



UNIVERSIDAD DE MÁLAGA

ESCUELA TÉCNICA SUPERIOR DE
INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

RECONOCIMIENTO DE PATRONES DE
HIPOPERFUSIÓN EN IMÁGENES PET.
APLICACIÓN AL DIAGNÓSTICO DE LA
ENFERMEDAD DE ALZHEIMER.

GRADO EN INGENIERÍA DE
TECNOLOGÍAS DE TELECOMUNICACIÓN

ANA CARRILLO TORONJO

MÁLAGA 2016

UNIVERSIDAD DE MÁLAGA
ESCUELA TÉCNICA SUPERIOR DE
INGENIERÍA DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

RECONOCIMIENTO DE PATRONES DE
HIPOPERFUSIÓN EN IMÁGENES PET. APLICACIÓN AL
DIAGNÓSTICO DE LA ENFERMEDAD DE ALZHEIMER.

GRADO EN INGENIERÍA DE
TECNOLOGÍA DE TELECOMUNICACIÓN

ANA CARRILLO TORONJO
MÁLAGA, 2016

**Reconocimiento de patrones de hipoperfusión en imágenes PET.
Aplicación al diagnóstico de la enfermedad de Alzheimer.**

Autor: Ana Carrillo Toronjo

Tutor: Andrés Ortiz García

Departamento: Ingeniería de Comunicaciones

Titulación: Grado en Ingeniería de Tecnologías de Telecomunicación

Palabras clave: Enfermedad de Alzheimer, SVM, PET, clasificador, SF.

Resumen

Este proyecto presenta un modelo estadístico para la detección de patrones de hipoperfusión en las imágenes PET (tomografía de emisión de positrones) para el diagnóstico asistido de las enfermedades neurológicas, como la enfermedad de Alzheimer. El estudio utiliza la base de datos ADNI, la cuál está constituida por 138 pacientes de ambas clases, utilizándose únicamente las regiones con mayor poder discriminativo en el diagnóstico precoz del Alzheimer. Se han utilizado dos procedimientos, uno para la extracción de las características más relevantes (SF) y el otro para la clasificación de dichas características basado en máquinas de soporte vectorial (SVM). Finalmente, se ha evaluado el rendimiento del sistema para distintos diseños, obteniendo resultados de hasta el 90% de precisión en la clasificación de pacientes controles y pacientes con Alzheimer.

**Recognition of hypoperfusion patterns in PET images.
Application to the diagnosis of Alzheimer's disease.**

Author: Ana Carrillo Toronjo

Supervisor: Andrés Ortiz García

Department: Ingeniería de Comunicaciones

Degree: Grado en Ingeniería de Tecnologías de Telecomunicación

Keywords: Alzheimer's Disease, SVM, PET, classifier, SF.

Abstract

This project presents an statistical model for the detection of hypoperfusion patterns in PET images (positron emission tomography) for the assisted diagnosis of neurological diseases, such as Alzheimer's disease. The study uses the ADNI database, which consists of 138 patients of both classes, however only regions with the greatest discriminative power in the early diagnosis of Alzheimer's have been used. Two methods have been designed, one for the extraction of the most relevant characteristics (SF) and the other for the classification of these features based on vector support machines (SVM). Finally, the performance of the system has been evaluated for different designs, obtaining results of up to 90% accuracy in the classification of control patients and patients with Alzheimer's disease.

Agradecimientos

Este trabajo ha supuesto un reto personal que no hubiera sido posible sin las personas que día a día me han apoyado.

Quisiera dar las gracias a mi familia, en especial a mis padres y a mi hermana por ayudarme a continuar, a mi pareja por su confianza, su apoyo y sus ganas de sacar lo mejor de mí.

Por último, a mi tutor, Andrés Ortiz, por su paciencia, cercanía y por ayudarme a comenzar este camino, teniendo siempre un hueco para atenderme.

*Solo los que se atreven a ir demasiado lejos,
se dan cuenta de lo lejos que pueden llegar*

Thomas Stearns Eliot

Índice

| | |
|---|-----------|
| Capítulo 1. Introducción..... | 1 |
| 1.1. Motivación | 1 |
| 1.2. Descripción breve y objetivos | 2 |
| 1.3. Plan de trabajo | 3 |
| 1.4. Estructura del trabajo..... | 3 |
| Capítulo 2. Enfermedad de Alzheimer | 5 |
| 2.1. Síntomas | 5 |
| 2.2. Diagnóstico..... | 6 |
| 2.2.1. Examen del Estado Mental Mínimo | 7 |
| 2.2.2. Examen de deterioro global..... | 7 |
| 2.3. Tratamiento..... | 8 |
| 2.4. Cambios neuroestructurales provocados por la enfermedad..... | 8 |
| Capítulo 3. Base de datos ADNI | 11 |
| 3.1. Imágenes médicas..... | 11 |
| 3.2. Adquisición de imágenes PET..... | 12 |
| 3.2.1. Radiofármacos..... | 12 |
| 3.2.2. Funcionamiento de las cámaras PET..... | 13 |
| Capítulo 4: Métodos. Caracterización y Clasificación..... | 15 |
| 4.1. Métodos de aprendizaje automático | 15 |
| 4.2. Sparse Filtering..... | 15 |
| 4.2.1. Escasez | 16 |
| 4.2.2. Aplicación de la escasez en Sparse Filtering..... | 17 |
| 4.2.3. Implementación de sparse filtering | 17 |
| 4.2.4. Aprendizaje de la distribución de características..... | 20 |
| 4.3. Support Vector Machine (SVM)..... | 21 |

| | |
|--|-----------|
| 4.3.1. SVM lineales con clases linealmente separables..... | 22 |
| 4.3.2. SVM lineales con clases no separables | 24 |
| 4.3.3. El Truco del Kernel | 26 |
| 4.4. Validación Cruzada..... | 27 |
| Capítulo 5. Sistemas de Clasificación | 29 |
| 5.1. Primer Diseño..... | 29 |
| 5.1.1. Base de datos ADNI | 30 |
| 5.1.3. Selección de regiones de interés..... | 30 |
| 5.1.4. Paralelepipedo más pequeño que contiene a cada región | 31 |
| 5.1.5. División en cubos | 32 |
| 5.1.6. Sparse Filtering..... | 32 |
| 5.1.7. Support Vector Machine..... | 33 |
| 5.2. Segundo Diseño..... | 34 |
| 5.2.1. Sparse Filtering | 35 |
| 5.2.2. Support Vector Machine | 36 |
| Capítulo 6. Resultados Experimentales..... | 37 |
| 6.1. Métodos de Evaluación del Clasificador | 37 |
| 6.1.1. Precisión, sensibilidad, especificidad | 37 |
| 6.1.2 Curva ROC | 38 |
| 6.2. Evaluación del clasificador | 39 |
| 6.2.1. Primer Diseño..... | 39 |
| 6.2.2. Segundo Diseño..... | 41 |
| 6.2.3. Curva ROC | 42 |
| Capítulo 7. Conclusiones y trabajos futuros..... | 45 |
| Bibliografía | 47 |

Índice de Figuras

| | |
|---|----|
| Figura 2.1: Cerebro sin AD (izquierda) , cerebro con AD (derecha) y comparación entre ambos (abajo). | 9 |
| Figura 2.2: Corte transversal por la mitad del cerebro entre los oídos de un cerebro sano (derecha) y un cerebro con AD (izquierda). | 9 |
| Figura 3.1: Fusión de las imágenes PET y RM de un sujeto sano. | 13 |
| Figura 4.1. Descomposición del algoritmo en sus etapas constituyentes. | 18 |
| Figura 4.2. Izquierda: Separación de clases linealmente. Derecha: Hiperplanos de máxima separación..... | 22 |
| Figura 4.3: Aparición del parámetro de error ξ en la clasificación | 24 |
| Figura 4.4: Utilización del truco del Kernel para transformar el espacio de datos | 26 |
| Figura 4.5: Proceso de validación cruzada. | 27 |
| Figura 5.1: Diagrama de bloques del sistema..... | 29 |
| Figura 5.2: Izquierda, imagen PET. Derecha, atlas PET [25]..... | 30 |
| Figura 5.3: Características sparse filtering | 33 |
| Figura 5.4: Diagrama de bloques tratando las clases independientemente | 35 |
| Figura 6.1: Representación de la curva ROC | 38 |
| Figura 6.2: Izquierda: Clasificación de accuracy para diferente número de características. Derecha: Precisión, sensibilidad, especificidad para una capa de 80 características y desviación típica entre los valores k de la validación cruzada. | 40 |
| Figura 6.3: Implementación de otra capa L2 a sparse filtering. | 40 |
| Figura 6.4: Accuracy obtenido para el entrenamiento con clases separadas y juntas utilizando cubos de 4x4x4. | 41 |
| Figura 6.5: Curva ROC para clasificación NC/AD utilizando capa L1 | 42 |
| Figura 6.6: Curva ROC para clasificación NC/AD utilizando capa L2 | 43 |

Índice de Tablas

| | |
|---|----|
| Tabla 4.1. Notación..... | 18 |
| Tabla 5.1. Demografía de los pacientes | 30 |
| Tabla 5.2: Regiones de interés [26] | 31 |
| Tabla 6.1: Resultados de clasificación. Valor medio \pm Desviación típica | 43 |
| Tabla 6.2: Resultados de clasificación. Valor medio \pm Desviación típica capa L2 | 43 |

Lista de Acrónimos

| | |
|--------------|---|
| <i>AD</i> | <i>Alzheimer's Disease</i> |
| <i>ADNI</i> | <i>Alzheimer's Disease Neuroimaging Initiative</i> |
| <i>CV</i> | <i>Cross Validation</i> |
| <i>MMSE</i> | <i>Mini-Mental State Examination</i> |
| <i>MRI</i> | <i>Magnetic Resonance Images</i> |
| <i>NIFTI</i> | <i>Neuroimaging Informatics Technology Initiative</i> |
| <i>PET</i> | <i>Tomografía de Emisión de Positrones</i> |
| <i>ROI</i> | <i>Region of Interest</i> |
| <i>SF</i> | <i>Sparse Filtering</i> |
| <i>SVM</i> | <i>Support Vector Machines</i> |

Capítulo 1. Introducción

En este primer capítulo se pretende poner de manifiesto la necesidad de desarrollar un sistema de ayuda al diagnóstico temprano de la enfermedad de Alzheimer (AD). El diagnóstico precoz permite que el paciente pueda planificar su vida, además de hacer posible la administración temprana de un tratamiento eficaz que consiga ralentizar los síntomas de la enfermedad. Otra ventaja de la detección temprana es la mejora de la calidad de vida del paciente y de su cuidador, retrasando la instalación del paciente en una residencia, contribuyendo así a mitigar los costes de la enfermedad [1].

Por otra parte, en este capítulo se describen los objetivos del presente proyecto, así como la estructura de los contenidos de esta memoria. Dentro de los objetivos, destacamos el desarrollo de un sistema de caracterización y clasificación sobre las regiones de interés de imágenes funcionales. Finalmente, se presenta una breve descripción de la organización de los contenidos de la memoria.

1.1. Motivación

La enfermedad de Alzheimer (AD) es la causa de demencia más común afectando a 30 millones de personas en todo el mundo. Debido al crecimiento de la esperanza de vida y al envejecimiento de la población en los países desarrollados, se espera que la AD afecte a 60 millones de personas en todo el mundo en los próximos 50 años.

Esta enfermedad no causa la muerte pero produce un deterioro paulatino de la capacidad cognitiva y funcional de la persona que lo presenta [2]. Entre las funciones corticales superiores que se ven afectadas por la evolución de esta enfermedad, la memoria es la más perceptible en sus primeras fases, siendo difícil diferenciar entre la enfermedad o los propios despistes cotidianos por el envejecimiento.

