavelet-based estimation for improving magnetic resonance images," in *2009 2nd International Congress on Image and Signal Processing*, IEEE, oct 2009.
- [163] P. Coupé, J. V. Manjón, E. Gedamu, D. Arnold, M. Robles, and D. L. Collins, "Robust rician noise estimation for MR images," *Medical Image Analysis*, vol. 14, pp. 483–493, aug 2010.
- [164] H. M. Golshan and R. P. Hasanzadeh, "An optimized LMMSE based method for 3d MRI denoising," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 12, pp. 861–870, jul 2015.

- [165] X. You, N. Cao, H. Lu, M. Mao, and W. Wanga, “Denoising of MR images with rician noise using a wider neural network and noise range division,” *Magnetic Resonance Imaging*, vol. 64, pp. 154–159, dec 2019.
- [166] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, “Image denoising using scale mixtures of gaussians in the wavelet domain,” *IEEE Transactions on Image Processing*, vol. 12, pp. 1338–1351, nov 2003.
- [167] T. Küstner, K. Armanious, J. Yang, B. Yang, F. Schick, and S. Gatidis, “Retrospective correction of motion-affected MR images using deep learning frameworks,” *Magnetic Resonance in Medicine*, vol. 82, pp. 1527–1540, may 2019.
- [168] M. Reuter, M. D. Tisdall, A. Qureshi, R. L. Buckner, A. J. van der Kouwe, and B. Fischl, “Head motion during MRI acquisition reduces gray matter volume and thickness estimates,” *NeuroImage*, vol. 107, pp. 107–115, feb 2015.
- [169] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, M. Moreno-Caballero, and M. G.-R. and, “Parametrization of textural patterns in 123i-ioflupane imaging for the automatic detection of parkinsonism,” *Medical Physics*, vol. 41, p. 012502, dec 2013.
- [170] F. Segovia, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, J. Levin, M. Schuberth, M. Brendel, A. Rominger, K. Bötzel, G. Garraux, and C. Phillips, “Multivariate analysis of 18f-DMFP PET data to assist the diagnosis of parkinsonism,” *Frontiers in Neuroinformatics*, vol. 11, p. 23, mar 2017.
- [171] K. Friston, *Statistical parametric mapping : the analysis of funtional brain images*. Amsterdam Boston: Elsevier/Academic Press, 2007.
- [172] P. Ripollés, J. Marco-Pallarés, R. de Diego-Balaguer, J. Miró, M. Falip, M. Juncadella, F. Rubio, and A. Rodriguez-Fornells, “Analysis of automated methods for spatial normalization of lesioned brains,” *NeuroImage*, vol. 60, pp. 1296–1306, apr 2012.
- [173] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, J. H. Song, M. Jenkinson, C. Lepage, D. Rueckert, P. Thompson, T. Vercauteren, R. P. Woods, J. J. Mann, and R. V. Parsey, “Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration,” *NeuroImage*, vol. 46, pp. 786–802, jul 2009.
- [174] J. Ashburner, “A fast diffeomorphic image registration algorithm,” *NeuroImage*, vol. 38, pp. 95–113, oct 2007.

- [175] B. Avants, C. Epstein, M. Grossman, and J. Gee, “Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain,” *Medical Image Analysis*, vol. 12, pp. 26–41, feb 2008.
- [176] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of ANTs similarity metric performance in brain image registration,” *NeuroImage*, vol. 54, pp. 2033–2044, feb 2011.
- [177] B. B. Avants, N. J. Tustison, M. Stauffer, G. Song, B. Wu, and J. C. Gee, “The insight ToolKit image registration framework,” *Frontiers in Neuroinformatics*, vol. 8, apr 2014.
- [178] J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, “Evaluating the impact of intensity normalization on MR image synthesis,” in *Medical Imaging 2019: Image Processing* (E. D. Angelini and B. A. Landman, eds.), SPIE, mar 2019.
- [179] P. Padilla, J. Górriz, J. Ramírez, D. Salas-Gonzalez, and I. Illán, “Intensity normalization in the analysis of functional DaTSCAN SPECT images: The α -stable distribution-based normalization method vs other approaches,” *Neurocomputing*, vol. 150, pp. 4–15, feb 2015.
- [180] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, and A. Ortiz, “Convolutional neural networks for neuroimaging in parkinson’s disease: Is preprocessing needed?,” *International Journal of Neural Systems*, vol. 28, p. 1850035, dec 2018.
- [181] M. G. M. Rahman, M. M. Islam, T. Tsujikawa, Y. Kiyono, and H. Okazawa, “Count-based method for specific binding ratio calculation in [i-123]FP-CIT SPECT analysis,” *Annals of Nuclear Medicine*, vol. 33, pp. 14–21, oct 2018.
- [182] C. Scherfler and M. Nocker, “Dopamine transporter SPECT: How to remove subjectivity?,” *Movement Disorders*, vol. 24, no. S2, pp. 721–724, 2009.
- [183] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.
- [184] S. M. Smith, “Fast robust automated brain extraction,” *Human Brain Mapping*, vol. 17, pp. 143–155, nov 2002.

- [185] G. Lin, U. Adiga, K. Olson, J. F. Guzowski, C. A. Barnes, and B. Roy-sam, “A hybrid 3d watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks,” *Cytometry*, vol. 56A, pp. 23–36, oct 2003.
- [186] F. Ségonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn, and B. Fischl, “A hybrid approach to the skull stripping problem in MRI,” *NeuroImage*, vol. 22, pp. 1060–1075, jul 2004.
- [187] K. Boesen, K. Rehm, K. Schaper, S. Stoltzner, R. Woods, E. Lüders, and D. Rottenberg, “Quantitative comparison of four brain extraction algorithms,” *NeuroImage*, vol. 22, pp. 1255–1261, jul 2004.
- [188] M. Jenkinson, M. Pechaud, S. Smith, *et al.*, “BET2: Mr-based estimation of brain, skull and scalp surfaces,” in *Eleventh annual meeting of the organization for human brain mapping*, vol. 17, p. 167, 2005.
- [189] M. Ganzetti, G. A. Taberna, and D. Mantini, “fMRI data processing in MRTOOL: to what extent does anatomical registration affect the reliability of functional results?,” *Brain Imaging and Behavior*, vol. 13, pp. 1538–1553, nov 2018.
- [190] J. E. Iglesias, C.-Y. Liu, P. M. Thompson, and Z. Tu, “Robust brain extraction across datasets and comparison with publicly available methods,” *IEEE Transactions on Medical Imaging*, vol. 30, pp. 1617–1634, sep 2011.
- [191] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, “Deep MRI brain extraction: A 3d convolutional neural network for skull stripping,” *NeuroImage*, vol. 129, pp. 460–469, apr 2016.
- [192] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, “Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging,” *IEEE Transactions on Medical Imaging*, vol. 36, pp. 2319–2330, nov 2017.
- [193] Z. Akkus, P. Kostandy, K. A. Philbrick, and B. J. Erickson, “Robust brain extraction tool for CT head images,” *Neurocomputing*, vol. 392, pp. 189–195, jun 2020.
- [194] F. Balsiger, C. Steindel, M. Arn, B. Wagner, L. Grunder, M. El-Koussy, W. Valenzuela, M. Reyes, and O. Scheidegger, “Segmentation of peripheral nerves from magnetic resonance neurography: A fully-automatic, deep learning-based approach,” *Frontiers in Neurology*, vol. 9, sep 2018.

- [195] D. Kennedy, P. Filipek, and V. Caviness, "Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 8, pp. 1–7, mar 1989.
- [196] H. Suzuki and J. ichiro Toriwaki, "Automatic segmentation of head mri images by knowledge guided thresholding," *Computerized Medical Imaging and Graphics*, vol. 15, pp. 233–240, jul 1991.
- [197] M. Brummer, R. Mersereau, R. Eisner, and R. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Transactions on Medical Imaging*, vol. 12, pp. 153–166, jun 1993.
- [198] S. Sandor and R. Leahy, "A 3d morphological algorithm for automated labelling of the cortex in magnetic resonance brain images," in *AAAI Spring Symposium Applications of Computer Vision in Medical Image Processing, Paulo Alto, CA*, 1994.
- [199] J. Fan, D. Yau, A. Elmagarmid, and W. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1454–1466, 2001.
- [200] M. Singh, P. Patel, D. Khosla, and T. Kim, "Segmentation of functional MRI by k-means clustering," *IEEE Transactions on Nuclear Science*, vol. 43, pp. 2030–2036, jun 1996.
- [201] D.-Q. Zhang and S.-C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation," *Artificial Intelligence in Medicine*, vol. 32, pp. 37–50, sep 2004.
- [202] K. V. Leemput, F. Maes, D. Vandermeulen, and P. Suetens, "Automated model-based tissue classification of MR images of the brain," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 897–908, 1999.
- [203] A. R. F. da Silva, "Bayesian mixture models of variable dimension for image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 94, pp. 1–14, apr 2009.
- [204] T. Kapur, W. L. Grimson, W. M. W. III, and R. Kikinis, "Segmentation of brain tissue from magnetic resonance images," *Medical Image Analysis*, vol. 1, pp. 109–127, jun 1996.
- [205] W. Wells, W. Grimson, R. Kikinis, and F. Jolesz, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging*, vol. 15, no. 4, pp. 429–442, 1996.

- [206] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, 2001.
- [207] J. Xie and H. Tsui, “Image segmentation based on maximum-likelihood estimation and optimum entropy-distribution (MLE-OED),” *Pattern Recognition Letters*, vol. 25, pp. 1133–1141, jul 2004.
- [208] Y.-F. Tsai, I.-J. Chiang, Y.-C. Lee, C.-C. Liao, and K.-L. Wang, “Automatic MRI meningioma segmentation using estimation maximization,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, IEEE, 2005.
- [209] P. Anbeek, K. L. Vincken, G. S. van Bochove, M. J. van Osch, and J. van der Grond, “Probabilistic segmentation of brain tissue in MR imaging,” *NeuroImage*, vol. 27, pp. 795–804, oct 2005.
- [210] H. A. Vrooman, C. A. Cocosco, F. van der Lijn, R. Stokking, M. A. Ikram, M. W. Vernooij, M. M. Breteler, and W. J. Niessen, “Multi-spectral brain tissue segmentation using automatically trained k-nearest-neighbor classification,” *NeuroImage*, vol. 37, pp. 71–81, aug 2007.
- [211] W.-H. Chao, Y.-Y. Chen, C.-W. Cho, S.-H. Lin, Y.-Y. I. Shih, and S. Tsang, “Improving segmentation accuracy for magnetic resonance imaging using a boosted decision tree,” *Journal of Neuroscience Methods*, vol. 175, pp. 206–217, nov 2008.
- [212] A. van Opbroek, F. van der Lijn, and M. de Bruijne, “Automated brain-tissue segmentation by multi-feature svm classification,” in *Proceedings of the MICCAI Workshops—The MICCAI Grand Challenge on MR Brain Image Segmentation (MRBrainS’13)*, 2013.
- [213] Z. Yi, A. Criminisi, J. Shotton, and A. Blake, “Discriminative, semantic segmentation of brain tissue in MR images,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009*, pp. 558–565, Springer Berlin Heidelberg, 2009.
- [214] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [215] D. Tian and L. Fan, “A brain MR images segmentation method based on SOM neural network,” in *2007 1st International Conference on Bioinformatics and Biomedical Engineering*, IEEE, 2007.

- [216] İ. Güler, A. Demirhan, and R. Karakiş, “Interpretation of MR images using self-organizing maps and knowledge-based expert systems,” *Digital Signal Processing*, vol. 19, pp. 668–677, jul 2009.
- [217] A. Ortiz, J. Górriz, J. Ramírez, D. Salas-González, and J. Llamas-Elvira, “Two fully-unsupervised methods for MR brain image segmentation using SOM-based strategies,” *Applied Soft Computing*, vol. 13, pp. 2668–2682, may 2013.
- [218] A. Ortiz, J. Gorriz, J. Ramirez, and D. Salas-Gonzalez, “Improving MR brain image segmentation using self-organising maps and entropy-gradient clustering,” *Information Sciences*, vol. 262, pp. 117–136, mar 2014.
- [219] C. J. J. Sheela and G. Suganthi, “Morphological edge detection and brain tumor segmentation in magnetic resonance (MR) images based on region growing and performance evaluation of modified fuzzy c-means (FCM) algorithm,” *Multimedia Tools and Applications*, vol. 79, pp. 17483–17496, feb 2020.
- [220] Y. Xia, Z. Ji, and Y. Zhang, “Brain MRI image segmentation based on learning local variational gaussian mixture models,” *Neurocomputing*, vol. 204, pp. 189–197, sep 2016.
- [221] K.-B. Liang, Y.-H. Guan, and Y.-T. Luo, “A brain MR image segmentation method based on gaussian model and markov random field,” in *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, IEEE, oct 2016.
- [222] A. Chan, I. A. Wood, and J. Fripp, “Maximum pseudolikelihood estimation for mixture-markov random field segmentation of the brain,” in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, IEEE, nov 2016.
- [223] M. Ismail, A. Soliman, M. Ghazal, A. E. Switala, G. Gimel’farb, G. N. Barnes, A. Khalil, and A. El-Baz, “A fast stochastic framework for automatic MR brain images segmentation,” *PLOS ONE*, vol. 12, nov 2017.
- [224] X. Wang, S. He, and Z. Tong, “Improved mixture model for markov random field and its application in magnetic resonance image segmentation,” *Journal of Medical Imaging and Health Informatics*, vol. 7, pp. 323–329, apr 2017.
- [225] M. Yaqub, M. K. Javaid, C. Cooper, and J. A. Noble, “Investigation of the role of feature selection and weighted voting in random forests for 3-d volumetric segmentation,” *IEEE Transactions on Medical Imaging*, vol. 33, pp. 258–271, feb 2014.