A pesar de todo esto, el AD hoy día no tiene cura, sin embargo se ha demostrado que los cerebros que cuentan con una reserva cognitiva pueden retrasar la aparición del Alzheimer, así como su intensidad en caso de que llegara a aparecer. Esta reserva cognitiva es la resistencia de nuestro cerebro frente a esta enfermedad, por tanto un diagnóstico temprano es necesario para poder aplicar un tratamiento que refuerce el mantenimiento de esa reserva cognitiva para la prevención de su avance.

En la actualidad el diagnóstico temprano del Alzheimer sigue suponiendo un reto. Los ensayos clínicos han permitido avanzar significativamente en el conocimiento de la enfermedad mostrándonos que dicha enfermedad se manifiesta entre 10 y 15 años antes de que aparezcan sus primeros síntomas. Debido a este reciente conocimiento y a las

técnicas de procesamiento de imágenes cerebrales, se han abierto numerosas líneas de investigación sobre el diagnóstico precoz. Las imágenes médicas actuales proporcionan gran cantidad de información sobre el estado del paciente, esto hace que el análisis de imágenes sea una técnica común y eficaz en el diagnóstico de la enfermedad [3].

Las nuevas técnicas de diagnóstico puede llevarse a cabo analizando los cambios funcionales o estructurales en el cerebro. Los cambios funcionales están asociados a los trastornos neurológicos que pueden deducirse mediante la tomografía de emisión de positrones (PET), ya que permite estudiar la activación de ciertas áreas del cerebro durante el desarrollo de tareas específicas. Por otro lado, los trastornos neurológicos también pueden ser descubiertos mediante el análisis de los cambios estructurales en el cerebro que suelen ser evaluados por resonancia magnética (MRI) [4]. Estos datos se pueden procesar de forma independiente o incluso fusionar con el fin de mejorar la precisión y disminuir la tasa de falsos positivos en las herramientas de ayuda al diagnóstico por ordenador o CAD. Tanto las imágenes MRI como las PET se han utilizado en numerosos trabajos para el diagnóstico automático. En el estudio [4] utilizan ambas imágenes cerebrales combinadas, cada una requiere de un procesamiento distinto, las imágenes PET son normalizadas en intensidad para poder compararlas mientras que las MRI requieren de un proceso de segmentación que proporciona información sobre las distribuciones de tejidos GM (Materia Gris) y WM (Materia Blanca), que pueden utilizarse para diferenciar los pacientes con AD de los pacientes CN, exponiendo un método novedosos para el diagnóstico de la AD combinando la clasificación SCR (Sparse Representation Classifiers) y obteniendo buenos resultados.

1.2. Descripción breve y objetivos

El presente trabajo tiene como objetivo la obtención de un algoritmo de parámetros óptimos que permitan mejorar la clasificación de los paciente que presentan la enfermedad de Alzheimer. Para lograr las características que conforman la enfermedad, partiremos de imágenes por tomografía por emisión de positrones (PET), de dos grupos diferentes; pacientes clasificados como sujetos control (CN) y pacientes diagnosticados de la enfermedad de Alzheimer (AD), teniendo únicamente en cuenta sus regiones de interés, es decir, aquellas que son más propensas a sufrir cambios funcionales conforme avanza la enfermedad. Posteriormente las características extraídas de estas regiones formarán la entrada al algoritmo de caracterización desarrollado, el cual reducirá el número de característica de entrada. Las características optimas obtenidas por el algoritmo de caracterización servirán de entrada al algoritmo de clasificación. El objetivo de este proyecto será alcanzar dichos parámetros óptimos que permitan extraer las característi-

cas que aportan mayor información de las imágenes y por tanto con ellas realizar una clasificación mejor de los pacientes.

1.3. Plan de trabajo

- Elaboración de una submuestra significativa de pacientes utilizando la base de datos ADNI (base de datos estandarizada que proporciona datos clínicos sobre AD).
- Identificar las regiones de interés para la enfermedad y transformarlas para su posterior análisis.
- Extracción de las características más relevantes para la enfermedad.
- Diseño de un clasificador basado en las características obtenidas.
- Documentación de los resultados obtenidos y conclusiones alcanzadas.

1.4. Estructura del trabajo

Introducción. Breve presentación del proyecto donde se explica las motivaciones y objetivos que se quieren alcanzar.

Enfermedad de Alzheimer. Se realiza una introducción de la enfermedad de Alzheimer, donde se describen los síntomas generales, las técnicas de diagnóstico y el tratamiento asociado a la enfermedad, además de los cambios neuroestructurales que se desarrollan en el cerebro durante la evolución de esta.

Base de datos ADNI. Se detalla la base de datos de “Alzheimer’s Disease Neuroimaging Initiative”. Técnicas utilizadas para el pre-procesamiento de las imágenes PET.

Métodos de extracción de características y clasificación. Se exponen las técnicas empleadas en este trabajo para la detección automática de patrones de hipoperfusión. Descripción y desarrollo de los métodos matemáticos utilizados para la selección de las características más representativas de cada región de interés así como el desarrollo de un algoritmo clasificador de dichas características.

Sistema de clasificación. Se expondrán los dos diseños de clasificación implementados, así como la base de datos usada y la descripción detallada de los bloques funcionales que componen el sistema.

Resultados experimentales. Resultados experimentales obtenidos por el clasificador, *Support Vector Machine* (SVM), basado en las características aprendidas por *Sparse Filtering* (SF). Se evaluará que hiperparámetros aportan características más representativas de la enfermedad y por tanto, clasifican mejor.

Conclusiones y trabajos futuros. Se realiza un resumen de las conclusiones a partir de los resultados obtenidos a lo largo del trabajo y las líneas de trabajo futuras que podrían mejorar el sistema.

Bibliografía. Referencias del proyecto.

Capítulo 2. Enfermedad de Alzheimer

En 1906 Alois Alzheimer, describió las lesiones cerebrales características del trastorno que recibió su nombre: placas seniles y ovillos neurofibrilares. La enfermedad de Alzheimer (AD) es ahora, 100 años después, la forma más común de demencia en el mundo. Debido al aumento de la esperanza de vida y al envejecimiento de la población, el AD se convertirá en un grave problema de salud pública en el futuro. Por ello, se requieren métodos urgentes de diagnóstico y tratamiento, ya que el diagnóstico definitivo de AD sigue siendo neuropatológico. Además la similitud de sus síntomas en los estadios iniciales de la degeneración celular con otros tipos de demencias, dificultan su diagnóstico. En los últimos 30 años esta enfermedad ha tomado una mayor relevancia, abriendo líneas de investigación sobre el tema y aprobándose el uso de fármacos que retarden el proceso de la enfermedad. Aún así, todavía no existe ningún tratamiento curativo o preventivo para el AD.

2.1. Síntomas

Los síntomas de la enfermedad de Alzheimer varían entre los individuos, presentándose en diferente orden y grado de afectación dependiendo del paciente. Podemos separar la enfermedad en tres ámbitos.

Síntomas cognitivos. Por lo general se ven afectadas la memoria, orientación temporal y espacial, orientación, lenguaje y capacidad de ejecución. A pesar de que la enfermedad se puede iniciar en cualquiera de estas áreas, el síntoma inicial más común es la incapacidad gradual para recordar nueva información es decir, la afectación de la memoria a corto plazo. Esto se debe a las lesiones cerebrales producidas en el hipocampo, que tienen un papel muy importante en la formación de nuevos recuerdos [5].

Síntomas psicopatológicos. Estos trastornos se deben a la grave afectación del lóbulo frontal que provocan una alteración de la conducta y de la personalidad del paciente. Los pacientes empiezan a presentar depresión, ansiedad, agresividad, trastorno del sueño, delirios...

Síntomas funcionales. La interrelación entre los síntomas cognitivos, psicológicos y conductuales provocan la incapacitación del paciente para realizar las tareas cotidianas habituales así como la limitación para emprender otras nuevas.

Estos síntomas varían dependiendo de la fase en la que se encuentre el paciente y acelerando su desarrollo dependiendo de la edad, nivel educativo, profesión, personalidad...

Los siguientes son síntomas más comunes de la enfermedad de Alzheimer:

- Cambios de memoria que dificultan la vida cotidiana.
- Dificultad para planificarse y resolver problemas.
- Dificultad para resolver tareas en la casa, en el trabajo o en el tiempo libre.
- Desorientación de tiempo o lugar.
- Dificulta para comprender imágenes visuales y cómo los objetos se relacionan uno al otro en el ambiente.
- Colocación de objetos fuera de lugar.
- Disminución o falta de buen juicio.
- Pérdida de iniciativa en el trabajo o en actividades sociales.
- Cambio en el humor o la personalidad.

2.2. Diagnóstico

La única forma de saber con certeza si el paciente ha sufrido AD es mediante un estudio microscópico de los tejidos cerebrales, una vez haya muerto. El cerebro del paciente con AD presenta un peso y volumen menor que el de un paciente normal, así como un aumento del tamaño de los ventrículos y de la profundidad de los surcos. Mientras el paciente está vivo solo se pueden hacer diagnósticos provisionales sobre la enfermedad [3], por tanto es de vital importancia un diagnóstico temprano que permita al paciente planificar el futuro y buscar redes de apoyo para su familia y para el propio paciente que le permita afrontar la enfermedad. La detección de la enfermedad en los primeros estadios puede lograr la autosuficiencia del paciente durante meses o incluso años. El diagnóstico del AD requiere una cuidadosa evaluación médica completa, ya que no existe una única y sencilla prueba para su valoración. El diagnóstico se concluirá por descarte de otros tipos de demencias y no será un diagnóstico definitivo. Para ello será necesario:

- **Evaluación clínica.** Entrevista con el paciente y sus familiares para descartar otras patologías que puedan ser la causa del deterioro.
- **Evaluación neuropsicológica.** Prueba que evalúa la memoria, la habilidad para resolver problemas y el funcionamiento emocional del paciente. La prueba más común para determinar el estado mental del paciente es el examen del estado mental mínimo (MMSE).
- **Pruebas para descartar otras causas.** Pruebas complementarias en los laboratorios para descartar otras demencias según los datos obtenidos en las pruebas anteriores.

Se buscan señales de derrames cerebrales, de Parkinson, tumores cerebrales, de acumulación de fluidos en el cerebro y de otras enfermedades que podrían afectar a la memoria. Para ello se evalúa los reflejos, la coordinación, la fuerza, el movimiento ocular, el habla y la sensibilidad.

2.2.1. Examen del Estado Mental Mínimo

El examen de estado mental mínimo (MMSE siglas en inglés de Mini Mental State Examination), fue desarrollado por Folstein en 1975 [6], consiste en una breve prueba que mide el deterioro cognitivo atendiendo a diferentes criterios como la edad, el nivel educativo y otras variables personales.

La prueba incluye 30 preguntas sobre orientación espacial y temporal, memoria diferida, atención, cálculo, lenguaje y capacidad visuoespacial. Esta prueba es utilizada internacionalmente, existiendo por tanto diversos criterios para la clasificación de los pacientes [7]. Por lo general su valor límite suele estar en torno a 26, sufriendo algún síntoma de demencia si la puntuación obtenida en esta prueba es menor; si es mayor se comprende que el paciente está sano. Esta es la clasificación más extendida:

- Puntuación 27 o mayor. Normal.
- Puntuación 26 a 21. Demencia leve.
- Puntuación 20 a 11. Demencia moderada.
- Puntuación 10 o menor. Demencia severa.

Aunque esta prueba no es definitiva, sus resultados nos pueden ayudar a determinar el estado mental del paciente o si son necesarias más evaluaciones. En promedio, la puntuación MMSE de una persona con Alzheimer decae cerca de dos a cuatro puntos cada año.

2.2.2. Examen de deterioro global

El examen de deterioro global de Reisberg se realiza para llevar un seguimiento del paciente con AD. Para determinar su grado de demencia se utiliza la escala de deterioro global (GDS), que clasifica al paciente en 7 etapas diferentes según su grado de deterioro cognitivo y funcional [8]. Según la evolución de la enfermedad se encontrarán en las siguientes escalas de GDS:

- GDS1. Ausencia de alteración cognitiva. Ausencia de trastornos de memoria.
- GDS2. Deterioro muy leve. Olvidos puntuales asociados con el envejecimiento. No hay evidencias de defectos objetivos en el trabajo o en situaciones sociales, pero se siente preocupación respecto a los síntomas.

- GDS3. Deterioro leve. Aparecen los primeros defectos de memoria tales como dificultad de concentración y desorientación al viajar al lugares nuevos, lo que puede provocar ansiedad. Esta etapa se encuentra aproximadamente 7 años antes del inicio de la demencia.
- GDS4. Deterioro moderado. Disminución de la capacidad para realizar tareas complejas como el control económico o la planificación personal.
- GDS5. Deterioro moderadamente severo. Necesidad de asistencia en determinadas tareas, como por ejemplo para elegir la ropa e incapacidad para ciertos datos de su vida cotidiana.
- GDS6. Deterioro severo. Desconocimiento de los acontecimientos y experiencias recientes de su vida.
- GDS7. Deterioro muy severo. Perdida progresiva del habla y de las habilidades psicomotoras básicas. En este estadio el paciente necesita asistencia para la higiene personal y la alimentación.

2.3. Tratamiento

Dado que actualmente no hay cura para la enfermedad de Alzheimer, el objetivo del tratamiento es mejorar, estabilizar o disminuir el deterioro cognitivo, funcional y conductual del paciente. La mejora de estos síntomas brinda a los pacientes bienestar, dignidad e independencia durante más tiempo, además de ayudar a las personas encargadas de atenderlos, ya que tienen que enfrentarse tanto a los cambios en las capacidades físicas del paciente como a los nuevos patrones de conducta adquiridos. Algunos de los medicamentos más utilizados son Razadyne (galantamina), Exelon (rivastigmina) y Aricept (donepezilo) llamados inhibidores de colinesterasa [9].

2.4. Cambios neuroestructurales provocados por la enfermedad

Las lesiones neuropatológicas empiezan a formarse probablemente años antes de la completa expresión de la demencia clínica, particularmente en los estadios de deterioro cognitivo leve que son a menudo precedentes al AD. No se sabe a ciencia cierta que comienza este proceso de degeneración celular o por qué algunos de los cambios normales asociados con el envejecimiento se vuelven mucho más extremos y destructivos en las personas con esta enfermedad. Lo que si se sabe es que aunque el desarrollo de la enfermedad no sea el mismo para cada paciente, los daños estructurales parecen sufrir las mismas etapas.

Patológicamente, la enfermedad de Alzheimer se caracteriza por los daños a las grandes neuronas corticales que abastecen la cognición, inicialmente en los lóbulos temporales y más tarde en las áreas restantes del neocórtex y las áreas asociadas. Se

caracteriza por la acumulación progresiva de fragmentos de proteínas beta-amiloide (placas seniles) fuera de las neuronas y cordones trenzados de la proteína tau (ovillos neurofibrilares) en el interior de las neuronas [3].

A medida que aumenta la enfermedad se va reducido el número de células nerviosas y las conexiones entre ellas, provocando que el cerebro se encoja como se puede ver en la Figura 2.1.

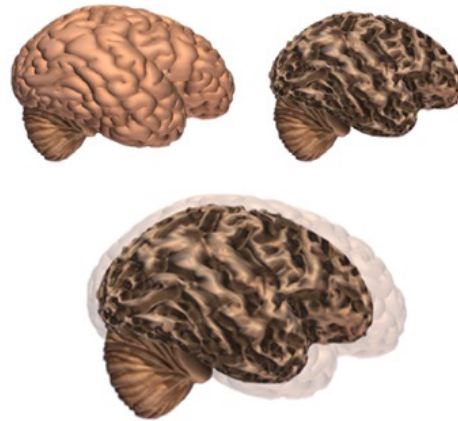


Figura 2.1: Cerebro sin AD (izquierda) , cerebro con AD (derecha) y comparación entre ambos (abajo).

En la Figura 2.2 se puede observar como la corteza se encoge, dañando las áreas del cerebro usadas para pensar, planear y recordar. El encogimiento es especialmente severo en el hipocampo, un área de la corteza que juega un papel importante en la formación de nuevos recuerdos y acciones. Por último, los ventrículos (espacios llenos de líquido dentro del cerebro) se vuelven más grandes [10].

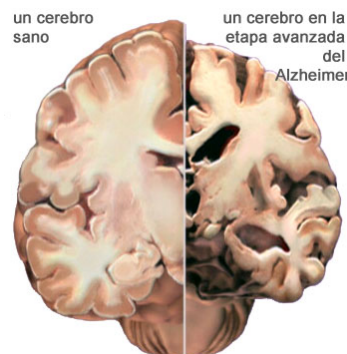


Figura 2.2: Corte transversal por la mitad del cerebro entre los oídos de un cerebro sano (derecha) y un cerebro con AD (izquierda).

Cuando se acerca la fase final de la enfermedad, los daños estructurales se han extendido ampliamente y los tejidos del cerebro se han deteriorado considerablemente afectando a muchas de las funciones del paciente.

Capítulo 3. Base de datos ADNI

Los datos utilizados para este estudio pertenecen a “Alzheimer’s Disease Neuroimaging Initiative” (ADNI). ADNI se lanzó en 2003 por el “National Institute on Aging” (NIA), el “National Institute of Biomedical Imaging and Bioengineering” (NI-BIB), la “Food and Drugs Administration” (FDA), compañías farmacéuticas privadas y organizaciones sin ánimo de lucro. El objetivo principal de ADNI ha sido probar si las imágenes de resonancia magnética (MRI), tomografías por emisión de positrones (PET), otros marcadores biológicos y la evaluación clínica y neuropsicológica se pueden combinar para medir la progresión del deterioro cognitivo leve (DCL) y la enfermedad de Alzheimer en etapas tempranas. ADNI es un repositorio de datos, generalmente accesible, que describe los cambios longitudinales en la estructura del cerebro y el metabolismo.

El investigador principal de esta iniciativa es Michael W. Weiner, MD, VA Medical Center y la Universidad de California, San Francisco. ADNI es el resultado de los esfuerzos de muchos investigadores de una amplia gama de instituciones académicas y empresas privadas, y los sujetos fueron reclutados de más de 50 sitios a través de los EE.UU. y Canadá. El objetivo inicial de ADNI era reclutar a 800 sujetos, pero ADNI ha sido seguida por ADNI-GO y ADNI-2. Hasta la fecha estos tres protocolos han reclutado a más de 1.500 adultos, entre 55 a 90 años, agrupados personas mayores sin daños cognitivos, personas con daños cognitivos leves y las personas con enfermedad de Alzheimer en etapas tempranas. La duración del seguimiento de cada grupo se especifica en los protocolos de ADNI-1, ADNI-2 y ADNI-GO. Los sujetos reclutados originalmente para ADNI-1 y ADNI- GO tenían la opción de seguir en ADNI-2. Para información actualizada, consulte www.adni-info.org. [11]

3.1. Imágenes médicas

Las imágenes neurológicas ayudan a estudiar los trastornos psiquiátricos de desarrollo neurológico. Además de permitir el estudio longitudinal en aquellos pacientes que sufren un deterioro cognitivo o funcional, así mismo también permite hacer una comparación con lo que se conoce como desarrollo neurológico normal, en aquellos casos que corran riesgo de presentarlo. Hay diferentes tipos de imágenes:

- Imágenes estructurales, **imágenes de resonancias magnéticas (MRI)**, esta técnica se basa en la aplicación de campos magnéticos sobre un cuerpo, desplazando sus átomos de hidrógeno. Cuando estos átomos vuelven a su posición inicial, se experimen-

ta un efecto que genera señales que son analizadas y convertidas en imágenes de la parte del cuerpo que este siendo examinada.

- Imágenes funcionales, **tomografía por emisión de positrones (PET)**, es una técnica no invasiva basada en la distribución tienen unas determinadas moléculas en el organismo, a estas moléculas se le llaman radiofármacos y estarán marcadas con isótopos emisores de positrones. Utilizando esta técnica, se puede realizar un estudio completo o parcial del sujeto en menos de una hora, siendo su principal inconveniente su elevado coste y por tanto su escasa disponibilidad.

3.2. Adquisición de imágenes PET

Antes de la adquisición de la imagen, el sujeto debe ayunar al menos 4 horas para que los niveles de glucosa en la sangre sean menores de 180 mg/dl (9,9 mmol/L). Tras la inyección del radiofármaco ([¹⁸F]-FDG o C-PIB), el paciente debe permanecer tumbado en silencio, en una habitación con luz tenue durante aproximadamente 45 min, tiempo que necesita el fármaco para distribuirse por el organismo. Posteriormente, se procede a la exploración en la que se tomarán datos con un escáner PET, en el caso de la iniciativa ADNI se trata de una cámara Siemens HR/HR + ECAT/ACCEL, durante aproximadamente 30 min [12].

3.2.1. Radiofármacos.

Los radionúclidos, isótopos radioactivos emisor de positrones (electrones cargados positivamente), son generalmente utilizados en medicina nuclear para el diagnóstico de enfermedades. Los más utilizados en las exploraciones PET son Carbono-11, Nitrógeno-13, Oxígeno-15, Fluor-18, Cobre-62, Galio-68, Rubidio-82. Estos isótopos son elegidos principalmente por su corto periodo de vida [13]. Los radionúclidos son incorporados en algún compuesto para expandirlo por el organismo, en el caso del Alzheimer lo más común es la glucosa, a estos compuestos se los conoce como radiofármacos. En la actualidad el radiofármaco más utilizado es el fluorodesoxiglucosa (FDG) donde el flúor de la molécula se convierte en F18. Este radiofármaco es el más utilizado debido a sus características metabólicas ya que algunos de sus compuestos están presentes en el cuerpo humano y también por su rápida expulsión del organismo sin provocar ningún efecto secundario. FDG es incorporado principalmente en las células con elevadas tasas de glucosa, como por ejemplo el cerebro, donde la fosforilación de la misma impide que sea liberada al metabolismo.

En el caso de Alzheimer, debido a su alta tasa de glucosa en la células cerebrales, la imagen PET muestra una disminución de glucosa en sus fases iniciales lo que nos permite identificar rápidamente la enfermedad. También se podrá conocer la efectividad

de los tratamientos, en cuyo caso se observará un aumento del metabolismo cerebral en relación con la situación inicial.