- [226] S. Pereira, A. Pinto, J. Oliveira, A. M. Mendrik, J. H. Correia, and C. A. Silva, “Automatic brain tissue segmentation in MR images using random forests and conditional random fields,” *Journal of Neuroscience Methods*, vol. 270, pp. 111–123, sep 2016.
- [227] S. C. V. D. Leemput, M. Meijs, A. Patel, F. J. A. Meijer, B. V. Ginneken, and R. Manniesing, “Multiclass brain tissue segmentation in 4d CT using convolutional neural networks,” *IEEE Access*, vol. 7, pp. 51557–51569, 2019.
- [228] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, “Automatic segmentation of MR brain images with a convolutional neural network,” *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1252–1261, may 2016.
- [229] S. Bao and A. C. S. Chung, “Multi-scale structured CNN with label consistency for brain MR image segmentation,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, pp. 113–117, jun 2016.
- [230] C. Wachinger, M. Reuter, and T. Klein, “DeepNAT: Deep convolutional neural network for segmenting neuroanatomy,” *NeuroImage*, vol. 170, pp. 434–445, apr 2018.
- [231] J. Bernal, K. Kushibar, M. Cabezas, S. Valverde, A. Oliver, and X. Llado, “Quantitative analysis of patch-based fully convolutional neural networks for tissue segmentation on brain magnetic resonance imaging,” *IEEE Access*, vol. 7, pp. 89986–90002, 2019.
- [232] Y. Chen, B. Shi, Z. Wang, P. Zhang, C. D. Smith, and J. Liu, “Hippocampus segmentation through multi-view ensemble ConvNets,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE, apr 2017.
- [233] C. G. B. Yogananda, B. C. Wagner, G. K. Murugesan, A. Madhuranthakam, and J. A. Maldjian, “A deep learning pipeline for automatic skull stripping and brain segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, apr 2019.
- [234] M. Luna and S. H. Park, “3d patchwise u-net with transition layers for mr brain segmentation,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, and T. van Walsum, eds.), (Cham), pp. 394–403, Springer International Publishing, jan 2019.

- [235] N. Yamanakkanavar and B. Lee, “Using a patch-wise m-net convolutional neural network for tissue segmentation in brain MRI images,” *IEEE Access*, vol. 8, pp. 120946–120958, 2020.
- [236] B. Fischl, “FreeSurfer,” *NeuroImage*, vol. 62, pp. 774–781, aug 2012.
- [237] L. Palumbo, P. Bosco, M. Fantacci, E. Ferrari, P. Oliva, G. Spera, and A. Retico, “Evaluation of the intra- and inter-method agreement of brain MRI segmentation software packages: A comparison between SPM12 and FreeSurfer v6.0,” *Physica Medica*, vol. 64, pp. 261–272, aug 2019.
- [238] M. Ganzetti, N. Wenderoth, and D. Mantini, “Quantitative evaluation of intensity inhomogeneity correction methods for structural MR brain images,” *Neuroinformatics*, vol. 14, pp. 5–21, aug 2015.
- [239] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, “A nonparametric method for automatic correction of intensity nonuniformity in mri data,” *IEEE transactions on medical imaging*, vol. 17, pp. 87–97, feb 1998.
- [240] D. N. Greve and B. Fischl, “Accurate and robust brain image alignment using boundary-based registration,” *NeuroImage*, vol. 48, pp. 63–72, oct 2009.
- [241] B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale, “Whole brain segmentation,” *Neuron*, vol. 33, pp. 341–355, jan 2002.
- [242] R. A. Morey, C. M. Petty, Y. Xu, J. P. Hayes, H. R. Wagner, D. V. Lewis, K. S. LaBar, M. Styner, and G. McCarthy, “A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes,” *NeuroImage*, vol. 45, pp. 855–866, apr 2009.
- [243] B. Fischl, “Automatically parcellating the human cerebral cortex,” *Cerebral Cortex*, vol. 14, pp. 11–22, jan 2004.
- [244] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, “An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest,” *NeuroImage*, vol. 31, pp. 968–980, jul 2006.
- [245] P. Raniga, P. Bourgeat, J. Fripp, O. Acosta, V. L. Villemagne, C. Rowe, C. L. Masters, G. Jones, G. O’Keefe, O. Salvado, and S. Ourselin,

- “Automated 11c-PiB standardized uptake value ratio,” *Academic Radiology*, vol. 15, pp. 1376–1389, nov 2008.
- [246] H. J. Son, J. S. Oh, M. Oh, S. J. Kim, J.-H. Lee, J. H. Roh, and J. S. Kim, “The clinical feasibility of deep learning-based classification of amyloid PET images in visually equivocal cases,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 47, pp. 332–341, dec 2019.
- [247] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, and B. Fischl, “Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer,” *NeuroImage*, vol. 32, pp. 180–194, aug 2006.
- [248] A. M. Dale, B. Fischl, and M. I. Sereno, “Cortical Surface-Based Analysis,” *NeuroImage*, vol. 9, pp. 179–194, feb 1999.
- [249] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, “Automatic 3d intersubject registration of mr volumetric data in standardized talairach space.,” *Journal of computer assisted tomography*, vol. 18, pp. 192–205, 1994.
- [250] J. L. Lancaster, M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten, and P. T. Fox, “Automated talairach atlas labels for functional brain mapping.,” *Human brain mapping*, vol. 10, pp. 120–131, jul 2000.
- [251] A. M. Dale and M. I. Sereno, “Improved localizadon of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach,” *Journal of Cognitive Neuroscience*, vol. 5, pp. 162–176, apr 1993.
- [252] B. Fischl and A. M. Dale, “Measuring the thickness of the human cerebral cortex from magnetic resonance images,” *Proceedings of the National Academy of Sciences*, vol. 97, pp. 11050–11055, sep 2000.
- [253] B. Fischl, A. Liu, and A. Dale, “Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 70–80, 2001.
- [254] N. J. Tustison, P. A. Cook, A. Klein, G. Song, S. R. Das, J. T. Duda, B. M. Kandel, N. van Strien, J. R. Stone, J. C. Gee, and B. B. Avants, “Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements,” *NeuroImage*, vol. 99, pp. 166–179, oct 2014.

- [255] P. A. Yushkevich, J. B. Pluta, H. Wang, L. Xie, S.-L. Ding, E. C. Gertje, L. Mancuso, D. Klot, S. R. Das, and D. A. Wolk, “Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment,” *Human Brain Mapping*, vol. 36, pp. 258–287, sep 2014.
- [256] R. Dahnke, R. A. Yotter, and C. Gaser, “Cortical thickness and central surface estimation,” *NeuroImage*, vol. 65, pp. 336–348, jan 2013.
- [257] R. Maani, Y. H. Yang, and S. Kalra, “Voxel-based texture analysis of the brain,” *PLOS ONE*, vol. 10, mar 2015.
- [258] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, pp. 610–621, nov 1973.
- [259] J. Zhang, C. Yu, G. Jiang, W. Liu, and L. Tong, “3d texture analysis on MRI images of alzheimer’s disease,” *Brain Imaging and Behavior*, vol. 6, pp. 61–69, nov 2011.
- [260] F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, I. A. Illán, and A. Ortiz, “Evaluating alzheimer’s disease diagnosis using texture analysis,” in *Communications in Computer and Information Science*, pp. 470–481, Springer International Publishing, 2017.
- [261] Y. Ding, C. Zhang, T. Lan, Z. Qin, X. Zhang, and W. Wang, “Classification of alzheimer’s disease based on the combination of morphometric feature and texture feature,” in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, nov 2015.
- [262] J. Ashburner and K. J. Friston, “Voxel-based morphometry - the methods,” *NeuroImage*, vol. 11, pp. 805–821, jun 2000.
- [263] K. Vaithinathan and L. Parthiban, “A novel texture extraction technique with t1 weighted MRI for the classification of alzheimer’s disease,” *Journal of Neuroscience Methods*, vol. 318, pp. 84–99, apr 2019.
- [264] J.-H. Cai, Y. He, X.-L. Zhong, H. Lei, F. Wang, G.-H. Luo, H. Zhao, and J.-C. Liu, “Magnetic resonance texture analysis in alzheimer’s disease,” *Academic Radiology*, vol. 27, pp. 1774–1783, dec 2020.
- [265] M. Termenon and M. Graña, “A two stage sequential ensemble applied to the classification of alzheimer’s disease based on MRI features,” *Neural Processing Letters*, vol. 35, pp. 1–12, oct 2011.
- [266] M. Liu, D. Zhang, and D. S. and, “View-centralized multi-atlas classification for alzheimer’s disease diagnosis,” *Human Brain Mapping*, vol. 36, pp. 1847–1865, jan 2015.

- [267] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, “Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer’s disease: a machine learning approach,” *Frontiers in Neuroscience*, vol. 9, sep 2015.
- [268] A. V. Lebedev, E. Westman, M. K. Beyer, M. G. Kramberger, C. Aguilar, Z. Pirtosek, and D. Aarsland, “Multivariate classification of patients with alzheimer’s and dementia with lewy bodies using high-dimensional cortical thickness measurements: an MRI surface-based morphometric study,” *Journal of Neurology*, vol. 260, pp. 1104–1115, dec 2012.
- [269] M. C. Campbell, J. Markham, H. Flores, J. M. Hartlein, A. M. Goate, N. J. Cairns, T. O. Videen, and J. S. Perlmutter, “Principal component analysis of PiB distribution in parkinson and alzheimer diseases,” *Neurology*, vol. 81, pp. 520–527, jul 2013.
- [270] R. K. Lama, J. Gwak, J.-S. Park, and S.-W. Lee, “Diagnosis of alzheimer’s disease based on structural MRI images using a regularized extreme learning machine and PCA features,” *Journal of Healthcare Engineering*, vol. 2017, pp. 1–11, 2017.
- [271] S. Alam and G.-R. K. and, “Alzheimer disease classification using KPCA, LDA, and multi-kernel learning SVM,” *International Journal of Imaging Systems and Technology*, vol. 27, pp. 133–143, jun 2017.
- [272] F. J. Martínez-Murcia, A. Ortiz, J.-M. Górriz, J. Ramírez, and D. Castillo-Barnes, “Studying the manifold structure of alzheimer’s disease: A deep learning approach using convolutional autoencoders,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 17–26, jan 2020.
- [273] M. Rönkkö, C. N. McIntosh, J. Antonakis, and J. R. Edwards, “Partial least squares path modeling: Time for some serious second thoughts,” *Journal of Operations Management*, vol. 47-48, pp. 9–27, jun 2016.
- [274] F. Segovia, J. M. Górriz, J. Ramírez, I. Álvarez, J. M. Jiménez-Hoyuela, and S. J. Ortega, “Improved parkinsonism diagnosis using a partial least squares based approach,” *Medical Physics*, vol. 39, pp. 4395–4403, 2012.
- [275] A. H. Andersen, W. S. Rayens, Y. Liu, and C. D. Smith, “Partial least squares for discrimination in fMRI data,” *Magnetic Resonance Imaging*, vol. 30, pp. 446–452, apr 2012.
- [276] R. Chaves, J. Ramírez, J. Górriz, and C. Puntonet, “Association rule-based feature selection method for alzheimer’s disease diagnosis,” *Expert Systems with Applications*, vol. 39, pp. 11766–11774, oct 2012.

- [277] P. Wang, K. Chen, L. Yao, B. Hu, X. Wu, J. Zhang, Q. Ye, X. Guo, and for the Alzheimer’s Disease Neuroimaging Initiative, “Multimodal classification of mild cognitive impairment based on partial least squares,” *Journal of Alzheimer’s Disease*, vol. 54, pp. 359–371, 2016.
- [278] K. Hett, V.-T. Ta, J. V. Manjón, and P. Coupé, “Adaptive fusion of texture-based grading for alzheimer’s disease classification,” *Computerized Medical Imaging and Graphics*, vol. 70, pp. 8–16, dec 2018.
- [279] R. Mehrotra, K. Namuduri, and N. Ranganathan, “Gabor filter-based edge detection,” *Pattern Recognition*, vol. 25, pp. 1479–1494, dec 1992.
- [280] P. Coupé, S. F. Eskildsen, J. V. Manjón, V. S. Fonov, J. C. Pruessner, M. Allard, and D. L. Collins, “Scoring by nonlocal image patch estimator for early detection of alzheimer’s disease,” *NeuroImage: Clinical*, vol. 1, no. 1, pp. 141–152, 2012.
- [281] L. Nanni, S. Brahmam, C. Salvatore, and I. Castiglioni, “Texture descriptors and voxels for the early diagnosis of alzheimer’s disease,” *Artificial Intelligence in Medicine*, vol. 97, pp. 19–26, jun 2019.
- [282] Y. Meyer, *Wavelets and operators*. Cambridge England New York: Cambridge University Press, 1992.
- [283] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Transactions on Image Processing*, vol. 19, pp. 1635–1650, jun 2010.
- [284] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial Intelligence*, vol. 97, pp. 273–324, dec 1997.
- [285] J. H. Gennari, P. Langley, and D. Fisher, “Models of incremental concept formation,” *Artificial Intelligence*, vol. 40, pp. 11–61, sep 1989.
- [286] C. Aggarwal, *Data Clustering : Algorithms and Applications*. Boca Raton, FL: Chapman and Hall/CRC, 2018.
- [287] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, “Benchmark for filter methods for feature selection in high-dimensional classification data,” *Computational Statistics & Data Analysis*, vol. 143, p. 106839, mar 2020.
- [288] S. Maldonado, R. Weber, and F. Famili, “Feature selection for high-dimensional class-imbalanced data sets using support vector machines,” *Information Sciences*, vol. 286, pp. 228–246, dec 2014.

- [289] S. Dimitriadis, D. Liparas, and M. N. Tsolaki, “Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healthy elderly, MCI, cMCI and alzheimer’s disease patients: From the alzheimer’s disease neuroimaging initiative (ADNI) database,” *Journal of Neuroscience Methods*, vol. 302, pp. 14–23, may 2018.
- [290] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, and C. G. Puntonet, “Selecting regions of interest for the diagnosis of alzheimer’s disease in brain SPECT images using welch’s t-test,” in *Lecture Notes in Computer Science*, pp. 965–972, Springer Berlin Heidelberg, 2009.
- [291] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. Álvarez, F. Segovia, R. Chaves, and C. G. Puntonet, “Computer-aided diagnosis of alzheimer’s disease using support vector machines and classification trees,” *Physics in Medicine and Biology*, vol. 55, pp. 2807–2817, apr 2010.
- [292] A. Savio and M. Graña, “Deformation based feature selection for computer aided diagnosis of alzheimer’s disease,” *Expert Systems with Applications*, vol. 40, pp. 1619–1628, apr 2013.
- [293] D. Salas-Gonzalez, F. Segovia, F. J. Martínez-Murcia, E. W. Lang, J. M. Górriz, and J. Ramírez, “An optimal approach for selecting discriminant regions for the diagnosis of alzheimer’s disease,” *Current Alzheimer Research*, vol. 13, pp. 838–844, may 2016.
- [294] H. Gunduz, “An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on parkinson’s disease classification,” *Biomedical Signal Processing and Control*, vol. 66, p. 102452, apr 2021.
- [295] Q. Gu, Z. Li, and J. Han, “Generalized fisher score for feature selection,” in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, UAI’11, (Arlington, Virginia, USA), pp. 266–273, AUAI Press, 2011.
- [296] F. Martínez-Murcia, J. Górriz, J. Ramírez, C. Puntonet, and D. Salas-González, “Computer aided diagnosis tool for alzheimer’s disease based on mann-whitney-wilcoxon u-test,” *Expert Systems with Applications*, vol. 39, pp. 9676–9685, aug 2012.
- [297] D. L. Campbell, H. Kang, and S. Shokouhi, “Application of haralick texture features in brain [18f]-florbetapir positron emission tomography without reference region normalization,” *Clinical Interventions in Aging*, vol. Volume 12, pp. 2077–2086, dec 2017.

- [298] S. Lahmiri and A. Shmuel, “Detection of parkinson’s disease based on voice patterns ranking and optimized support vector machine,” *Biomedical Signal Processing and Control*, vol. 49, pp. 427–433, mar 2019.
- [299] J. Ramírez, J. Górriz, A. Ortiz, F. Martínez-Murcia, F. Segovia, D. Salas-Gonzalez, D. Castillo-Barnes, I. Illan, and C. Puntonet, “Ensemble of random forests one vs. rest classifiers for MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares,” *Journal of Neuroscience Methods*, vol. 302, pp. 47–57, may 2018.
- [300] N. Madeira, J. V. Duarte, R. Martins, G. N. Costa, A. Macedo, and M. Castelo-Branco, “Morphometry and gyrification in bipolar disorder and schizophrenia: A comparative MRI study,” *NeuroImage: Clinical*, vol. 26, p. 102220, 2020.
- [301] C. Aggarwal, *Data classification : algorithms and applications*. Boca Raton: CRC Press, 2015.
- [302] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of relieff and rrelieff,” *Machine Learning*, vol. 53, pp. 23–69, 2003.
- [303] W. Lee, B. Park, and K. Han, “SVM-based classification of diffusion tensor imaging data for diagnosing alzheimer’s disease and mild cognitive impairment,” in *Intelligent Computing Theories and Methodologies*, pp. 489–499, Springer International Publishing, 2015.
- [304] B. Schölkopf, *Learning with kernels - support vector machines, regularization, optimization, and beyond*. Cambridge, Mass: MIT Press, 2002.
- [305] V. Vapnik, *Estimation of Dependences Based on Empirical Data*. No. 1, Springer New York, 2006.
- [306] A. Farzan, S. Mashohor, A. R. Ramli, and R. Mahmud, “Boosting diagnosis accuracy of alzheimer’s disease using high dimensional recognition of longitudinal brain atrophy patterns,” *Behavioural Brain Research*, vol. 290, pp. 124–130, sep 2015.
- [307] B. Peng, S. Wang, Z. Zhou, Y. Liu, B. Tong, T. Zhang, and Y. Dai, “A multilevel-ROI-features-based machine learning method for detection of morphometric biomarkers in parkinson’s disease,” *Neuroscience Letters*, vol. 651, pp. 88–94, jun 2017.

- [308] L. Sørensen and M. Nielsen, “Ensemble support vector machine classification of dementia using structural MRI and mini-mental state examination,” *Journal of Neuroscience Methods*, vol. 302, pp. 66–74, may 2018.
- [309] S. Anita and P. A. Priya, “Three dimensional analysis of SPECT images for diagnosing early parkinson’s disease using radial basis function kernel - extreme learning machine,” *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 15, pp. 461–470, jun 2019.
- [310] S. Anita, “Improved classification accuracy for diagnosing the early stage of parkinson’s disease using alpha stable distribution,” *IETE Journal of Research*, pp. 1–12, apr 2021.
- [311] P.-H. Chen, C.-J. Lin, and B. Schölkopf, “A tutorial on ν -support vector machines,” *Applied Stochastic Models in Business and Industry*, vol. 21, pp. 111–136, mar 2005.
- [312] M. Abdar, U. R. Acharya, N. Sarrafzadegan, and V. Makarenkov, “NE-nu-SVC: A new nested ensemble clinical decision support system for effective diagnosis of coronary artery disease,” *IEEE Access*, vol. 7, pp. 167605–167620, 2019.
- [313] P. Domingos and M. Pazzani, “On the optimality of the simple bayesian classifier under zero-one loss,” *Machine Learning*, vol. 29, no. 2/3, pp. 103–130, 1997.
- [314] D. Berg, R. B. Postuma, C. H. Adler, B. R. Bloem, P. Chan, B. Dubois, T. Gasser, C. G. Goetz, G. Halliday, L. Joseph, A. E. Lang, I. Liepelt-Scarfone, I. Litvan, K. Marek, J. Obeso, W. Oertel, C. W. Olanow, W. Poewe, M. Stern, and G. Deuschl, “MDS research criteria for prodromal parkinson’s disease,” *Movement Disorders*, vol. 30, pp. 1600–1611, oct 2015.
- [315] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, “High-accuracy detection of early parkinson’s disease through multimodal features and machine learning,” *International Journal of Medical Informatics*, vol. 90, pp. 13–21, jun 2016.
- [316] M. Dyrba, M. Ewers, M. Wegrzyn, I. Kilimann, C. Plant, A. Oswald, T. Meindl, M. Pievani, A. L. W. Bokde, A. Fellgiebel, M. Filippi, H. Hampel, S. Klöppel, K. Hauenstein, T. Kirste, and S. J. T. and, “Robust automated detection of microstructural white matter degeneration in alzheimer’s disease using machine learning classification of multicenter DTI data,” *PLoS ONE*, vol. 8, may 2013.

- [317] F. Martínez-Murcia, J. Górriz, J. Ramírez, C. Puntonet, and I. Illán, “Functional activity maps based on significance measures and independent component analysis,” *Computer Methods and Programs in Biomedicine*, vol. 111, pp. 255–268, jul 2013.
- [318] J. Dukart, F. Sambataro, A. Bertolino, and for the Alzheimer’s Disease Neuroimaging Initiative, “Accurate prediction of conversion to alzheimer’s disease using imaging, genetic, and neuropsychological biomarkers,” *Journal of Alzheimer’s Disease*, vol. 49, pp. 1143–1159, 2015.
- [319] A. Khazaei, A. Ebrahimzadeh, and A. Babajani-Feremi, “Classification of patients with MCI and AD from healthy controls using directed graph measures of resting-state fMRI,” *Behavioural Brain Research*, vol. 322, pp. 339–350, mar 2017.
- [320] B. S. Wade, S. H. Joshi, B. A. Gutman, and P. M. Thompson, “Machine learning on high dimensional shape data from subcortical brain surfaces: A comparison of feature selection and classification methods,” *Pattern Recognition*, vol. 63, pp. 731–739, mar 2017.
- [321] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [322] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, “Random forest-based similarity measures for multi-modal classification of alzheimer’s disease,” *NeuroImage*, vol. 65, pp. 167–175, jan 2013.
- [323] J. Ramírez, J. M. Górriz, A. Ortiz, P. Padilla, and F. J. Martínez-Murcia, “Ensemble tree learning techniques for magnetic resonance image analysis,” in *Innovation in Medicine and Healthcare 2015*, pp. 395–404, Springer International Publishing, aug 2015.
- [324] M. R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. T. Inan, and H. Liao, “Neuroimaging and machine learning for dementia diagnosis: Recent advancements and future prospects,” *IEEE Reviews in Biomedical Engineering*, vol. 12, pp. 19–33, 2019.
- [325] J. Liu, S. Shang, K. Zheng, and J.-R. Wen, “Multi-view ensemble learning for dementia diagnosis from neuroimaging: An artificial neural network approach,” *Neurocomputing*, vol. 195, pp. 112–116, jun 2016.
- [326] N. Bhagwat, J. Pipitone, A. N. Voineskos, and M. M. C. and, “An artificial neural network model for clinical score prediction in alzheimer disease using structural neuroimaging measures,” *Journal of Psychiatry and Neuroscience*, vol. 44, pp. 246–250, jul 2019.

- [327] J. Tang, B. Yang, M. P. Adams, N. N. Shenkov, I. S. Klyuzhin, S. Fotouhi, E. Davoodi-Bojd, L. Lu, H. Soltanian-Zadeh, V. Sossi, and A. Rahmim, “Artificial neural network-based prediction of outcome in parkinson’s disease patients using DaTscan SPECT imaging features,” *Molecular Imaging and Biology*, vol. 21, pp. 1165–1173, mar 2019.
- [328] G. E. Hinton, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, pp. 504–507, jul 2006.
- [329] V. K. Ithapu, V. Singh, O. C. Okonkwo, R. J. Chappell, N. M. Dowling, and S. C. J. and, “Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment,” *Alzheimer’s & Dementia*, vol. 11, pp. 1489–1499, jun 2015.
- [330] P. Bhatkoti and M. Paul, “Early diagnosis of alzheimer’s disease: A multi-class deep learning framework with modified k-sparse autoencoder classification,” in *2016 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, IEEE, nov 2016.
- [331] C. V. Dolph, M. Alam, Z. Shboul, M. D. Samad, and K. M. Iftekharrudin, “Deep learning of texture and structural features for multiclass alzheimer’s disease classification,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, may 2017.
- [332] D. Lu, , K. Popuri, G. W. Ding, R. Balachandar, and M. F. Beg, “Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer’s disease using structural MR and FDG-PET images,” *Scientific Reports*, vol. 8, apr 2018.
- [333] J. Islam and Y. Zhang, “A novel deep learning based multi-class classification method for alzheimer’s disease detection using brain MRI data,” in *Brain Informatics*, pp. 213–222, Springer International Publishing, 2017.
- [334] E. M. Alkabawi, A. R. Hilal, and O. A. Basir, “Computer-aided classification of multi-types of dementia via convolutional neural networks,” in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, may 2017.
- [335] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, “Refining diagnosis of parkinson’s disease with deep learning-based interpretation of dopamine transporter imaging,” *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.
- [336] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, and X. Q. and, “Convolutional neural networks-based MRI