3.2.2. Funcionamiento de las cámaras PET

Una vez que el radiofármaco se encuentra en el interior del paciente, los positrones emitidos, tras un recorrido pequeño (2–3 mm), son atraídos fuertemente por los electrones del medio siendo aniquilados. La aniquilación de los positrones genera una energía en forma de par de fotones de 511 keV, en la misma dirección y sentidos opuestos [16]. Cada cámara PET presenta varios anillos adyacentes (entre 4 y 16) que tienen incorporados un circuito de detectores de coincidencias (cristales de centelleo) situados uno frente al otro separados por una línea de coincidencia (LC), capaces de valorar si los fotones han llegado en la misma dirección, con la misma energía y en coincidencia, generando un suceso “verdadero”. Desde su emisión hasta alcanzar el fotodetector, uno o los dos fotones puede sufrir el efecto Compton, desviándose de su trayectoria pero manteniendo su energía dentro de la ventana energética (511 keV), este suceso afecta a la calidad de la imagen. Cada detección de un fotón de aniquilación se denomina suceso “sencillo”, produciéndose un suceso “verdadero” cuando dos detectores registran en coincidencia ambos sucesos sencillos causados por los dos fotones del mismo suceso de aniquilación, en un intervalo de tiempo entre 6 y 12 ns [14].

También se pueden producir sucesos coincidentes “aleatorios” y de “dispersión” en las que se producen asignaciones incorrectas de la LC, esto degrada tanto la calidad de la imagen final como su valor cuantitativo. Los cristales de centelleo están acoplados a varios tubos fotomultiplicadores (TFM). La luz detectada por los TFM se utiliza para localizar el detector en el que ha interactuado el fotón incidente. Cada línea de detección viene determinada por una coordenada de muestreo angular y radial. Esto permite almacenar los sucesos coincidentes en matrices o sinogramas. A partir de la información de cada sinograma y por medio de la reconstrucción tomográfica se puede obtener los cortes axial, coronal y sagital de todas las regiones del cuerpo (Figura 3.1).

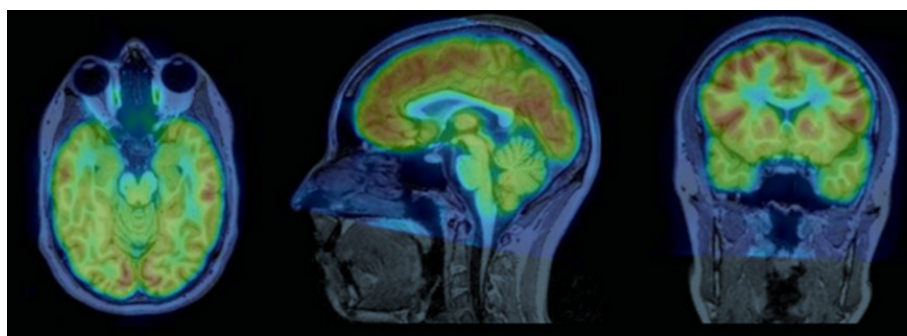


Figura 3.1: Fusión de las imágenes PET y RM de un sujeto sano.

Todos los datos de imágenes PET son procesados en formato DICOM (Digital Imaging and Communication in Medicine), que tiene como objetivo:

- Uniformidad de los datos PET disponibles.
- Utilización de imágenes PET de sistemas similares.

Finalmente, se convertirán a formato NIFTI, (Neuroimaging Informatics Technology Initiative), para su posterior análisis mediante SPM (Statistical Parametric Mapping).

El objetivo principal de NIFTI es proporcionar servicios coordinados y dirigidos, capacitación e investigación para acelerar el desarrollo y mejorar la utilidad de las herramientas informáticas relacionadas con la neuroimagen [15]. NIFTI, es una estructura nii que contiene tanto el archivo de encabezado, como la imagen 3D. Los encabezados contienen la información sobre los datos, tales como el sexo y la edad del paciente, el tamaño de los voxels, etc., todos ellos almacenados en una estructura de información y los datos en sí, como una matriz 3D, usualmente de tipo único. Todos los archivos de imágenes médica NIFTI que se encuentran en la base de datos ADNI siguen el mismo estándar.

Capítulo 4: Métodos. Caracterización y Clasificación

En este capítulo se introducirán los fundamentos matemáticos en los que están basados los algoritmos que se utilizan. Se usarán dos métodos de aprendizaje automático, uno de carácter no supervisado, como *sparse filtering* y otro supervisado, como es *support vector machine*. También se introducirá el concepto de validación cruzada que se utilizará para separar las muestras de test y de entrenamiento y los resultados que se obtienen.

4.1. Métodos de aprendizaje automático

El aprendizaje automático es una rama de la Inteligencia Artificial que permite a la maquina aprender de un conjunto de muestras de entrenamiento, creando un modelo de las categorías en las que se agrupan los datos, de tal forma que se puedan asignar a nuevos ejemplos una o más categorías de manera automática mediante analogía de patrones. Dependiendo del tipo de aprendizaje se puede clasificar en dos tipos:

- **No Supervisados.** El proceso de aprendizaje se lleva a cabo sobre un conjunto de ejemplos formados por las entradas del sistema de las que a priori no sabemos cuales serán las salidas. Estos modelos tienen la habilidad de caracterizar los datos de entrada, generando patrones de dichos datos.
- **Supervisados.** Aprendizaje a través de patrones de entrenamiento. Este aprendizaje es controlado por un supervisor que determina la respuesta que debe generar la red a la salida a partir de una entrada dada. Los algoritmos de aprendizaje supervisado han sido los modelos más utilizados desde los inicios de estos diseños y los que utilizaremos para clasificar las muestras en este trabajo.

4.2. *Sparse Filtering*

Sparse Filtering (SF) forma parte de los algoritmos no supervisados por lo que es capaz de caracterizar un conjunto de datos sin un conocimiento a priori de estos datos. El éxito de este algoritmo radica en la capacidad que posee para preservar la información relevante transmitida en la distribución de los datos. SF es eficiente y capaz de manejar datos de entrada de alta dimensionalidad [16]. En nuestro caso lo utilizaremos para generar un número reducido de características, las cuales serán posteriormente clasificadas por otro algoritmo de aprendizaje, en este caso de aprendizaje supervisado, *support vector machine* (SVM).

4.2.1. Escasez

Se puede definir el aprendizaje de distribución de características como cualquier algoritmo de aprendizaje no supervisado que se ocupa de generar una función de densidad de probabilidad de las características ($pdf\ p(Z)$) que tiene un conjunto de propiedades deseables. Los algoritmos de aprendizaje de distribución de características pasan por alto el problema de estimar la distribución verdadera $p(X)$ y en cambio, se centran en la formación de la $pdf\ p(Z)$ aprendida según los criterios elegidos. El algoritmo más representativo de esta familia es SF. Propiedades como la escasez son útiles al modelar la variabilidad de una distribución y por tanto pueden ser impuestas durante el aprendizaje de distribución de características [17].

- **Escasez.** Dado un vector genérico \mathbf{v} en un espacio N -dimensional, decimos que \mathbf{v} es escaso si un pequeño número de componentes del vector representa la mayor parte de la energía del vector. Siendo la energía del vector la procedente del procesamiento de la señal, es decir:

$$energy(\mathbf{v}) = \sum_{i=1}^N |v_i|^2 \quad (1)$$

Prácticamente, decimos que el vector \mathbf{v} es escaso si $n \ll N$, siendo n los componentes del vector \mathbf{v} que están activos (tienen un valor diferente de cero) mientras que los restantes $N-n$ componentes son inactivos (tienen un valor cero). Un vector \mathbf{v} es k -escaso si exactamente k componentes están activos. Por analogía, podemos definir las matrices de escasez y de variables aleatorias.

- **Medidas de escasez.** Una de las medidas más comunes de la escasez es la norma ℓ_1 . Dado un vector genérico \mathbf{v} en un espacio N -dimensional, su norma ℓ_1 se define como:

$$\ell_1(\mathbf{v}) = \sum_{i=1}^N |v_i| \quad (2)$$

En SF, la norma ℓ_1 a menudo aparece como activación. Dada una representación $\mathbf{z}^{(i)}$, se puede cuantificar su escasez calculando $\ell_1(\mathbf{z}^{(i)})$ o su activación ($\mathbf{z}^{(i)}$). Minimizar la activación de la representación aprendida $\mathbf{z}^{(i)}$ maximizará la escasez de $\mathbf{z}^{(i)}$.

Por tanto el objetivo de *sparse filtering* será aprender un $pdf\ p(Z)$ que maximice la escasez de las representaciones aprendidas $\mathbf{z}^{(i)}$.

4.2.2. Aplicación de la escasez en *Sparse Filtering*

La representación de la escasez aprendida $z^{(i)}$ se obtiene mediante el cumplimiento de tres restricciones sobre la matriz de representaciones aprendidas Z :

- **Escasez de población.** Cada muestra $z^{(i)}$ con $1 \leq i \leq N$, se requiere que sea descrita sólo por algunas características. La escasez de una muestra $z^{(i)}$ se calcula como su activación:

$$\ell_1(z^{(i)}) = \sum_{j=1}^L |z_j^{(i)}| \quad (3)$$

- **Escasez de por vida.** Cada característica z_j con $1 \leq j \leq L$, se requiere que sea descrito sólo por algunas muestras. La escasez de una característica z_j se calcula como su activación:

$$\ell_1(z_j) = \sum_{i=1}^N |z_j^{(i)}| \quad (4)$$

- **Alta Dispersión.** Todas las características z_j con $1 \leq j \leq L$, deben tener aproximadamente la misma activación. La dispersión de las características z_j se calcula como la varianza de su activación: $Var [activación(z_j)]$. La menor varianza corresponderá a una mayor dispersión.

La aplicación de estas tres propiedades equivale a una representación del aprendizaje escaso no degenerado.

4.2.3. Implementación de sparse filtering

Sparse filtering puede implementarse en seis pasos como se ilustra en la Figura 4.2.

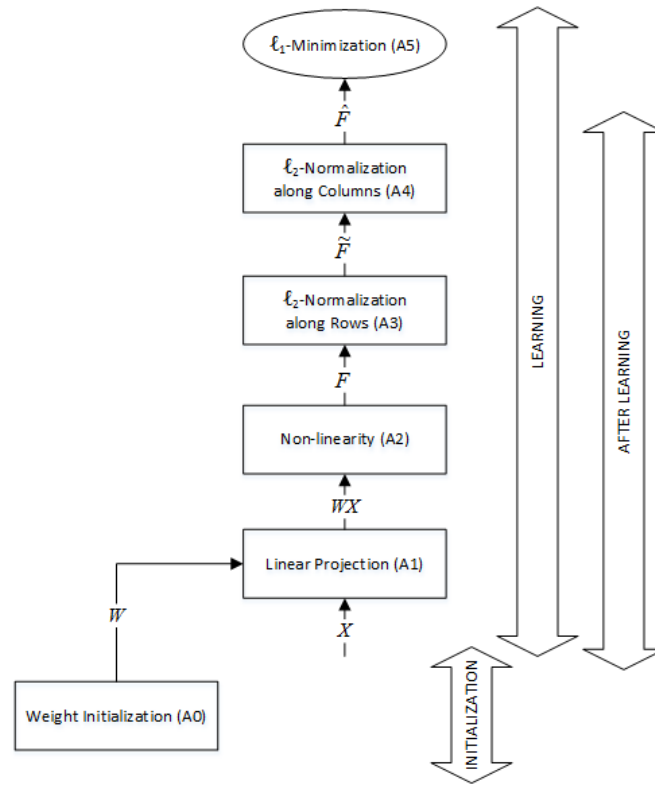


Figura 4.1. Descomposición del algoritmo en sus etapas constituyentes.

Primero vamos a ver la notación que se seguirá, Tabla 4.1, para explicar brevemente cada uno de los pasos que se sigue en la implementación del algoritmo.

| Principales identificadores utilizados | |
|--|---|
| O | Dimensionalidad original de las muestras. |
| L | Dimensionalidad aprendida de las muestras. |
| N | Número de muestras |
| W | Matriz de pesos con dominio $\mathbb{R}^{L \times O}$ |
| X | Matriz de representaciones originales con dominio $\mathbb{R}^{O \times N}$ |
| Z | Matriz de representaciones aprendidas con dominio $\mathbb{R}^{L \times N}$ |

Tabla 4.1. Notación

A0. Inicialización de los pesos: Se inicializa la matriz de pesos **W** con dimensiones (L x O) por muestreo de cada componente de una distribución normal $N(0, 1)$.

A1. Proyección lineal de los datos originales: $f_{A1}(\mathbf{X}) = \mathbf{W}\mathbf{X}$. La matriz de pesos \mathbf{W} puede interpretarse como un diccionario, donde cada fila es un filtro aplicado a cada muestra en las columnas de \mathbf{X} .

A2. Transformación no lineal: $F = f_{A2}(\mathbf{W}\mathbf{X})$, donde $f_{A2}(\mathbf{W}\mathbf{X})$ perteneciente al dominio real, es una función no lineal. Aunque esta función no lineal puede ser elegida arbitrariamente, por lo general se utiliza una función con el valor absoluto de los elementos $f(x)=|x|$. Por razones prácticas, esta no linealidad se implementa como $f(x)=(x^2+e)^{1/2}$, donde e será el error de cada elemento.

A3. Normalización ℓ_2 a lo largo de las características (filas):

$$\tilde{F} = f_{A3}(F) = \frac{F_j^{(i)}}{\sqrt{\sum_{i=1}^N (F_j^{(i)})^2}} \quad (5)$$

En este paso, cada característica es normalizada para que su ajuste de activación sea uno, es decir,

$$\sum_{i=1}^N (\tilde{F}_j^{(i)})^2 = 1 \quad (6)$$

A4. Normalización ℓ_2 a lo largo de las muestras (columnas):

$$Z = \hat{F} = f_{A4}(\tilde{F}) = \frac{\tilde{F}_j^{(i)}}{\sqrt{\sum_{j=1}^L (\tilde{F}_j^{(i)})^2}} \quad (7)$$

En este paso, como antes, cada muestra se normaliza de modo que el ajuste de activación es uno, es decir,

$$\sum_{j=1}^L (\hat{F}_j^{(i)})^2 = 1 \quad (8)$$

A5. Minimización ℓ_1 :

$$\sum_{ij} \hat{F}_j^{(i)} \quad (9)$$

Esta minimización es el objetivo del *sparse filtering*; minimizando la activación global de la matriz F , maximizamos la dispersión de la matriz de las representaciones aprendidas.

El paso A0 se ejecuta sólo durante la inicialización del algoritmo, mientras que el paso A5 se ejecuta sólo durante el aprendizaje. Después de aprender, los nuevos datos X' , se procesan entre los pasos A1 y A4, es decir, $Z' = f_{A1:A4}(X') = f_{A4}(f_{A3}(f_{A2}(WX')))$.

Como se explica en [18], la combinación de la minimización ℓ_1 con dos normalizaciones ℓ_2 garantiza el aprendizaje de una representación con las propiedades explicadas anteriormente (escasez de población, escasez de por vida y alta dispersión).

4.2.4. Aprendizaje de la distribución de características

En términos generales podemos decir que una buena representación no supervisada puede ser obtenida satisfaciendo dos requisitos teóricos:

- Maximizar la información entre la entrada y la salida (principio infomax).
- Maximizar la información de la salida de manera que conduzca a la extracción de información útil del ruido (principio de informatividad).

Según esto, el objetivo del aprendizaje de distribución de características puede expresarse como pura optimización del principio de informatividad. Es decir, tratamos de aprender una transformación f tal que la función de densidad de probabilidad de las características (*pdf* $p(z)$) de las nuevas representaciones $z^{(i)}$ tengan máxima informatividad. Maximizar la informatividad puede expresarse simplemente como la maximización de la entropía relativa (también llamada divergencia de Kullback-Leibler, es una medida no simétrica de la similitud o diferencia entre dos funciones de distribuciones de probabilidad) entre la *pdf* aprendida $P(Z)$ y la maximización de la entropía *pdf* q :

$$\max D[p(Z) \parallel q] \quad (10)$$

Donde $D[.]$ es la medida de la divergencia entre las *pdfs*.

Sin embargo, el objetivo de maximizar la información no es suficiente para conducir a un aprendizaje útil o significativo. La solución óptima del problema de maximizar $D[p(Z) \parallel q]$ es trivialmente el aprendizaje de una *pdf* con la forma de una delta de Dirac. Si trazamos todas las muestras originales $x^{(i)}$ sobre una representación arbitraria \mathbf{z} , entonces la *pdf* $p(Z)$ tendrá la forma de una función delta de Dirac centrada en \mathbf{z} . Esta *pdf* tiene entropía mínima y, por lo tanto, máxima informatividad. Sin embargo, mapeando arbitrariamente todas las muestras $x^{(i)}$ a una representación constante \mathbf{z} no tiene ningún

significado, por tanto maximizaremos la información en $p(Z)$ pero perdemos toda la información transportada por $p(X)$ sobre $P(X^*)$.

$$\max D[p(X,Z) \parallel p(X)p(Z)] \quad (11)$$

Donde $D[.]$ es otra vez la medida de la divergencia entre las *pdfs*.

Al igual que cualquier algoritmo de aprendizaje sin supervisión, el aprendizaje de distribución de características tiene que negociar el equilibrio entre el principio infomax y el principio de informatividad:

$$\underbrace{\max D[p(X,Z) \parallel p(X)p(Z)]}_{\text{Principio de infomax}} + \underbrace{\max D[p(Z) \parallel q]}_{\text{Principio de informatividad}} \quad (12)$$

Incluso si la definición de aprendizaje de distribución de rasgos no hace referencia al principio infomax, argumentamos que debe tenerse en cuenta de alguna manera.

- ***Aprendizaje de la distribución de datos.*** En relación con la ecuación 10, podemos definir el aprendizaje de distribución de datos como cualquier aprendizaje sin supervisión en el que el objetivo principal es maximizar el principio infomax (primer término en la ecuación 10), mientras que la maximización del principio de informatividad (segundo término de la ecuación 10) se contabiliza a través de restricciones.
- ***Aprendizaje de la distribución de características.*** En relación con la ecuación 10, podemos definir el aprendizaje de distribución de características como cualquier aprendizaje no supervisado en el que el objetivo principal es maximizar el principio de informatividad (segundo término de la ecuación 10), mientras que la maximización del principio infomax (primer término de la ecuación 10) se contabiliza a través de restricciones.

Por tanto, *Sparse Filtering* esta basado en la búsqueda de un diccionario D no conocido, lo que hace que este algoritmo sea de aprendizaje no supervisado, que proporcione una representación de la dispersión minimizando las entradas normalizadas en una matriz de valor de característica.

4.3. Support Vector Machine (SVM)

Support Vector Machine se basan en un conjunto de algoritmos de aprendizaje supervisado que fue desarrollado por Vladimir Vapnik y que esta relacionado con problemas de clasificación y regresión [18]. Este algoritmo se basa en dos ideal claves :

- Maximización del margen de separación.
- El truco del kernel, un método de expansión un clasificador lineal a uno no lineal.

Para este proyecto se utilizara como clasificador *Support Vector Machine* (SVM), debido al reducido número de muestras disponibles en comparación con su dimensionalidad. SVM separa un determinado conjunto de datos de entrenamiento, con lo que se conoce como hiperplanos, definido como el vector que separa los dos puntos mas cercanos de distintas clases, en nuestro caso una clase es AD y la otra clase normal. El objetivo será construir una función que clasifique correctamente a los nuevos ejemplos.

4.3.1. SVM lineales con clases linealmente separables

Dado un conjunto de entrenamiento formado por:

$$x_i \in \mathbb{R}^d \text{ con } y_i \in \{+1, -1\} \text{ para } i = 1, \dots, n,$$

Se puede representar el hiperplano de separación entre las dos clases ($y_i = \pm 1$), como en la Figura 4.2, donde \mathbf{w} es un vector de coeficientes reales de dimensión d , mientras que b es un coeficiente real [19].

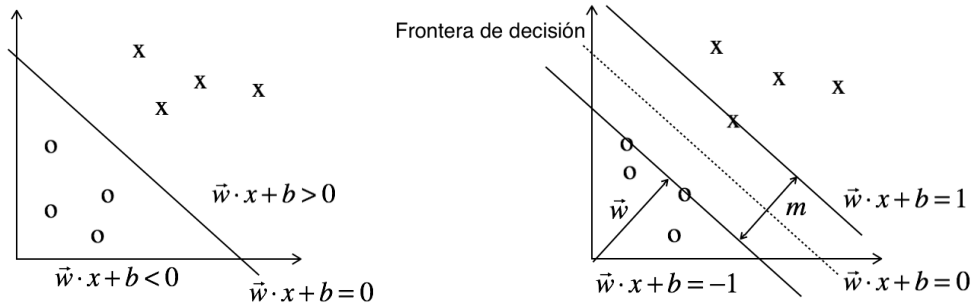


Figura 4.2. Izquierda: Separación de clases linealmente. Derecha: Hiperplanos de máxima separación

La distancia perpendicular desde el hiperplano al punto, $y_i = +1$, más cercano de una clase se denota por x_+ , mientras que la distancia mas cerca desde el hiperplano a un punto de la otra clase, $y_i = -1$, se denotará como x_- . La anchura del margen se define como $\min(x_+, x_-)$ y la solución SVM busca el vector de pesos que maximiza este margen.

Imponiendo algunas restricciones:

$$\left. \begin{aligned} y_i = +1 &\rightarrow \vec{w} \cdot x_+ + b \geq +1 \\ y_i = -1 &\rightarrow \vec{w} \cdot x_- + b \leq -1 \end{aligned} \right\} y_i(x_i \cdot \vec{w} + b) \geq 1 \quad (13)$$

Estas se pueden combinar en un solo conjunto de desigualdades:

$$y_i = \{1, -1\} \rightarrow y_i(x_i \cdot \vec{w} + b) \geq 1 \quad (14)$$

Haciendo consideraciones geométricas, x_+ será igual a x_- , de modo que hay al menos un punto en cada clase para la cual:

$$y_i(x_i \cdot \vec{w} + b) = 1 \quad (15)$$

Dada las expresiones anterior, se puede deducir que habrá infinitos hiperplanos capaces de separar las muestra linealmente, es por tanto se necesitará una condición adicional capaz de definir un hiperplano de separación óptimo. Para ello, vamos a tener en cuenta el margen, Por geometría, se sabe que la distancia entre ambos hiperplanos (x_+, x_-), viene dado por:

$$m = (x_+ - x_-) \cdot \frac{\vec{w}}{\|\vec{w}\|} = \frac{2}{\|\vec{w}\|} \quad (16)$$

Los elementos que delimitan el margen reciben el nombre de vectores de soporte, estos elementos son los más difíciles de clasificar y por tanto serán las únicas muestras que tendremos en cuenta para delimitar el hiperplano.

La búsqueda del margen máximo para el caso de clases separadas linealmente se reduce a encontrar w y b que minimiza la función.

$$\max \frac{2}{\|\vec{w}\|} = \min \frac{1}{2} \|\vec{w}\|^2 \quad (17)$$

Este problema de optimización con restricciones corresponde a un problema de programación cuadrática y se aborda mediante la teoría de la optimización. Dicha teoría establece que un problema de optimización, tiene una forma dual si la función a optimizar y las restricciones son funciones estrictamente convexas.