- image analysis for the alzheimer’s disease prediction from mild cognitive impairment,” *Frontiers in Neuroscience*, vol. 12, nov 2018.
- [337] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [338] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84–90, may 2017.
- [339] C. Feng, A. Elazab, P. Yang, T. Wang, B. Lei, and X. Xiao, “3d convolutional neural network and stacked bidirectional recurrent neural network for alzheimer’s disease diagnosis,” in *PRedictive Intelligence in Medicine*, pp. 138–146, Springer International Publishing, 2018.
- [340] R. Cui, M. Liu, and G. Li, “Longitudinal analysis for alzheimer’s disease diagnosis using RNN,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, apr 2018.
- [341] F. Li and M. Liu, “A hybrid convolutional and recurrent neural network for hippocampus analysis in alzheimer’s disease,” *Journal of Neuroscience Methods*, vol. 323, pp. 108–118, jul 2019.
- [342] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, “Classification of alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning,” in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, jun 2018.
- [343] T.-D. Vu, N.-H. Ho, H.-J. Yang, J. Kim, and H.-C. Song, “Non-white matter tissue extraction and deep convolutional neural network for alzheimer’s disease detection,” *Soft Computing*, vol. 22, pp. 6825–6833, aug 2018.
- [344] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, “Convolutional neural network based alzheimer’s disease classification from magnetic resonance brain images,” *Cognitive Systems Research*, vol. 57, pp. 147–159, oct 2019.
- [345] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, “Classification and visualization of alzheimer’s disease using volumetric convolutional neural network and transfer learning,” *Scientific Reports*, vol. 9, dec 2019.
- [346] D. H. Kim, H. Wit, and M. Thurston, “Artificial intelligence in the diagnosis of parkinson’s disease from ioflupane-123 single-photon emission computed tomography dopamine transporter scans using transfer

- learning,” *Nuclear Medicine Communications*, vol. 39, pp. 887–893, oct 2018.
- [347] G. Martí-Juan, G. Sanroma-Guell, and G. Piella, “A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in alzheimer’s disease,” *Computer Methods and Programs in Biomedicine*, vol. 189, p. 105348, jun 2020.
- [348] G. M. Fitzmaurice and C. Ravichandran, “A primer in longitudinal data analysis,” *Circulation*, vol. 118, pp. 2005–2010, nov 2008.
- [349] J. Ye, M. Farnum, E. Yang, R. Verbeeck, V. Lobanov, N. Raghavan, G. Novak, A. DiBernardo, and V. A. Narayan, “Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data,” *BMC Neurology*, vol. 12, jun 2012.
- [350] D. Zhang and D. S. and, “Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers,” *PLoS ONE*, vol. 7, mar 2012.
- [351] H. Wang, F. Nie, H. Huang, J. Yan, S. Kim, S. L. Risacher, A. J. Saykin, and L. Shen, “High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer’s disease progression prediction,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS’12*, (Red Hook, NY, USA), pp. 1277–1285, Curran Associates Inc., 2012.
- [352] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, “Modeling disease progression via multi-task learning,” *NeuroImage*, vol. 78, pp. 233–248, sep 2013.
- [353] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, pp. 267–288, jan 1996.
- [354] L. Huang, Y. Jin, Y. Gao, K.-H. Thung, and D. Shen, “Longitudinal clinical score prediction in alzheimer’s disease with soft-split sparse regression based random forest,” *Neurobiology of Aging*, vol. 46, pp. 180–191, oct 2016.
- [355] B. Lei, F. Jiang, S. Chen, D. Ni, and T. Wang, “Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning,” *Frontiers in Aging Neuroscience*, vol. 9, mar 2017.
- [356] B. Jie, M. Liu, J. Liu, D. Zhang, and D. Shen, “Temporally constrained group sparse learning for longitudinal data analysis in alzheimer’s disease,” *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 238–249, jan 2017.

- [357] L. M. Aksman, M. A. Scelsi, A. F. Marquand, D. C. Alexander, S. Ourselin, and A. A. and, “Modeling longitudinal imaging biomarkers with parametric bayesian multi-task learning,” *Human Brain Mapping*, vol. 40, pp. 3982–4000, jun 2019.
- [358] D. A. Rastegar, N. Ho, G. M. Halliday, and N. Dzamko, “Parkinson’s progression prediction using machine learning and serum cytokines,” *npj Parkinson’s Disease*, vol. 5, jul 2019.
- [359] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, pp. 301–320, apr 2005.
- [360] B. Cheng, , M. Liu, H.-I. Suk, D. Shen, and D. Zhang, “Multimodal manifold-regularized transfer learning for MCI conversion prediction,” *Brain Imaging and Behavior*, vol. 9, pp. 913–926, feb 2015.
- [361] H.-I. Suk, S.-W. Lee, and D. Shen, “Deep ensemble learning of sparse regression models for brain disease diagnosis,” *Medical Image Analysis*, vol. 37, pp. 101–113, apr 2017.
- [362] A. Ortiz, , J. Munilla, F. J. Martínez-Murcia, J. M. Górriz, and J. Ramírez, “Learning longitudinal MRI patterns by SICE and deep learning: Assessing the alzheimer’s disease progression,” in *Communications in Computer and Information Science*, pp. 413–424, Springer International Publishing, 2017.
- [363] B. A. Pearlmutter, “Learning state space trajectories in recurrent neural networks,” *Neural Computation*, vol. 1, pp. 263–269, jun 1989.
- [364] N. Bhagwat, J. D. Viviano, A. N. Voineskos, and M. M. C. and, “Modeling and prediction of clinical symptom trajectories in alzheimer’s disease using longitudinal data,” *PLOS Computational Biology*, vol. 14, sep 2018.
- [365] D. Chicco, “Siamese neural networks: An overview,” in *Methods in Molecular Biology*, pp. 73–94, Springer US, aug 2020.
- [366] C. Wu, , S. Guo, Y. Hong, B. Xiao, Y. Wu, and Q. Zhang, “Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks,” *Quantitative Imaging in Medicine and Surgery*, vol. 8, pp. 992–1003, nov 2018.
- [367] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituiev, T. P. Copeland, M. S. Aboian, C. M. Aparici, S. C. Behr, R. R. Flavell, S.-Y. Huang, K. A. Zalocusky, L. Nardo, Y. Seo, R. A. Hawkins, M. H. Pampaloni, D. Hadley, and B. L. Franc, “A

- deep learning model to predict a diagnosis of alzheimer disease by using 18f-FDG PET of the brain,” *Radiology*, vol. 290, pp. 456–464, feb 2019.
- [368] M. M. Ghazi, M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, and L. Sørensen, “Training recurrent neural networks robust to incomplete data: Application to alzheimer’s disease progression modeling,” *Medical Image Analysis*, vol. 53, pp. 39–46, apr 2019.
- [369] H. Li and Y. Fan, “Early prediction of alzheimer’s disease dementia based on baseline hippocampal MRI and 1-year follow-up cognitive measures using deep recurrent neural networks,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, apr 2019.
- [370] H. Li, M. Habes, D. A. Wolk, and Y. F. and, “A deep learning model for early prediction of alzheimer’s disease dementia based on hippocampal magnetic resonance imaging data,” *Alzheimer’s & Dementia*, vol. 15, pp. 1059–1070, jun 2019.
- [371] C. Nikias, *Signal processing with alpha-stable distributions and applications*. New York: Wiley, 1995.
- [372] D. Salas-Gonzalez, J. Górriz, J. Ramírez, M. Schloegl, E. Lang, and A. Ortiz, “Parameterization of the distribution of white and grey matter in MRI using the α -stable distribution,” *Computers in Biology and Medicine*, vol. 43, pp. 559–567, jun 2013.
- [373] M. Jas, T. Dupré la Tour, U. Simsekli, and A. Gramfort, “Learning the morphology of brain signals using alpha-stable convolutional sparse coding,” in *Advances in Neural Information Processing Systems 30* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), pp. 1099–1108, Curran Associates, Inc., 2017.
- [374] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, and E. W. Lang, “Linear intensity normalization of fp-cit spect brain images using the α -stable distribution,” *NeuroImage*, vol. 65, no. Supplement C, pp. 449–455, 2013.
- [375] J.-H. Kang, , B. Mollenhauer, C. S. Coffey, J. B. Toledo, D. Weintraub, D. R. Galasko, D. J. Irwin, V. V. Deerlin, A. S. Chen-Plotkin, C. Caspell-Garcia, T. Waligórska, P. Taylor, N. Shah, S. Pan, P. Zero, M. Frasier, K. Marek, K. Kiebertz, D. Jennings, C. M. Tanner, T. Simuni, A. Singleton, A. W. Toga, S. Chowdhury, J. Q. Trojanowski, and L. M. Shaw, “CSF biomarkers associated with disease heterogeneity in early parkinson’s disease: the parkinson’s progression markers

- initiative study,” *Acta Neuropathologica*, vol. 131, pp. 935–949, mar 2016.
- [376] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. C. Morris, R. C. Petersen, A. J. Saykin, L. M. Shaw, A. W. Toga, and J. Q. Trojanowski, “Recent publications from the alzheimer’s disease neuroimaging initiative: Reviewing progress toward improved AD clinical trials,” *Alzheimer’s & Dementia*, vol. 13, pp. 1–85, apr 2017.
- [377] K. Gwinn, K. K. David, C. Swanson-Fischer, R. Albin, C. S. Hillaire-Clarke, B.-A. Sieber, C. Lungu, F. D. Bowman, R. N. Alcalay, D. Babcock, T. M. Dawson, R. B. Dewey, T. Foroud, D. German, X. Huang, V. Petyuk, J. A. Potashkin, R. Saunders-Pullman, M. Sutherland, D. R. Walt, A. B. West, J. Zhang, A. Chen-Plotkin, C. R. Scherzer, D. E. Vaillancourt, and L. S. Rosenthal, “Parkinson’s disease biomarkers: perspective from the NINDS parkinson’s disease biomarkers program,” *Biomarkers in Medicine*, vol. 11, pp. 451–473, may 2017.
- [378] L. Farotti, S. Paciotti, A. Tasegian, P. Eusebi, and L. Parnetti, “Discovery, validation and optimization of cerebrospinal fluid biomarkers for use in parkinson’s disease,” *Expert Review of Molecular Diagnostics*, vol. 17, pp. 771–780, jun 2017.
- [379] S. Saiki, T. Hatano, M. Fujimaki, K.-I. Ishikawa, A. Mori, Y. Oji, A. Okuzumi, T. Fukuhara, T. Koinuma, Y. Imamichi, M. Nagumo, N. Furuya, S. Nojiri, T. Amo, K. Yamashiro, and N. Hattori, “Decreased long-chain acylcarnitines from insufficient beta-oxidation as potential early diagnostic markers for parkinson’s disease,” *Scientific Reports*, vol. 7, p. 7328, aug 2017.
- [380] B. Mollenhauer, C. J. Caspell-Garcia, C. S. Coffey, P. Taylor, L. M. Shaw, J. Q. Trojanowski, A. Singleton, M. Frasier, K. Marek, D. Galasko, and F. the Parkinson’s Progression Marker Initiative, “Longitudinal csf biomarkers in patients with early parkinson disease and healthy controls,” *Neurology*, vol. 89, pp. 1959–1969, oct 2017.
- [381] R. L. Handels, S. J. Vos, M. G. Kramberger, V. Jelic, K. Blennow, M. van Buchem, W. van der Flier, Y. Freund-Levi, H. Hampel, M. O. Rikkert, A. Oleksik, Z. Pirtosek, P. Scheltens, H. Soininen, C. Teunissen, M. Tsolaki, A. K. Wallin, B. Winblad, F. R. Verhey, and P. J. Visser, “Predicting progression to dementia in persons with mild cognitive impairment using cerebrospinal fluid markers,” *Alzheimer’s & Dementia*, vol. 13, no. 8, pp. 903–912, 2017.

- [382] A. D. Andersen, M. Binzer, E. Stenager, and J. B. Gramsbergen, “Cerebrospinal fluid biomarkers for parkinson’s disease – a systematic review,” *Acta Neurologica Scandinavica*, vol. 135, pp. 34–56, 2017.
- [383] T. F. Tropea and A. S. Chen-Plotkin, “Unlocking the mystery of biomarkers: A brief introduction, challenges and opportunities in parkinson disease,” *Parkinsonism & Related Disorders*, vol. 46, pp. 15–18, jan 2018. Proceedings of XXII World Congress of the International Association of Parkinsonism and Related Disorders, Ho Chi Minh City, Vietnam, 12-15 November 2017.
- [384] N. Abbasi, B. Mohajer, S. Abbasi, P. Hasanabadi, A. Abdolalizadeh, and R. Rajimehr, “Relationship between cerebrospinal fluid biomarkers and structural brain network properties in parkinson’s disease,” *Movement Disorders*, vol. 33, no. 3, pp. 431–439, 2018.
- [385] A. Siderowf, S. X. Xie, H. Hurtig, D. Weintraub, J. Duda, A. Chen-Plotkin, L. M. Shaw, V. V. Deerlin, J. Q. Trojanowski, and C. Clark, “CSF amyloid 1-42 predicts cognitive decline in parkinson disease,” *Neurology*, vol. 75, pp. 1055–1061, aug 2010.
- [386] S. Gallegos, C. Pacheco, C. Peters, C. M. Opazo, and L. G. Aguayo, “Features of alpha-synuclein that could explain the progression and irreversibility of parkinson’s disease,” *Frontiers in Neuroscience*, vol. 9, p. 59, 2015.
- [387] V. C. Constantinides, G. P. Paraskevas, E. Emmanouilidou, O. Petropoulou, A. Bougea, K. Vekrellis, I. Evdokimidis, E. Stamboulis, and E. Kapaki, “Csf biomarkers beta-amyloid, tau proteins and a-synuclein in the differential diagnosis of parkinson-plus syndromes,” *Journal of the Neurological Sciences*, vol. 382, no. Supplement C, pp. 91–95, 2017.
- [388] M. Delgado-Alvarado, B. Gago, A. Gorostidi, H. Jiménez-Urbieto, R. Dacosta-Aguayo, I. Navalpotro-Gómez, J. Ruiz-Martínez, A. Bergareche, J. F. Martí-Massó, P. Martínez-Lage, A. Izagirre, and M. C. Rodríguez-Oroz, “Tau/alpha-synuclein ratio and inflammatory proteins in parkinson’s disease: An exploratory study,” *Movement Disorders*, vol. 32, pp. 1066–1073, may 2017.
- [389] A. Garg, S. Appel-Cresswell, K. Popuri, M. J. McKeown, and M. F. Beg, “Morphological alterations in the caudate, putamen, pallidum, and thalamus in parkinson’s disease,” *Frontiers in Neuroscience*, vol. 9, p. 101, 2015.
- [390] J. Dukart, F. Sambataro, and A. Bertolino, “Distinct role of striatal functional connectivity and dopaminergic loss in parkinson’s symptoms,” *Frontiers in Aging Neuroscience*, vol. 9, p. 151, 2017.

- [391] U. Saeed, J. Compagnone, R. I. Aviv, A. P. Strafella, S. E. Black, A. E. Lang, and M. Masellis, “Imaging biomarkers in parkinson’s disease and parkinsonian syndromes: current and emerging concepts,” *Translational Neurodegeneration*, vol. 6, p. 8, mar 2017.
- [392] C. Owens-Walton, D. Jakabek, X. Li, F. A. Wilkes, M. Walterfang, D. Velakoulis, D. van Westen, J. C. Looi, and O. Hansson, “Striatal changes in parkinson disease: An investigation of morphology, functional connectivity and their relationship to clinical symptoms,” *Psychiatry Research: Neuroimaging*, vol. 275, pp. 5–13, 2018.
- [393] F. Segovia, J. M. Górriz, J. Ramírez, F. J. Martínez-Murcia, D. Castillo-Barnes, I. A. Illán, A. Ortiz, and D. Salas-Gonzalez, “Automatic separation of parkinsonian patients and control subjects based on the striatal morphology,” in *Natural and Artificial Computation for Biomedicine and Neuroscience: International Work-Conference on the Interplay Between Natural and Artificial Computation, IWINAC 2017, Corunna, Spain, June 19-23, 2017, Proceedings, Part I* (J. M. Ferrández Vicente, J. R. Álvarez-Sánchez, F. de la Paz López, J. Toledo Moreo, and H. Adeli, eds.), (Cham), pp. 345–352, Springer International Publishing, 2017.
- [394] J. Stoeckel, N. Ayache, G. Malandain, P. M. Koulibaly, K. P. Ebmeier, and J. Darcourt, “Automatic Classification of SPECT Images of Alzheimer’s Disease Patients and Control Subjects,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI*, vol. 3217 of *Lecture Notes in Computer Science*, pp. 654–662, Springer, 2004.
- [395] M. Molina-Molina, D. Castillo-Barnes, F. J. Martinez-Murcia, F. Segovia, J. M. Gorriz, J. Ramirez, and D. Salas-Gonzalez, “[123i]FP-CIT SPECT brain imaging for parkinson’s diagnosis using contour lines,” in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, p. 1, IEEE, nov 2018.
- [396] M. P. Fay and M. A. Proschan, “Wilcoxon-mann-whitney or t-test? on assumptions for hypothesis tests and multiple interpretations of decision rules,” *Statistics Surveys*, vol. 4, no. 0, pp. 1–39, 2010.
- [397] D. Freedman, R. Pisani, and R. Purves, *Statistics*. W W Norton & Co., mar 2007.
- [398] M. A. Lindquist, B. Caffo, and C. Crainiceanu, “Ironing out the statistical wrinkles in “ten ironic rules”,” *NeuroImage*, vol. 81, pp. 499–502, nov 2013.

- [399] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, “Machine learning methods for quantitative radiomic biomarkers,” *Scientific Reports*, vol. 5, aug 2015.
- [400] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, E. Liu, J. C. Morris, R. C. Petersen, A. J. Saykin, M. E. Schmidt, L. Shaw, J. A. Siuciak, H. Soares, A. W. Toga, and J. Q. T. and, “The alzheimer’s disease neuroimaging initiative: A review of papers published since its inception,” *Alzheimer’s & Dementia*, vol. 8, pp. S1–S68, nov 2011.
- [401] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, W. Poewe, B. Mollenhauer, P.-E. Klinik, T. Sherer, M. Frasier, C. Meunier, A. Rudolph, C. Casaceli, J. Seibyl, S. Mendick, N. Schuff, Y. Zhang, A. Toga, K. Crawford, A. Ansbach, P. D. Blasio, M. Piovella, J. Trojanowski, L. Shaw, A. Singleton, K. Hawkins, J. Eberling, D. Brooks, D. Russell, L. Leary, S. Factor, B. Sommerfeld, P. Hogarth, E. Pighetti, K. Williams, D. Standaert, S. Guthrie, R. Hauser, H. Delgado, J. Jankovic, C. Hunter, M. Stern, B. Tran, J. Leverenz, M. Baca, S. Frank, C.-A. Thomas, I. Richard, C. Deeley, L. Rees, F. Sprenger, E. Lang, H. Shill, S. Obradov, H. Fernandez, A. Winters, D. Berg, K. Gauss, D. Galasko, D. Fontaine, Z. Mari, M. Gerstenhaber, D. Brooks, S. Malloy, P. Barone, K. Longo, T. Comery, B. Ravina, I. Grachev, K. Gallagher, M. Collins, K. L. Widnell, S. Ostrowizki, P. Fontoura, T. Ho, J. Luthman, M. van der Brug, A. D. Reith, and P. Taylor, “The parkinson progression marker initiative (PPMI),” *Progress in Neurobiology*, vol. 95, pp. 629–635, dec 2011.
- [402] J. M. Ringman, J. O’Neill, D. Geschwind, L. Medina, L. G. Apostolova, Y. Rodriguez, B. Schaffer, A. Varpetian, B. Tseng, F. Ortiz, J. Fitten, J. L. Cummings, and G. Bartzokis, “Diffusion tensor imaging in preclinical and presymptomatic carriers of familial alzheimer’s disease mutations,” *Brain*, vol. 130, pp. 1767–1776, may 2007.
- [403] T. L. S. Benzinger, T. Blazey, C. R. Jack, R. A. Koeppe, Y. Su, C. Xiong, M. E. Raichle, A. Z. Snyder, B. M. Ances, R. J. Bateman, N. J. Cairns, A. M. Fagan, A. Goate, D. S. Marcus, P. S. Aisen, J. J. Christensen, L. Ercole, R. C. Hornbeck, A. M. Farrar, P. Aldea, M. S. Jaszec, C. J. Owen, X. Xie, R. Mayeux, A. Brickman, E. McDade, W. Klunk, C. A. Mathis, J. Ringman, P. M. Thompson, B. Ghetti, A. J. Saykin, R. A. Sperling, K. A. Johnson, S. Salloway, S. Correia, P. R. Schofield, C. L. Masters, C. Rowe, V. L. Villemagne, R. Martins, S. Ourselin, M. N. Rossor, N. C. Fox, D. M. Cash, M. W. Weiner, D. M. Holtzman, V. D. Buckles, K. Moulder, and J. C.

- Morris, “Regional variability of imaging biomarkers in autosomal dominant alzheimer’s disease,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 47, pp. 4502–4509, 2013.
- [404] D. M. Cash, G. R. Ridgway, Y. Liang, N. S. Ryan, K. M. Kinnunen, T. Yeatman, I. B. Malone, T. L. Benzinger, C. R. Jack, P. M. Thompson, B. F. Ghetti, A. J. Saykin, C. L. Masters, J. M. Ringman, S. P. Salloway, P. R. Schofield, R. A. Sperling, N. J. Cairns, D. S. Marcus, C. Xiong, R. J. Bateman, J. C. Morris, M. N. Rossor, S. Ourselin, and N. C. Fox, “The pattern of atrophy in familial alzheimer disease,” *Neurology*, vol. 81, no. 16, pp. 1425–1433, 2013.
- [405] J. M. Ringman, A. Goate, C. L. Masters, N. J. Cairns, A. Danek, N. Graff-Radford, B. Ghetti, and J. C. Morris, “Genetic heterogeneity in alzheimer disease and implications for treatment strategies,” *Current Neurology and Neuroscience Reports*, vol. 14, p. 499, Sep 2014.
- [406] Q. YT, S. AP, C. K, and et al, “Brain imaging and blood biomarker abnormalities in children with autosomal dominant alzheimer disease: A cross-sectional study,” *JAMA Neurology*, vol. 72, no. 8, pp. 912–919, 2015.
- [407] G. Brown, J. Wyatt, R. Harris, and X. Yao, “Diversity creation methods: a survey and categorisation,” *Information Fusion*, vol. 6, pp. 5–20, mar 2005.
- [408] Y. Cao, T. A. Geddes, J. Y. H. Yang, and P. Yang, “Ensemble deep learning in bioinformatics,” *Nature Machine Intelligence*, vol. 2, pp. 500–508, aug 2020.
- [409] L. Breiman, “Bagging predictors,” *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [410] E. Varol, B. Gaonkar, G. Erus, R. Schultz, and C. Davatzikos, “Feature ranking based nested support vector machine ensemble for medical image classification,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, may 2012.
- [411] A. Savio and M. Graña, “An ensemble of classifiers guided by the AAL brain atlas for alzheimer’s disease detection,” in *Advances in Computational Intelligence*, pp. 107–114, Springer Berlin Heidelberg, 2013.
- [412] L. Nanni, C. Salvatore, A. Cerasa, and I. Castiglioni, “Combining multiple approaches for the early diagnosis of alzheimer’s disease,” *Pattern Recognition Letters*, vol. 84, pp. 259–266, dec 2016.

- [413] F. Segovia, J. M. Górriz, J. Ramírez, C. Phillips, and . for the Alzheimer's Disease Neuroimaging, "Combining feature extraction methods to assist the diagnosis of alzheimer's disease," *Current Alzheimer Research*, vol. 13, pp. 831–837, may 2016.
- [414] A. Lebedev, E. Westman, G. V. Westen, M. Kramberger, A. Lunder-vold, D. Aarsland, H. Soininen, I. Kłoszewska, P. Mecocci, M. Tsola-ki, B. Vellas, S. Lovestone, and A. Simmons, "Random forest ensembles for detection and prediction of alzheimer's disease with a good between-cohort robustness," *NeuroImage: Clinical*, vol. 6, pp. 115–125, 2014.
- [415] J. Rodriguez, J. Ramirez, J. M. Gorriz, P. Padilla, and A. Ortiz, "Short-term MCI-to-AD prediction using MRI, neuropsychological scores and ensemble tree learning techniques," in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, oct 2015.
- [416] M. Fratello, G. Caiazzo, F. Trojsi, A. Russo, G. Tedeschi, R. Tagliaferri, and F. Esposito, "Multi-view ensemble classification of brain connectivity images for neurodegeneration type discrimination," *Neuroinformatics*, vol. 15, pp. 199–213, feb 2017.
- [417] R. Armananzas, M. Iglesias, D. A. Morales, and L. Alonso-Nanclares, "Voxel-based diagnosis of alzheimer's disease using classifier ensembles," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 778–784, may 2017.
- [418] A. K. Patra, R. Ray, A. A. Abdullah, and S. R. Dash, "Prediction of parkinson's disease using ensemble machine learning classification from acoustic analysis," *Journal of Physics: Conference Series*, vol. 1372, p. 012041, nov 2019.
- [419] M. Cesari, J. A. Christensen, M.-L. Muntean, B. Mollenhauer, F. Sixel-Döring, H. B. Sorensen, C. Trenkwalder, and P. Jennum, "A data-driven system to identify REM sleep behavior disorder and to predict its progression from the prodromal stage in parkinson's disease," *Sleep Medicine*, apr 2020.
- [420] Z. Barzegar and M. Jamzad, "A reliable ensemble-based classification framework for glioma brain tumor segmentation," *Signal, Image and Video Processing*, vol. 14, pp. 1591–1599, may 2020.
- [421] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis*, vol. 38, pp. 367–378, feb 2002.