De este modo se reduce el problema de optimización sin restricciones utilizando la función Lagrangiana:

$$L(w, b, \alpha) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i(\vec{w} \cdot x_i + b) - 1] \quad (18)$$

donde α_i son coeficientes no negativos denominados multiplicadores de Lagrange.

$$\frac{\partial L}{\partial \vec{w}} = \vec{w} - \sum_{i=1}^n \alpha_i y_i x_i = 0 \rightarrow \vec{w} = \sum_{i=1}^n \alpha_i y_i x_i \quad (19)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n \alpha_i y_i = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \quad (20)$$

Siendo α_i cero para cada x_i excepto aquellos que se encuentran en los hiperplanos. Observe que la solución es una combinación lineal de x_i . Esto indicará que el vector \mathbf{w} es una suma lineal de todas las muestras o algunas de las muestras.

Con las restricciones obtenidas anteriormente se puede expresar la función Lagrangiana solo en función de α_i , siendo:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad \text{para} \quad \sum_{i=1}^n \alpha_i y_i = 0 \quad (21)$$

Así mismo, se observa que la optimización depende solo del producto escalar de los pares de muestras y por tanto también el margen.

$$\sum_{i=1}^n \alpha_i y_i x_i \cdot x + b \geq 0 \quad (22)$$

Esto indicará que el vector \mathbf{w} es una suma lineal de todas o algunas de las muestras.

4.3.2. SVM lineales con clases no separables

Normalmente se encuentran elementos de una clase dentro de la región de otra clase, por tanto no se pueden separar las clases con hiperplanos. En estos casos los conjuntos no serán linealmente separables.

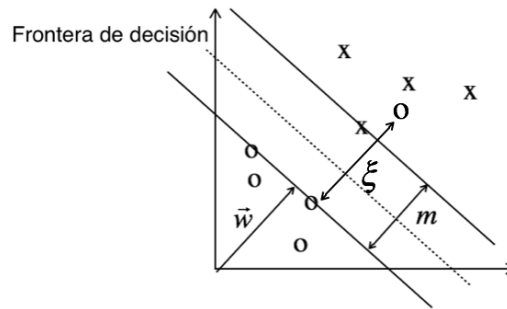


Figura 4.3: Aparición del parámetro de error ξ en la clasificación

Para poder resolver este problema, ampliaremos las ideas generales del caso separable introduciendo una variable de “relajación”, $\xi \geq 0$, para cada ejemplo de entrenamiento.

$$y_i = +1 \rightarrow x_i \cdot \vec{w} + b \geq +1 - \xi_i \quad (23)$$

$$y_i = -1 \rightarrow x_i \cdot \vec{w} + b \leq -1 + \xi_i \quad (24)$$

Se tiene ahora que para que se produzca un error en la clasificación de un vector de entrenamiento es necesario que el valor correspondiente a ξ sea superior a la unidad.

De esta manera se le asigna a la función de costes un coste adicional que penaliza los errores. La función de optimización vendrá dada por:

$$\min \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^n \xi_i^k \quad (25)$$

Donde C es la cantidad máxima de pérdidas permitidas, la cuál se puede establecer, por ejemplo, mediante validación cruzada. k es un entero cuyo valor normalmente será 1. Planteando el problema de manera similar al de variables separables llegamos a:

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (26)$$

bajo la condición de: $\sum_{i=1}^n \alpha_i y_i = 0$ y donde $0 \leq \alpha_i \leq C$ para $i=1, \dots, n$.

Donde la única diferencia con lo obtenido para variables separadas es el límite C . En este caso cuando $0 < \alpha_i < C$, el punto estará en el margen, es decir $x_i = 0$ y cuando $\alpha_i = C$, el punto estará dentro del margen si $\xi \leq 1$, o en el lado equivocado de la frontera de decisión si $\xi > 1$. Teniendo en cuenta esto, la función de decisión vuelve a ser:

$$\sum_{i=1}^n \alpha_i y_i x_i \cdot x + b \quad \text{con } 0 \leq \alpha_i \leq C \quad (27)$$

El valor de C debe ser asignado para ejecutar el algoritmo. Este parámetro controla indirectamente la anchura del margen del clasificador. Existe un compromiso entre el ancho del margen y el número de errores de clasificación aceptados. No existe un procedimiento óptimo para asignar este parámetro, pero se espera que si C es grande, los errores de clasificación sean relevantes y el margen debe ser más estrecho. Por otro lado, si

C es pequeño, los errores de clasificación será menos relevantes y se espera que haya un margen más amplio.

4.3.3. El Truco del Kernel

Como se ha podido observar en los problemas duales, la única manera en que los puntos de datos aparecen en los problemas es como productos escalares $x_i \cdot x_j$. Supondremos que se llevan los datos a un nuevo espacio F (espacio de características) mediante una transformación φ como se muestra en la Figura 4.4. Se transformará el espacio de entrada a otro de mayor dimensión en el que los datos son separables linealmente. De esta manera se podrá aplicar sobre este nuevo espacio el procedimiento para variables separadas linealmente [20].

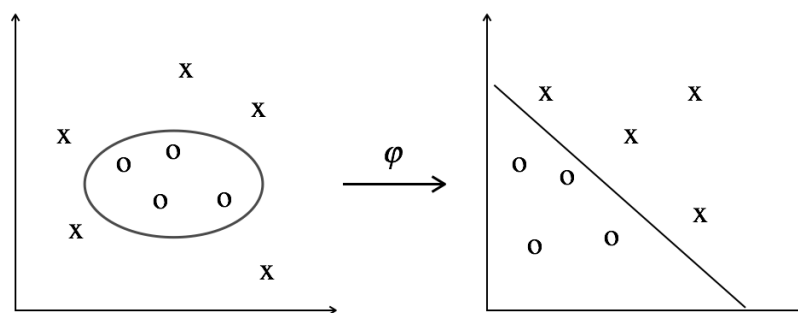


Figura 4.4: Utilización del truco del Kernel para transformar el espacio de datos

En este nuevo espacio el algoritmo de entrenamiento solo dependerá de los datos a través de los productos escalares, es decir, de funciones de la forma $\varphi(x_i) \cdot \varphi(x_j)$. Así el truco de Kernel es reemplazar $\varphi(x_i) \cdot \varphi(x_j)$ por una función conocida como función de Kernel, $k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$, donde podríamos utilizar únicamente k en el algoritmo (para ambas fases de entrenamiento y de test) sin necesidad de conocer φ .

Para caracterizar las funciones de kernel se utilizará la “condición de Mercer”, permitiendo determinar cuales son las condiciones que deben cumplirse para que el producto escalar en el espacio de salida se pueda escribir a partir de un cierto núcleo $k(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$.

En este estudio utilizaremos una de las funciones de kernel típicas, la lineal, $k(x_i, x_j) = x_i \cdot x_j$. En general con esta función de Kernel se obtienen peores resultados si los datos no son linealmente separables pero nos permitirá reducir significativamente el coste computacional en tiempo y en memoria, tanto del aprendizaje como de la clasificación.

4.4. Validación Cruzada

En este estudio, para la construcción de los conjuntos de test y de entrenamiento, se usará el método conocido como validación cruzada. En este método los datos se dividen en k subconjuntos iguales, realizando k iteraciones. En cada iteración se toma uno de los subconjuntos como datos de prueba y el resto ($k-1$) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones, con cada uno de los subconjuntos como datos de prueba. El error se calcula como la media aritmética de los errores de cada iteración para obtener un único resultado [21].

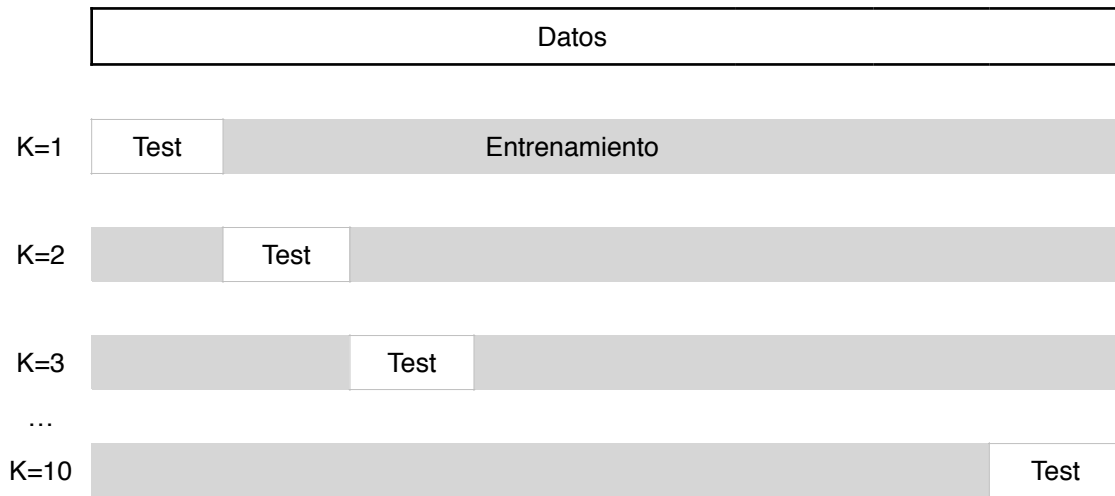


Figura 4.5: Proceso de validación cruzada.

Capítulo 5. Sistemas de Clasificación

En este capítulo se propone el diseño de dos sistemas de clasificación empleando los conceptos teóricos explicados en el capítulo 4. Ambos diseños están basados en la obtención de unas características, determinadas por un algoritmo de aprendizaje no supervisado. Dichas características serán adquiridas de distinto modo para cada diseño, de esta manera podremos comparar los conjuntos aportados por ambos modelos, para lograr las características más relevantes a la hora de discernir si un paciente nuevo se engloba dentro de pacientes diagnosticados de Alzheimer (AD) o de sujetos de control (NC).

Este procedimiento se inicializará con el procesamiento de las imágenes PET de la base de datos ADNI para posteriormente explicar con detalle los bloques funcionales de los que se componen ambos diseños.

5.1. Primer Diseño

En este apartado se expondrán los bloques funcionales de los cuales se compone este diseño, estos bloques vienen representados en la Figura 4.6, a lo largo de este apartado se explicará la función de cada uno de ellos.