- [422] P. Bühlmann and T. Hothorn, “Boosting algorithms: Regularization, prediction and model fitting,” *Statist. Sci.*, vol. 22, pp. 477–505, 11 2007.
- [423] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, “Evaluation of handwriting kinematics and pressure for differential diagnosis of parkinson’s disease,” *Artificial Intelligence in Medicine*, vol. 67, pp. 39–46, feb 2016.
- [424] L. Nanni, A. Lumini, and N. Zaffonato, “Ensemble based on static classifier selection for automated diagnosis of mild cognitive impairment,” *Journal of Neuroscience Methods*, vol. 302, pp. 42–46, may 2018.
- [425] M. Jin and W. Deng, “Predication of different stages of alzheimer’s disease using neighborhood component analysis and ensemble decision tree,” *Journal of Neuroscience Methods*, vol. 302, pp. 35–41, may 2018.
- [426] B. Braithwaite, J. Paananen, H. Taipale, A. Tanskanen, J. Tiihonen, S. Hartikainen, and A.-M. Tolppanen, “Detection of medications associated with alzheimer’s disease using ensemble methods and cooperative game theory,” *International Journal of Medical Informatics*, vol. 141, p. 104142, sep 2020.
- [427] J. Kim, Y. Park, S. Park, H. Jang, H. J. Kim, D. L. Na, H. Lee, and S. W. Seo, “Prediction of tau accumulation in prodromal alzheimer’s disease using an ensemble machine learning approach,” *Scientific Reports*, vol. 11, mar 2021.
- [428] M. M. Khan, S. K. Chalup, and A. Mendes, “Parkinson’s disease data classification using evolvable wavelet neural networks,” in *Lecture Notes in Computer Science*, pp. 113–124, Springer International Publishing, 2016.
- [429] J. Islam and Y. Zhang, “Brain MRI analysis for alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks,” *Brain Informatics*, vol. 5, may 2018.
- [430] R. Wang, H. Li, R. Lan, S. Luo, and X. Luo, “Hierarchical ensemble learning for alzheimer’s disease classification,” in *2018 7th International Conference on Digital Home (ICDH)*, IEEE, nov 2018.
- [431] Y. Chen, H. Jia, Z. Huang, and Y. Xia, “Early identification of alzheimer’s disease using an ensemble of 3d convolutional neural networks and magnetic resonance imaging,” in *Advances in Brain Inspired Cognitive Systems*, pp. 303–311, Springer International Publishing, 2018.

- [432] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. KIM, G.-R. Kwon, K. H. Lee, and H. Y. Jung, “Ensembles of patch-based classifiers for diagnosis of alzheimer diseases,” *IEEE Access*, vol. 7, pp. 73373–73383, 2019.
- [433] S. Ahmed, B. C. Kim, K. H. Lee, and H. Y. J. and, “Ensemble of ROI-based convolutional neural network classifiers for staging the alzheimer disease spectrum from magnetic resonance imaging,” *PLOS ONE*, vol. 15, p. e0242712, dec 2020.
- [434] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, “Early detection of alzheimer’s disease using magnetic resonance imaging: A novel approach combining convolutional neural networks and ensemble learning,” *Frontiers in Neuroscience*, vol. 14, may 2020.
- [435] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, “Volumetric feature-based alzheimer’s disease diagnosis from sMRI data using a convolutional neural network and a deep neural network,” *IEEE Access*, vol. 9, pp. 29870–29882, 2021.
- [436] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, “Ensembles of deep learning architectures for the early diagnosis of the alzheimer’s disease,” *International Journal of Neural Systems*, vol. 26, p. 1650025, aug 2016.
- [437] E. Jabason, M. O. Ahmad, and M. Swamy, “Classification of alzheimer’s disease from MRI data using an ensemble of hybrid deep convolutional neural networks,” in *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, aug 2019.
- [438] L. Nanni, M. Interlenghi, S. Brahnám, C. Salvatore, S. Papa, R. Nenni, and I. C. and, “Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of alzheimer’s disease,” *Frontiers in Neurology*, vol. 11, nov 2020.
- [439] J. Ruiz, M. Mahmud, M. Modasshir, M. S. Kaiser, and for the Alzheimer’s Disease Neuroimaging In, “3d DenseNet ensemble in 4-way classification of alzheimer’s disease,” in *Brain Informatics*, pp. 85–96, Springer International Publishing, 2020.
- [440] W. M. Association, “World medical association declaration of helsinki,” *JAMA*, vol. 310, p. 2191, nov 2013.
- [441] K. Marek, S. Chowdhury, A. Siderowf, S. Lasch, C. S. Coffey, C. Caspell-Garcia, T. Simuni, D. Jennings, C. M. Tanner, J. Q. Trojanowski, *et al.*, “The parkinson’s progression markers initiative (PPMI) - establishing a PD biomarker cohort,” *Annals of Clinical and Translational Neurology*, vol. 5, pp. 1460–1477, oct 2018.

- [442] R. V. Muñoz, F. Florence, C. Pierrick, and D. Michel, “No structural differences are revealed by vbm in “de novo” parkinsonian patients,” *Studies in Health Technology and Informatics*, vol. 264, no. MEDINFO 2019: Health and Wellbeing e-Networks for All, pp. 268–272, 2019.
- [443] K. Wyman-Chick, P. Martin, M. Minár, and R. Schroeder, “Cognition in patients with a clinical diagnosis of parkinson disease and scans without evidence of dopaminergic deficit (swedd): 2-year follow-up,” *Cognitive and Behavioral Neurology*, vol. 29, pp. 190–196, dec 2016.
- [444] J. C. Morris, P. S. Aisen, R. J. Bateman, T. L. S. Benzinger, N. J. Cairns, A. M. Fagan, B. Ghetti, A. M. Goate, D. M. Holtzman, W. E. Klunk, E. McDade, D. S. Marcus, R. N. Martins, C. L. Masters, R. Mayeux, A. Oliver, K. Quaid, J. M. Ringman, M. N. Rossor, S. Salloway, P. R. Schofield, N. J. Selsor, R. A. Sperling, M. W. Weiner, C. Xiong, K. L. Moulder, and V. D. Buckles, “Developing an international network for alzheimer research: The dominantly inherited alzheimer network,” *Clinical investigation*, vol. 2, pp. 975–984, Oct 2012.
- [445] F. Roselli, B. Tartaglione, F. Federico, V. Lepore, G. Defazio, and P. Livrea, “Rate of MMSE score change in alzheimer’s disease: Influence of education and vascular risk factors,” *Clinical Neurology and Neurosurgery*, vol. 111, pp. 327–330, may 2009.
- [446] A. S. Fauci, E. Braunwald, D. L. Kasper, S. L. Hauser, D. L. Longo, J. L. Jameson, and J. Loscalzo, *Harrison’s Principles of Internal Medicine (Spanish Edition)*, vol. 2. McGraw-Hill, 17 ed., 2008.
- [447] L. Berg, “Clinical dementia rating (cdr),” *Psychopharmacology Bulletin*, vol. 24, pp. 637–639, 1988.
- [448] J. C. Morris, “Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type,” *International Psychogeriatrics*, vol. 9, no. S1, pp. 173–176, 1997.
- [449] A. M. Fagan, C. Xiong, M. S. Jaszec, R. J. Bateman, A. M. Goate, T. L. S. Benzinger, B. Ghetti, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, S. Salloway, P. R. Schofield, R. A. Sperling, D. Marcus, N. J. Cairns, V. D. Buckles, J. H. Ladenson, J. C. Morris, D. M. Holtzman, *et al.*, “Longitudinal change in csf biomarkers in autosomal-dominant alzheimer’s disease,” *Science Translational Medicine*, vol. 6, pp. 226ra30–226ra30, mar 2014.
- [450] L. Yan, C. Y. Liu, K.-P. Wong, S.-C. Huang, W. J. Mack, K. Jann, G. Coppola, J. M. Ringman, and D. J. Wang, “Regional association of pcal-mri with fdg-pet and pib-pet in people at risk for autosomal

- dominant alzheimer's disease," *NeuroImage: Clinical*, vol. 17, pp. 751 – 760, 2018.
- [451] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski, "Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade," *The Lancet Neurology*, vol. 9, no. 1, pp. 119 – 128, 2010.
- [452] M. Brant-Zawadzki, G. D. Gillan, and W. R. Nitz, "MP RAGE: a three-dimensional, t1-weighted, gradient-echo sequence—initial experience in the brain.," *Radiology*, vol. 182, pp. 769–775, mar 1992.
- [453] B. Fischl, M. I. Sereno, R. B. Tootell, and A. M. Dale, "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Human Brain Mapping*, vol. 8, no. 4, pp. 272–284, 1999.
- [454] B. Fischl, D. H. Salat, A. J. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale, "Sequence-independent segmentation of magnetic resonance images," *NeuroImage*, vol. 23, pp. 69–84, jan 2004.
- [455] J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. MacFall, B. Fischl, and A. Dale, "Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data," *NeuroImage*, vol. 30, pp. 436–443, apr 2006.
- [456] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *NeuroImage*, vol. 61, no. 4, pp. 1402–1418, 2012.
- [457] D. J. Rowland, J. R. Garbow, R. Laforest, and A. Z. Snyder, "Registration of [18f]FDG microPET and small-animal MRI," *Nuclear Medicine and Biology*, vol. 32, pp. 567–572, aug 2005.
- [458] S. A. Eisenstein, J. M. Koller, M. Piccirillo, A. Kim, J. A. V. Antenor-Dorsey, T. O. Videen, A. Z. Snyder, M. Karimi, S. M. Moerlein, K. J. Black, J. S. Perlmutter, and T. Hershey, "Characterization of extrastriatal d2 in vivo specific binding of [18f](n-methyl)benperidol using PET," *Synapse*, vol. 66, pp. 770–780, may 2012.
- [459] T. Benzinger, A. Fagan, J. Chhatwal, J. C. Morris, M. Rossor, and R. Bateman, "Regional variability of imaging biomarkers in DIAN," *Alzheimer's & Dementia*, vol. 10, p. P127, jul 2014.
- [460] P. A. Filipek, D. N. Kennedy, and V. S. Caviness, "Volumetric analyses of central nervous system neoplasm based on MRI," *Pediatric Neurology*, vol. 7, pp. 347–351, sep 1991.

- [461] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, “Assessing the accuracy of prediction algorithms for classification: an overview,” *Bioinformatics*, vol. 16, pp. 412–424, may 2000.
- [462] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, pp. 861–874, jun 2006.
- [463] P. I. Good, *Resampling Methods*. Springer Basel AG, dec 2006.
- [464] H. Fofack and J. P. Nolan, “Tail behavior, modes and other characteristics of stable distributions,” *Extremes*, vol. 2, no. 1, pp. 39–58, 2000.
- [465] J. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez, “GMM based SPECT image classification for the diagnosis of alzheimer’s disease,” *Applied Soft Computing*, vol. 11, pp. 2313–2325, mar 2011.
- [466] D. M. H. Nguyen, H. T. Vu, H. Q. Ung, and B. T. Nguyen, “3d-brain segmentation using deep neural network and gaussian mixture model,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, mar 2017.
- [467] M. Wiper, D. R. Insua, and F. Ruggeri, “Mixtures of gamma distributions with applications,” *Journal of Computational and Graphical Statistics*, vol. 10, pp. 440–454, sep 2001.
- [468] E. G. Tsionas, “Bayesian analysis of finite mixtures of weibull distributions,” *Communications in Statistics - Theory and Methods*, vol. 31, pp. 37–48, jan 2002.
- [469] C. Fernández and P. J. Green, “Modelling spatially correlated data via mixtures: A bayesian approach,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 805–826, 2002.
- [470] D. Salas-Gonzalez, E. E. Kuruoglu, and D. P. Ruiz, “Finite mixture of α -stable distributions,” *Digital Signal Processing*, vol. 19, no. 2, pp. 250–264, 2009.
- [471] D. Salas-Gonzalez, E. E. Kuruoglu, and D. P. Ruiz, “Modelling with mixture of symmetric stable distributions using gibbs sampling,” *Signal Processing*, vol. 90, no. 3, pp. 774–783, 2010.
- [472] D. P. Geoffrey J. McLachlan, *Finite Mixture Models*. Wiley, 10 2000.
- [473] G. J. McLachlan and S. Rathnayake, “On the number of components in a gaussian mixture model,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, pp. 341–355, sep 2014.

- [474] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, dec 1974.
- [475] G. Schwarz *et al.*, “Estimating the dimension of a model,” *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [476] J. P. Nolan, “Maximum likelihood estimation and diagnostics for stable distributions,” in *Lévy Processes*, pp. 379–400, Birkhäuser Boston, 2001.
- [477] I. A. Koutrouvelis, “Regression-type estimation of the parameters of stable laws,” *Journal of the American Statistical Association*, vol. 75, no. 372, pp. 918–928, 1980.
- [478] I. A. Koutrouvelis, “An iterative procedure for the estimation of the parameters of stable laws,” *Communications in Statistics - Simulation and Computation*, vol. 10, no. 1, pp. 17–28, 1981.
- [479] J. H. McCulloch, “Simple consistent estimators of stable distribution parameters,” *Communications in Statistics - Simulation and Computation*, vol. 15, no. 4, pp. 1109–1136, 1986.
- [480] D. Normile, “Coronavirus cases have dropped sharply in South Korea. What’s the secret to its success?,” *Science*, mar 2020.
- [481] European Centre For Disease Prevention And Control, “Sonraí faoi choróinvíreas COVID-19,” 2020.
- [482] L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology*, vol. 26, pp. 297–302, jul 1945.
- [483] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, I. A. Illán, P. Padilla, F. J. Martínez-Murcia, and E. W. Lang, “Building a FP-CIT SPECT brain template using a posterization approach,” *Neuroinformatics*, vol. 13, pp. 391–402, mar 2015.
- [484] I. Sobel and G. Feldman, “A 3x3 isotropic gradient operator for image processing,” *Pattern Classification and Scene Analysis*, pp. 271–272, 1973.
- [485] C.-C. Chang and C.-J. Lin, “LIBSVM,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, apr 2011.
- [486] M. Imre, J. Tao, and C. Wang, “Identifying nearly equally spaced isosurfaces for volumetric data sets,” *Computers & Graphics*, vol. 72, pp. 82–97, 2018.

- [487] T. Lewiner, H. Lopes, A. W. Vieira, and G. Tavares, “Efficient implementation of marching cubes’ cases with topological guarantees,” *Journal of Graphics Tools*, vol. 8, no. 2, pp. 1–15, 2003.
- [488] W. E. Lorensen and H. E. Cline, “Marching cubes: A high resolution 3d surface construction algorithm,” in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH ’87, (New York, NY, USA), pp. 163–169, ACM, 1987.
- [489] E. V. Chernyaev, “Marching cubes 33: Construction of topologically correct isosurfaces. technical report cern cn 95-17,” tech. rep., CERN, 1995. Presented at GRAPHI CON ’95, Saint-Petersburg, Russia, 3-7 July, 1995.
- [490] B. H. Roune and E. S. de Cabezón, “Complexity and algorithms for euler characteristic of simplicial complexes,” *Journal of Symbolic Computation*, vol. 50, pp. 170–196, mar 2013.
- [491] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, “Cross-validation pitfalls when selecting and assessing regression and classification models,” *Journal of Cheminformatics*, vol. 6, mar 2014.
- [492] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of machine learning research*, vol. 13, no. Feb, pp. 281–305, 2012.
- [493] T. F. Chan, G. H. Golub, and R. J. LeVeque, “Updating formulae and a pairwise algorithm for computing sample variances,” in *COMPSTAT 1982 5th Symposium held at Toulouse 1982*, pp. 30–41, Springer, 1982.
- [494] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML’15, p. 448–456, JMLR.org, 2015.
- [495] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” in *NIPS 2014 Workshop on Deep Learning, December 2014*, 2014.
- [496] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” jul 2012.
- [497] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

- [498] L. Prechelt, “Early stopping - but when?,” in *Lecture Notes in Computer Science*, pp. 55–69, Springer Berlin Heidelberg, 1998.
- [499] A. S. Fleisher, K. Chen, Y. T. Quiroz, L. J. Jakimovich, M. G. Gomez, C. M. Langois, J. B. Langbaum, N. Ayutyanont, A. Roontiva, P. Thiyyagura, W. Lee, H. Mo, L. Lopez, S. Moreno, N. Acosta-Baena, M. Giraldo, G. Garcia, R. A. Reiman, M. J. Huentelman, K. S. Kosik, P. N. Tariot, F. Lopera, and E. M. Reiman, “Florbetapir pet analysis of amyloid- β deposition in the presenilin 1 e280a autosomal dominant alzheimer’s disease kindred: a cross-sectional study,” *The Lancet Neurology*, vol. 11, no. 12, pp. 1057–1065, 2012.
- [500] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, and J. Q. Trojanowski, “Tracking pathophysiological processes in alzheimer’s disease: an updated hypothetical model of dynamic biomarkers,” *The Lancet Neurology*, vol. 12, no. 2, pp. 207–216, 2013.
- [501] P. H. St George-Hyslop, J. L. Haines, L. A. Farrer, R. Polinsky, C. V. Broeckhoven, A. Goate, D. R. C. McLachlan, H. Orr, A. C. Bruni, S. Sorbi, I. Rainero, J. F. Foncin, D. Pollen, J.-M. Cantu, R. Tupper, N. Voskresenskaya, R. Mayeux, J. Growdon, V. A. Fried, R. H. Myers, L. Nee, H. Backhovens, J.-J. Martin, M. Rossor, M. J. Owen, M. Mullan, M. E. Percy, H. Karlinsky, S. Rich, L. Heston, M. Montesi, M. Mortilla, N. Nacmias, J. F. Gusella, and J. A. Hardy, “Genetic linkage studies suggest that alzheimer’s disease is not a single homogeneous disorder,” *Nature*, vol. 347, p. 194, 09 1990.
- [502] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan, and N. Silverberg, “NIA-AA research framework: Toward a biological definition of alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 14, pp. 535–562, apr 2018.
- [503] N. S. Ryan, J. M. Nicholas, P. S. J. Weston, Y. Liang, T. Lashley, R. Guerreiro, G. Adamson, J. Kenny, J. Beck, L. Chavez-Gutierrez, B. de Strooper, T. Revesz, J. Holton, S. Mead, M. N. Rossor, and N. C. Fox, “Clinical phenotype and genetic associations in autosomal dominant familial alzheimer’s disease: a case series,” *The Lancet Neurology*, vol. 15, no. 13, pp. 1326–1335, 2016.
- [504] R. I. Schill, G. R. Ridgway, J. W. Bartlett, J. Barnes, N. S. Ryan, S. Mead, J. Beck, M. J. Clarkson, S. J. Crutch, J. M. Schott, S. Our-

- selin, J. D. Warren, J. Hardy, M. N. Rossor, and N. C. Fox, “Genetic influences on atrophy patterns in familial Alzheimer’s disease: a comparison of APP and PSEN1 mutations,” *Journal of Alzheimer’s Disease*, vol. 35, no. 1, pp. 199–212, 2013.
- [505] N. S. Ryan, G.-J. Biessels, L. Kim, J. M. Nicholas, P. A. Barber, P. Walsh, P. Gami, H. R. Morris, A. J. Bastos-Leite, J. M. Schott, J. Beck, S. Mead, L. Chavez-Gutierrez, B. de Strooper, M. N. Rossor, T. Revesz, T. Lashley, and N. C. Fox, “Genetic determinants of white matter hyperintensities and amyloid angiopathy in familial alzheimer’s disease,” *Neurobiology of Aging*, vol. 36, pp. 3140–3151, dec 2015.
- [506] M. Suarez-Calvet, M. A. A. Caballero, G. Kleinberger, R. J. Bateman, A. M. Fagan, J. C. Morris, J. Levin, A. Danek, M. Ewers, and C. Haass, “Early changes in CSF sTREM2 in dominantly inherited alzheimers disease occur after amyloid deposition and neuronal injury,” *Science Translational Medicine*, vol. 8, pp. 369ra178–369ra178, dec 2016.
- [507] G. Di Fede, M. Catania, E. Maderna, R. Ghidoni, L. Benussi, E. Tonoli, G. Giaccone, F. Moda, A. Paterlini, I. Campagnani, S. Sorrentino, L. Colombo, A. Kubis, E. Bistaffa, B. Ghetti, and F. Tagliavini, “Molecular subtypes of Alzheimer’s disease,” *Scientific reports.*, vol. 8, no. 1, 2018.
- [508] C. A. Lemere, F. Lopera, K. S. Kosik, C. L. Lendon, J. Ossa, T. C. Saido, H. Yamaguchi, A. Ruiz, A. Martinez, L. Madrigal, L. Hincapie, J. C. A. L., D. C. Anthony, E. H. Koo, A. M. Goate, D. J. Selkoe, and J. C. A. V., “The e280a presenilin 1 alzheimer mutation produces increased $a\beta_{42}$ deposition and severe cerebellar pathology,” *Nature Medicine*, vol. 2, pp. 1146–1150, oct 1996.
- [509] M. N. Haque, N. Noman, R. Berretta, and P. Moscato, “Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification,” *PLOS ONE*, vol. 11, p. e0146116, jan 2016.
- [510] “Kolmogorov-smirnov test,” in *The Concise Encyclopedia of Statistics*, pp. 283–287, Springer New York, 2008.
- [511] J. M. Górriz, J. Ramirez, and J. Suckling, “On the computation of distribution-free performance bounds: Application to small sample sizes in neuroimaging,” *Pattern Recognition*, vol. 93, pp. 1–13, 2019.
- [512] J. Macqueen, “Some methods for classification and analysis of multivariate observations,” in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

- [513] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A k-means clustering algorithm,” *Applied Statistics*, vol. 28, no. 1, p. 100, 1979.
- [514] R. J. Bateman, P. S. Aisen, B. D. Strooper, N. C. Fox, C. A. Lemere, J. M. Ringman, S. Salloway, R. A. Sperling, M. Windisch, and C. Xiong, “Autosomal-dominant alzheimer’s disease: a review and proposal for the prevention of alzheimer’s disease,” *Alzheimer’s Research & Therapy*, vol. 3, no. 1, p. 1, 2010.
- [515] E. E. Tripoliti, D. I. Fotiadis, and G. Manis, “Automated diagnosis of diseases based on classification: Dynamic determination of the number of trees in random forests algorithm,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 615–622, jul 2012.
- [516] X. Guo, X. Wang, and H. Ling, “Exclusivity regularized machine: A new ensemble SVM classifier,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence Organization, aug 2017.
- [517] B. Lei, P. Yang, Y. Zhuo, F. Zhou, D. Ni, S. Chen, X. Xiao, and T. Wang, “Neuroimaging retrieval via adaptive ensemble manifold learning for brain disease diagnosis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 1661–1673, jul 2019.
- [518] N. An, H. Ding, J. Yang, R. Au, and T. F. Ang, “Deep ensemble learning for alzheimer’s disease classification,” *Journal of Biomedical Informatics*, vol. 105, p. 103411, may 2020.
- [519] G. Maguolo, L. Nanni, and S. Ghidoni, “Ensemble of convolutional neural networks trained with different activation functions,” *Expert Systems with Applications*, vol. 166, p. 114048, mar 2021.
- [520] A. Ebadi, J. L. D. da Rocha, D. B. Nagaraju, F. Tovar-Moll, I. Bramati, G. Coutinho, R. Sitaram, and P. Rashidi, “Ensemble classification of alzheimer’s disease and mild cognitive impairment based on complex graph measures from diffusion tensor images,” *Frontiers in Neuroscience*, vol. 11, feb 2017.
- [521] D. W. Opitz, , and J. W. Shavlik, “Actively searching for an effective neural network ensemble,” *Connection Science*, vol. 8, no. 3-4, pp. 337–354, 1996.
- [522] J. A. Santiago and J. A. Potashkin, “Evaluation of rna blood biomarkers in individuals at risk of parkinson’s disease,” *Journal of Parkinson’s Disease*, vol. 7, pp. 653–660, Nov. 2017.
- [523] P. Good, *Permutation Tests: a Practical Guide to Resampling Methods for Testing Hypotheses*. New York, NY: Springer New York, 1994.