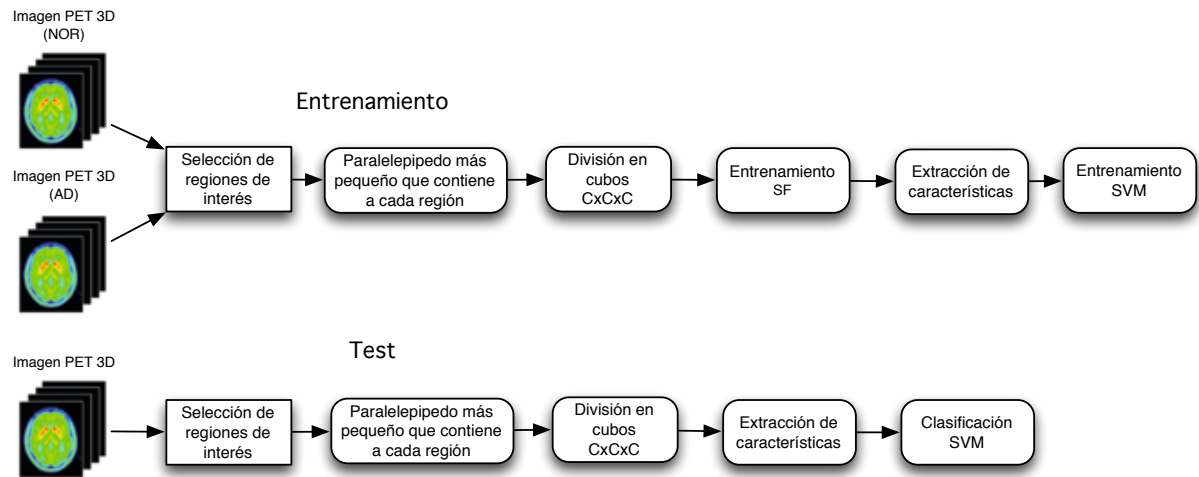


Figura 5.1: Diagrama de bloques del sistema

5.1.1. Base de datos ADNI

La base de datos utilizada en este trabajo contiene imágenes PET de 138 sujetos, que se comprende de 68 sujetos controles (NC) y 70 diagnosticados de la enfermedad de Alzheimer (AD), esta base de datos fue la descrita en el capítulo 3. Este repositorio fue creado para estudiar el avance de la enfermedad de Alzheimer, recolectando una gran cantidad de imágenes por tomografía por emisión de positrones (PET), así como biomarcadores de sangre y análisis del líquido cefalorraquídeo. El objetivo principal de esta base de datos es proporcionar una forma de diagnosticar temprano de la enfermedad de Alzheimer. La demografía de los pacientes empleada en este trabajo se muestra en la Tabla 5.1.

| Evaluación | Sexo (M/F) | Edad media \pm Std | Resultado MMSE \pm Std |
|------------|------------|----------------------|--------------------------|
| NC | 43/25 | 78.81 \pm 4.93 | 29.06 \pm 1.08 |
| AD | 46/24 | 75.33 \pm 7.17 | 22.84 \pm 2.91 |

Tabla 5.1. Demografía de los pacientes

5.1.3. Selección de regiones de interés.

Un aspecto clave de este trabajo consiste en dividir el cerebro en diferentes regiones que serán clasificadas por separado, para ello se utiliza un atlas que define las diferentes regiones del cerebro. En particular, se ha trabajado con un atlas de 116 regiones (AAL de sus siglas en inglés Automated Anatomical Labeling Atlas) para extraer las características. Estas regiones cerebrales correspondientes a diferentes áreas neuroanatómicas. A modo de ejemplo, la Figura 5.2, muestra la imagen PET y su atlas correspondiente.

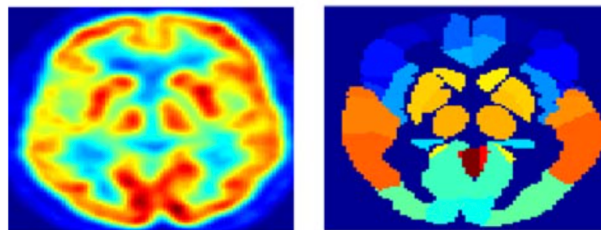


Figura 5.2: Izquierda, imagen PET. Derecha, atlas PET [25]

Para este estudio, de las 116 regiones que tiene el atlas, sólo se han seleccionado 42 regiones, distribuidas en los lóbulos frontal, parietal, occipital y temporal, ya que se con-

sideran potencialmente relacionadas con la AD [26]. En la Tabla 5.2 se muestran las 42 regiones utilizadas ("Left" significa que la región del cerebro está situada en el hemisferio izquierdo, "Right" significa hemisferio derecho) y sus etiquetas:

| Frontal lobe | | Parietal lobe | | Occipital lobe | | Temporal lobe | |
|--------------|--|---------------|---------------------------------|----------------|--------------------------------|---------------|--|
| 3 | Left Superior Frontal gyrus | 59 | Left Superior Parietal gyrus | 49 | Left Superior Occipital gyrus | 81 | Left Superior Temporal gyrus |
| 4 | Right Superior Frontal gyrus | 60 | Right Superior Parietal gyrus | 50 | Right Superior Occipital gyrus | 82 | Right Superior Temporal gyrus |
| 7 | Left Middle Frontal gyrus | 61 | Left Inferior Parietal | 51 | Left Middle Occipital gyrus | 83 | Temporal pole: Left Superior Temporal gyrus |
| 8 | Right Middle Frontal gyrus | 62 | Right Inferior Parietal | 52 | Right Middle Occipital gyrus | 84 | Temporal pole: Right Superior Temporal gyrus |
| 23 | Left Superior Frontal gyrus, medial | 67 | Left Precuneus | 53 | Left Inferior Occipital gyrus | 85 | Left Middle Temporal gyrus |
| 24 | Right Superior Frontal gyrus, medial | 68 | Right Precuneus | 54 | Right Inferior Occipital gyrus | 86 | Right Middle Temporal gyrus |
| 25 | Left Frontal gyrus, medial orbital | 35 | Left Posterior cingulate gyrus | | | 87 | Temporal pole: Right Middle Temporal gyrus |
| 26 | Right Frontal gyrus, medial orbital | 36 | Right Posterior cingulate gyrus | | | 88 | Temporal pole: Left Middle Temporal gyrus |
| 27 | Left Gyrus Rectus | | | | | 89 | Left Inferior temporal gyrus |
| 28 | Right Gyrus Rectus | | | | | 90 | Right Inferior temporal gyrus |
| 31 | Left Anterior Cingulate and Paracingulate gyrus | | | | | 55 | left Fusiform gyrus |
| 32 | Right Anterior Cingulate and Paracingulate gyrus | | | | | 56 | Right Inferior temporal gyrus |
| | | | | | | 37 | Left Hippocampus |
| | | | | | | 38 | Right Hippocampus |
| | | | | | | 39 | Left ParaHippocampal gyrus |
| | | | | | | 40 | Right ParaHippocampal gyrus |

Tabla 5.2: Regiones de interés [26]

5.1.4. Paralelepipedo más pequeño que contiene a cada región

El atlas mencionado anteriormente contiene todas las regiones del cerebro discriminativas para el AD. Con objeto de obtener las regiones de interés para este estudio se dispuso de 42 máscaras, una por cada región de interés, que contienen un identificador de los voxels de cada región, entendiéndose voxels como un punto en una imagen

3D. Una vez localizadas las regiones de interés, cada región es encerrada en el paralelepipedo más pequeño que las contengan, para su posterior procesamiento. Esto se hará para cada paciente y para cada región.

5.1.5. División en cubos

Con el fin de seleccionar las mejores características, de cada paralelepipedo, se hará una división en ventanas, es decir se extraerán cubos de distintos tamaños $C \times C \times C$. El objetivo de dividir cada región en trozos, es buscar patrones en cada región (un patrón es algo que se repite). Estos cubos (ventanas que hacen que se extraigan características locales) deben ser lo suficientemente pequeños para no perder resolución y lo suficientemente grandes para capturar esos patrones. Los diferentes tamaños de cubos extraídos para este estudio fueron $3 \times 3 \times 3$, $4 \times 4 \times 4$, $5 \times 5 \times 5$ y $6 \times 6 \times 6$ sin solapamiento entre ellos.

5.1.6. *Sparse Filtering*.

Tras dividir cada región en cubos de tamaño $C \times C \times C$ se pasa a la extracción de las características con mayor información sobre la enfermedad que ayudarán a clasificar a los pacientes de Alzheimer mediante SF. Los hiperparámetros que necesitará *sparse filtering* serán:

- El número de capas.
- Número de características de cada capa.

En este estudio se hará el diseño con una capa L1 y posteriormente, con los hiperparámetros que mejores resultados se obtengan, se diseñará otra capa L2, para comparar los resultados de clasificación alcanzados con ambas capas. Los datos se dividirán en fase de entrenamiento y fase de test mediante validación cruzada. Cada sujeto del conjunto de entrenamiento estará etiquetado (AD o CN), mientras que los sujetos del conjunto de test, no estarán etiquetados. Las etiquetas de los sujetos test se usarán después de la clasificación, esto nos permitirá conocer la validez de los resultados obtenidos.

5.1.6.1. Entrenamiento *Sparse Filtering*

En la **fase de entrenamiento**, los cubos extraídos de las regiones de interés formarán la matriz de datos de entrada para SF. Sparse filtering calculará los pesos de esta matriz de características que serán interpretados como un diccionario, donde las filas serán el número de características a las que convergerá el algoritmo. Esto lo realizará para cada región de cada paciente.

5.1.6.2. Extracción de Características mediante *Sparse Filtering*

Una vez realizado el entrenamiento de sparse filtering, se continuará con la extracción de características que contienen mayor información. Para identificar dichas características se aplicará el SF sobre la entrada, esto es, los voxels de los cubos de cada región que forman la matriz de datos inicial. Con esto se consigue reducir el conjunto de características, mejorando la tasa de detección y acelerando el proceso de clasificación, ya que SF potencia las características más representativas.

En la *fase de test*, se hará uso de *sparse filtering* previamente entrenado con pacientes sin etiquetar. En esta fase el sistema ya estará previamente entrenado por lo que solo tenemos que obtener las características mediante el paso de feedforward, que nos permitirán obtener las características que posteriormente utilizaremos para la clasificación de los pacientes AD y CN.

A priori, como se explico en el capítulo 4, no podemos saber cuales son los elementos que caracterizan nuestro diccionario y por tanto no sabemos que se tomará como características. Una capa puede “aprender” a reconocer formas simples mientras que la otra capa podría aprender a combinar esas formas. La Figura 5.3 muestra las unidades de agrupación aprendidas en una segunda capa utilizando sparse filtering. Se muestran las unidades de la primera capa más fuertemente conectadas para cada unidad de segunda capa.



Figura 5.3: Características *sparse filtering*

5.1.7. *Support Vector Machine*.

Una vez obtenidas todas las características para cada región y cada paciente se pasa a la clasificación mediante SVM con aprendizaje supervisado. Este método tiene dos fases: entrenamiento y test, estos han sido delimitados mediante validación cruzada en la cual el resultado final que obtenemos en la clasificación, no se encuentra sesgado a

las muestras de clasificación, ya que debido al carácter finito de las muestras debemos maximizar su uso para entrenar y clasificar.

- *Entrenamiento.* Para la fase de entrenamiento, se realiza un mapeo del conjunto de entrenamiento al espacio Kernel, definido por una función lineal. El conjunto de entrenamiento estará formado por una matriz de características significativas de cada una de las clases, AD y NC, obtenido mediante *sparse filtering*. Para la realización del entrenamiento mediante SVM es necesario disponer de un conocimiento previo de las clases que se quieren clasificar para poder establecer las restricciones en el clasificador, para esto el clasificador dispone de las muestras etiquetadas (pacientes de cada clase), para poder modelar la frontera de decisión según nuestros criterios. Crearemos un clasificador para cada región de interés de cada paciente que sea capaz de diferenciar entre ambas clases.
- *Test.* En esta fase utilizaremos el clasificador previamente entrenado con pacientes sin etiquetar, esto nos dará una referencia sobre cuanto de bueno es el clasificador basándonos en los resultados obtenidos.

Se realizará un SVM de cada paciente para cada región del cerebro discriminante, clasificando cada región como 1 si es un sujeto AD y como 0 si es un sujeto CN. Para determinar si un paciente sufre AD se realizará un “voto por mayoría”, es decir, si dado un sujeto, la mayoría de sus regiones del cerebro son clasificadas como 1, el sujeto será AD, mientras que si la mayoría son clasificadas como 0, el sujeto será CN. Estos resultados posteriormente serán comparados con las etiquetas de cada paciente para determinar como de bueno es el sistema de caracterización y clasificación.

5.2. Segundo Diseño

En este apartado se expondrán los bloques funcionales para realizar un segundo diseño, estos bloques vienen representados en la Figura 5.4. La diferencia con el diseño anterior radica en la obtención de las características utilizadas para realizar la clasificación.