- [524] H. Bhaskar, L. Mihaylova, and A. Achim, “Video foreground detection based on symmetric alpha-stable mixture models,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 1133–1138, aug 2010.
- [525] Y. C. Bechtel, C. Bonaiti-Pellie, N. Poisson, J. Magnette, and P. R. Bechtel, “A population and family study of n-acetyltransferase using caffeine urinary metabolites,” *Clinical pharmacology and therapeutics*, vol. 54, no. 2, pp. 134–141, 1993.
- [526] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Röcken, R. E. Nutt, H.-J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler, “Simultaneous PET-MRI: a new approach for functional and morphological imaging,” *Nature Medicine*, vol. 14, pp. 459–465, mar 2008.
- [527] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, A.-M. Dagonowski, M. Ernst, D. Fair, M. Hampson, M. J. Hoptman, J. S. Hyde, V. J. Kiviniemi, R. Kotter, S.-J. Li, C.-P. Lin, M. J. Lowe, C. Mackay, D. J. Madden, K. H. Madsen, D. S. Margulies, H. S. Mayberg, K. McMahon, C. S. Monk, S. H. Mostofsky, B. J. Nagel, J. J. Pekar, S. J. Peltier, S. E. Petersen, V. Riedl, S. A. R. B. Rombouts, B. Rypma, B. L. Schlaggar, S. Schmidt, R. D. Seidler, G. J. Siegle, C. Sorg, G.-J. Teng, J. Veijola, A. Villringer, M. Walter, L. Wang, X.-C. Weng, S. Whitfield-Gabrieli, P. Williamson, C. Windischberger, Y.-F. Zang, H.-Y. Zhang, F. X. Castellanos, and M. P. Milham, “Toward discovery science of human brain function,” *Proceedings of the National Academy of Sciences*, vol. 107, pp. 4734–4739, feb 2010.
- [528] C. R. Jack, “Alzheimer disease: New concepts on its neurobiology and the clinical role imaging will play,” *Radiology*, vol. 263, pp. 344–361, may 2012.
- [529] C. Salvatore, I. Castiglioni, and A. Cerasa, “Radiomics approach in the neurodegenerative brain,” *Aging Clinical and Experimental Research*, aug 2019.
- [530] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, “High-accuracy classification of parkinson’s disease through shape analysis and surface fitting in 123i-ioflupane SPECT imaging,” *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 794–802, may 2017.
- [531] M. Martínez-Ibanez, A. Ortiz, J. Munilla, D. Salas-Gonzalez, J. M. Górriz, and J. Ramírez, “Isosurface modelling of DatSCAN images

- for parkinson disease diagnosis,” in *Understanding the Brain Function and Emotions*, pp. 360–368, Springer International Publishing, 2019.
- [532] M. C. A. Santos, L. S. Campos, R. P. Guimarães, C. C. Piccinin, P. C. Azevedo, L. G. Piovesana, B. M. D. Campos, A. C. S. Amato-Filho, F. Cendes, and A. D’Abreu, “Does side of onset influence the pattern of cerebral atrophy in parkinson’s disease?,” *Frontiers in Neurology*, vol. 7, sep 2016.
- [533] D. Salat, A. J. Noyce, A. Schrag, and E. Tolosa, “Challenges of modifying disease progression in prediagnostic parkinson’s disease,” *The Lancet Neurology*, vol. 15, pp. 637–648, may 2016.
- [534] B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, K. Broich, E. Cavado, S. Crutch, J.-F. Dartigues, C. Duyckaerts, S. Epelbaum, G. B. Frisoni, S. Gauthier, R. Genthon, A. A. Gouw, M.-O. Habert, D. M. Holtzman, M. Kivipelto, S. Lista, J.-L. Molinuevo, S. E. O’Bryant, G. D. Rabinovici, C. Rowe, S. Salloway, L. S. Schneider, R. Sperling, M. Teichmann, M. C. Carrillo, J. Cummings, and C. R. Jack, “Preclinical alzheimer’s disease: Definition, natural history, and diagnostic criteria,” *Alzheimer’s & Dementia*, vol. 12, pp. 292–323, mar 2016.
- [535] P. H. Lockett, A. McCullough, B. A. Gordon, J. Strain, S. Flores, A. Dincer, J. McCarthy, T. Kuffner, A. Stern, K. L. Meeker, S. B. Berman, J. P. Chhatwal, C. Cruchaga, A. M. Fagan, M. R. Farlow, N. C. Fox, M. Jucker, J. Levin, C. L. Masters, H. Mori, J. M. Noble, S. Salloway, P. R. Schofield, A. M. Brickman, W. S. Brooks, D. M. Cash, M. J. Fulham, B. Ghetti, C. R. Jack, J. Vöglein, W. Klunk, R. Koeppe, H. Oh, Y. Su, M. Weiner, Q. Wang, L. Swisher, D. Marcus, D. Koudelis, N. Joseph-Mathurin, L. Cash, R. Hornbeck, C. Xiong, R. J. Perrin, C. M. Karch, J. Hassenstab, E. McDade, J. C. Morris, T. L. Benzinger, R. J. Bateman, and B. M. A. and, “Modeling autosomal dominant alzheimer’s disease with machine learning,” *Alzheimer’s & Dementia*, jan 2021.
- [536] S. Lee, F. Viqar, M. E. Zimmerman, A. Narkhede, G. Tosto, T. L. Benzinger, D. S. Marcus, A. M. Fagan, A. Goate, N. C. Fox, N. J. Cairns, D. M. Holtzman, V. Buckles, B. Ghetti, E. McDade, R. N. Martins, A. J. Saykin, C. L. Masters, J. M. Ringman, N. S. Ryan, S. Förster, C. Laske, P. R. Schofield, R. A. Sperling, S. Salloway, S. Correia, C. Jack, M. Weiner, R. J. Bateman, J. C. Morris, R. Mayeux, and A. M. B. and, “White matter hyperintensities are a core feature of alzheimer’s disease: Evidence from the dominantly inherited alzheimer network,” *Annals of Neurology*, vol. 79, pp. 929–939, apr 2016.

- [537] J. Meireles and J. Massano, “Cognitive impairment and dementia in parkinson’s disease: Clinical features, diagnosis, and management,” *Frontiers in Neurology*, vol. 3, p. 88, 2012.
- [538] A. Klettner, A. Tholey, A. Wiegandt, E. Richert, B. Nölle, G. Deuschl, J. Roider, and S. A. Schneider, “Reduction of gapdh in lenses of parkinson’s disease patients: A possible new biomarker,” *Movement Disorders*, vol. 32, no. 3, pp. 459–462, 2016.
- [539] C.-Y. Xu, W.-Y. Kang, Y.-M. Chen, T.-F. Jiang, J. Zhang, L.-N. Zhang, J.-Q. Ding, J. Liu, and S.-D. Chen, “DJ-1 inhibits α -synuclein aggregation by regulating chaperone-mediated autophagy,” *Frontiers in Aging Neuroscience*, vol. 9, sep 2017.
- [540] B. C. Vanle, V. R. Florang, D. J. Murry, A. L. Aguirre, and J. A. Doorn, “Inactivation of glyceraldehyde-3-phosphate dehydrogenase by the dopamine metabolite, 3,4-dihydroxyphenylacetaldehyde,” *Biochemical and Biophysical Research Communications*, vol. 492, pp. 275–281, oct 2017.
- [541] M. Grassi, N. Rouleaux, D. Caldirola, D. Loewenstein, K. Schruers, G. Perna, and M. D. and, “A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to alzheimer’s disease using socio-demographic characteristics, clinical information, and neuropsychological measures,” *Frontiers in Neurology*, vol. 10, jul 2019.
- [542] E. Lella, A. Paziienza, D. Lofù, R. Anglani, and F. Vitulano, “An ensemble learning approach based on diffusion tensor imaging measures for alzheimer’s disease classification,” *Electronics*, vol. 10, p. 249, jan 2021.
- [543] H. G. Schnack and R. S. Kahn, “Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters,” *Frontiers in Psychiatry*, vol. 7, mar 2016.
- [544] H. Wang, F. Nie, H. Huang, S. L. Risacher, A. J. Saykin, and L. S. and, “Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning,” *Bioinformatics*, vol. 28, pp. i127–i136, jun 2012.
- [545] V. Parekh and M. A. Jacobs, “Radiomics: a new application from established techniques,” *Expert Review of Precision Medicine and Drug Development*, vol. 1, pp. 207–226, mar 2016.
- [546] P. A. Penczek, “Fundamentals of three-dimensional reconstruction from projections,” in *Methods in Enzymology*, pp. 1–33, Elsevier, 2010.

- [547] B. A. Ardekani, E. Bermudez, A. M. Mubeen, A. H. Bachman, and for the Alzheimer’s Disease Neuroimaging Initiative, “Prediction of incipient alzheimer’s disease dementia in patients with mild cognitive impairment,” *Journal of Alzheimer’s Disease*, vol. 55, pp. 269–281, 2016.
- [548] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, “The quickhull algorithm for convex hulls,” *ACM Transactions on Mathematical Software*, vol. 22, pp. 469–483, dec 1996.
- [549] J. H. Gilmore, R. C. Knickmeyer, and W. Gao, “Imaging structural and functional brain development in early childhood,” *Nature Reviews Neuroscience*, vol. 19, pp. 123–137, feb 2018.
- [550] R. Raeper, A. Lisowska, and I. Rekik, “Cooperative correlational and discriminative ensemble classifier learning for early dementia diagnosis using morphological brain multiplexes,” *IEEE Access*, vol. 6, pp. 43830–43839, 2018.

Abreviaturas y Siglas

| | |
|-------|--|
| ADNI | Alzheimer's Disease Neuroimaging Initiative (Institución) |
| ANOVA | Análisis de la varianza (<i>ANalysis Of VAriance</i>) |
| APOE | Gen que regula la apolipoproteína-E |
| APP | Gen que regula la proteína precursora de amiloide (<i>Amyloid Precursor Protein</i>) |
| AUC | Área bajo una curva ROC (<i>Area Under the Curve</i>) |
| BR | <i>Binding Ratio</i> (Técnica de normalización en intensidad) |
| CNN | Redes neuronales convolucionales (<i>Convolutional Neural Networks</i>) |
| CSF | Líquido ceforraquídeo o marcadores presentes en el líquido ceforraquídeo (<i>CerebroSpinal Fluid</i>) |
| DATs | Transportadores presinápticos de dopamina (<i>presynaptic Dopamine Transporters</i>) |
| DCL | Deterioro Cognitivo Leve |
| DL | Aprendizaje profundo (<i>Deep Learning</i>) |
| EA | Enfermedad de Alzheimer |
| EM | Algoritmo de maximización de la esperanza (<i>Expectation-Maximization algorithm</i>) |
| EP | Enfermedad de Parkinson |
| EYO | Años estimados hasta la aparición de los primeros síntomas de una demencia (<i>Estimated Years to Onset</i>) |
| GMM | Modelo de mezclas de distribuciones Gaussianas (<i>Gaussian Mixture Model</i>) |
| ISBR | <i>Internet Brain Segmentation Repository</i> (Repositorio) |

| | |
|-------|---|
| KNN | Algoritmo de aprendizaje supervisado en función de los k vecinos más cercanos (<i>K-Nearest Neighbor</i>) |
| LMCA | Algoritmo de Lewiner para el cálculo de isosuperficies (<i>Lewiner Marching Cubes Algorithm</i>) |
| LME | Modelos lineales mixtos (<i>Linear-Mixed Effects</i>) |
| LOO | Esquema de validación cruzada de tipo <i>Leave-One-Out</i> |
| MAP | Criterio de probabilidad máxima a posteriori (<i>Maximum A Posteriori probability</i>) |
| MC | Paciente portador de alguna mutación en los genes APP, PSEN1, PSEN2 (<i>Mutation-Carrier</i>) |
| ML | Aprendizaje máquina (<i>Machine Learning</i>) |
| MLP | Perceptrón multicapa (<i>Multi-Layer Perceptron</i>) |
| MMSE | Test neuropsicológico (<i>Mini Mental Status Examination</i>) |
| MRF | Campos aleatorios de Markov (<i>Markov Random Fields</i>) |
| MRI | Imagen por resonancia magnética (<i>Magnetic Resonance Imaging</i>) |
| MV | Voto por mayoría (<i>Majority Voting</i>) |
| NC | Paciente libre de mutaciones en los genes APP, PSEN1, PSEN2 (<i>Non-Carrier</i>) |
| NIB | Marcadores que no proceden de pruebas de imagen (<i>Non-Imaging Biomarkers</i>) |
| NN | Redes neuronales (<i>Neural Networks</i>) |
| PCA | Análisis de componentes principales (<i>Principal Component Analysis</i>) |
| PET | Tomografía por emisión de positrones (<i>Positron Emission Tomography</i>) |
| PLS | Algoritmo de los mínimos cuadrados parciales (<i>Partial Least Squares</i>) |
| PPMI | <i>Parkinson's Progression Markers Initiative</i> (Institución) |
| PSEN1 | Gen de la Presenilina-1 |
| PSEN2 | Gen de la Presenilina-2 |

| | |
|-------|---|
| RBF | Kernel de función base radial (<i>Radial Basis Function kernel</i>) |
| ReLU | <i>Rectifier Linear Unit</i> (función de activación utilizada en redes neuronales) |
| RF | Bosques aleatorios (<i>Random Forests</i>) |
| RNN | Redes neuronales recurrentes (<i>Recurrent Neural Networks</i>) |
| ROC | Curva ROC (<i>Receiver Operating Curves</i>) |
| ROI | Región de interés (<i>Region Of Interest</i>) |
| SNR | Relación señal a ruido (<i>Signal to Noise Ratio</i>) |
| SOA | Algoritmo del Operador de Sobel-Feldman (<i>Sobel-Feldman Operator Algorithm</i>) |
| SOM | Redes neuronales de mapas auto-organizados (<i>Self-Organizing Map neural networks</i>) |
| SPECT | Tomografía computarizada por emisión de fotón único (<i>Single-Photon Emission Computed Tomography</i>) |
| SPM | Statistical Parametric Mapping (<i>Software</i>) |
| SUVR | Marcadores de intensidad calculados mediante el ratio de captación normalizado (<i>Standardized Uptake Value Ratio</i>) |
| SVM | Máquinas de vectores de soporte (<i>Support Vector Machine</i>) |
| VAF | <i>Voxels As Features</i> |
| VBM | Método <i>Voxel-Based-Morphometry</i> propuesto por Ashburner et al. en [262] |
| WMV | Voto por mayoría ponderado (<i>Weighted Majority Voting</i>) |