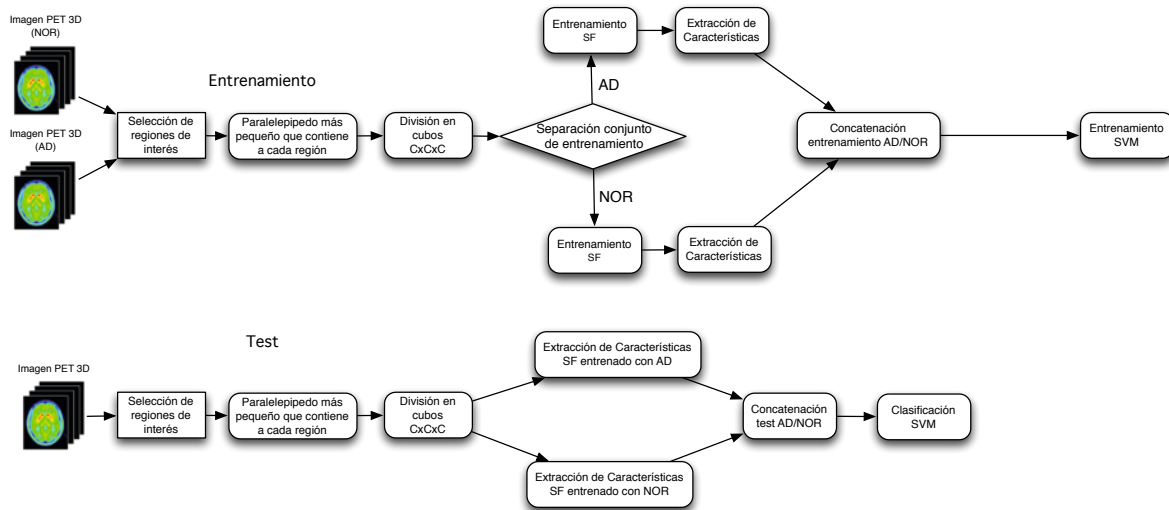


Figura 5.4: Diagrama de bloques tratando las clases independientemente

Las primeras fases: base de datos ADNI (donde se encuentran las imágenes PET utilizadas), selección de regiones de interés, encerrar en las regiones en el paralelepipedo más pequeño que las contienen y la extracción de cubos de los paralelepipedos se han explicado en el apartado 5.1, por tanto en este apartado se explicarán las diferencias significativas entre ambos diseños.

5.2.1. Sparse Filtering

Tras dividir la región en cubos de características de tamaño $C \times C \times C$, como se explica en el apartado anterior, se pasa al entrenamiento de SF y a la extracción de las características más significativas. Los conjuntos de entrenamiento y de test han sido delimitados mediante validación cruzada.

5.2.1.1. Entrenamiento *Sparse Filtering*

En este diseño, la **fase de entrenamiento** de *sparse filtering* se realizará a ambas clases AD y CN de forma independiente, es decir, se realizarán dos entrenamientos para SF, un entrenamiento con las clases AD y otro entrenamiento con las clases CN.

5.2.1.2. Extracción de Características mediante *Sparse Filtering*

La extracción de características se realizará mediante los dos *sparse filtering* entrenados. Por una parte, se extraerán características para SF entrenado con la clase AD y por otra parte, se extraerán características para SF entrenado con características CN. Cada conjunto de entrenamiento aportará características que permitirán diferenciar a esa clase, de esta manera se obtendrán el doble de características que con el diseño anterior, ya que cada entrenamiento con SF convergerá a un número de características fijado como parámetro de configuración.

En la **fase de test** se utilizará *sparse filtering* previamente entrenado con ambas clases, es decir, dado una región de un paciente, se aplicará sobre ella tanto el SF entrenado mediante los pacientes AD como el SF entrenado mediante los CN, realizando la extracción de características. Esto, como se ha explicado, aporta el doble de características que con el diseño anterior. Estas características deben ser concatenadas en el mismo orden en que se encuentran las etiquetas de los sujetos test, de manera que tendremos los mismos sujetos, pero el doble de características. Posteriormente, serán clasificados por el algoritmo SVM.

5.2.2. Support Vector Machine

Una vez obtenidas todas las características para cada región y cada paciente se continuará con la clasificación mediante SVM. Este método, como ya hemos visto en el anterior diseño, tiene dos fases: entrenamiento y test. Estos han sido delimitados mediante validación cruzada en la cual el resultado final de la clasificación, no se encuentra sesgado a las muestras de clasificación.

- *Entrenamiento*. La principal diferencia con la propuesta anterior radica en el conjunto de entrenamiento. En este caso estará formado por la concatenación de las características más significativas tanto de la clase AD, como de la clase NC, obtenido mediante *sparse filtering*.
- *Test*. En esta fase el clasificador utilizará como parámetros de entradas la concatenación de las características aportadas por ambas clases mediante SF, como se ha dicho, tendremos el doble de características. A priori, debería dar buenos resultados ya que el entrenamiento se realiza con el doble de características que el diseño anterior.

Capítulo 6. Resultados Experimentales

En este capítulo se explicarán los fundamentos teóricos realizados para la evaluación de la clasificación con ambas propuestas de diseño, así como su aplicación en los diseños y la comparativa de ambos.

6.1. Métodos de Evaluación del Clasificador

Hay diferentes medidas de precisión diagnóstica que se relacionan con los diferentes aspectos del procedimiento de diagnóstico. Algunas se utilizan para evaluar la propiedad discriminativa de la prueba y otras para evaluar su capacidad de predicción, siendo esta última muy sensible a las características de las poblaciones evaluadas. Para valorar la calidad del test de diagnóstico es necesario saber cuán buena y confiable es una prueba. Esta cuantificación es llevada a cabo por las medidas estadísticas de sensibilidad, especificidad y precisión. Siendo de gran importancia en el diagnóstico de la enfermedad, por tanto, es imprescindible saber cómo interpretarlas, así como cuándo y bajo qué condiciones usarlas [24][25].

6.1.1. Precisión, sensibilidad, especificidad

La sensibilidad, especificidad y precisión son los términos que miden estadísticamente el rendimiento del test y son independientes de la población sometida al test.

- **Sensibilidad.** La sensibilidad evalúa la eficacia de la prueba para detectar una enfermedad.

$$ESPECIFICIDAD = \frac{TN}{TN + FP} \quad (28)$$

- **Especificidad.** La especificidad estima la probabilidad de que los pacientes sin enfermedad puedan descartarse correctamente.

$$SENSIBILIDAD = \frac{TP}{TP + FN} \quad (29)$$

- **Precisión.** La precisión mide la corrección de una prueba diagnóstica que identifica y excluye una condición dada.

$$PRECISIÓN = \frac{TN}{TN + FP} \quad (30)$$

Donde TP, TN, FP y FN son el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

6.1.2 Curva ROC

La curva ROC (Receiver Operating Characteristic) es una técnica gráfica que nos permite evaluar la precisión del modelo estadístico para clasificar dos clases, AD y NOR. La curva se obtiene calculando la sensibilidad (proporción de resultados positivos verdaderos) y la especificidad del modelo en cada punto de corte posible, y trazando la sensibilidad frente a 1-especificidad (proporción de resultados falsos positivos).

Cada punto en el espacio ROC muestra el equilibrio entre la sensibilidad y la especificidad, es decir, que el aumento de sensibilidad va acompañado de una disminución en la especificidad. Así, la localización de un punto en el espacio ROC representa si la clasificación diagnóstica es buena o no. Siendo una clasificación ideal cuando el punto determinado por la relación entre la sensibilidad y 1-especificidad se encuentra en las coordenadas (0,1), esto indica que la prueba discrimina perfectamente entre AD y NOR. Esta clasificación será aleatoria cuando se encuentra en la diagonal trazada entre los puntos (0,0) y (1,0), de manera que cualquier punto por encima de la diagonal nos dará un buen diagnóstico mientras que si cae por debajo tendríamos un mal diagnóstico. En la Figura 6.1 se puede ver representado lo dicho anteriormente, siendo el “cut-point” el punto de funcionamiento óptimo en la curva ROC.

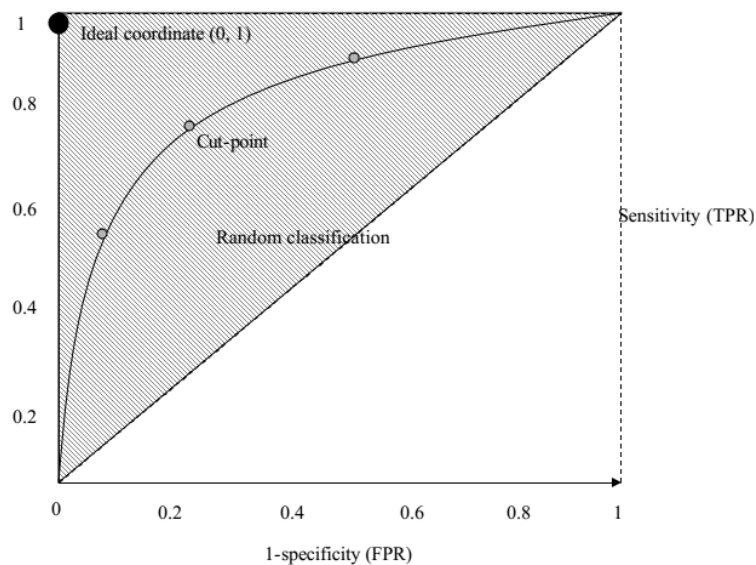


Figura 6.1: Representación de la curva ROC

Cada punto de corte de una prueba de diagnóstico define un único punto en el espacio ROC, los diferentes puntos posibles definen la curva ROC. Esto es análogo a lo dicho para un solo punto, por tanto cuanto más se acerquen los puntos de la curva ROC a la coordenada ideal, más exacta será la prueba y viceversa, cuanto más se aleje peores resultados obtendremos.

El área bajo la curva ROC se denomina AUC y se interpreta como el promedio de precisiones positivas y negativas. Este índice es especialmente útil en los estudio comparativo de pruebas de diagnóstico. Siendo deseable comparar toda la curva ROC en lugar de en un punto particular. Si tiene un valor máximo, $AUC = 1$ significa que la prueba diagnóstica es perfecta en la diferenciación de los sujetos enfermos y los sanos. Esto ocurre cuando la distribución de los resultados de las pruebas no se superponen. Cuando $AUC = 0.5$ la discriminación es casual, la curva se encuentra en la línea diagonal del espacio ROC. El AUC mínimo debe considerarse un nivel de probabilidad, es decir, $AUC = 0.5$, mientras que si $AUC = 0$ significa que la prueba clasifica incorrectamente a todos los sujetos con enfermedad, clasificando mal tanto a todos los sujetos enfermos como sano, esto es extremadamente improbable que suceda en la práctica clínica [26].

6.2. Evaluación del clasificador

Para la obtención de estos resultados se han utilizado todas las imágenes PET de la base de datos ADNI, que incluyen pacientes AD (diagnosticados de Enfermedad de Alzheimer) y NC (control). El procedimiento seguido tiene como objetivo evaluar la clasificación de los pacientes nuevos utilizando como entrada la matriz de características generada por *sparse filtering*. Estos experimentos fueron realizados empleando diferentes parámetros de configuración, uno fue el tamaño de los cubos extraídos de las regiones de interés utilizados como entrada para SF y otro, el número de características a las que converge SF en sus capas, tomando valores desde 10 hasta 100 con paso de 10.

6.2.1. Primer Diseño

Para evaluar el clasificador a partir del primer esquema de diseño, descrito en la Figura 5.6, se realizará el análisis de especificidad, precisión y sensibilidad. Se han elaborado distintos experimentos cambiando tanto el tamaño de los cubos de entrada, como las características utilizadas por la capa L1, Figura 6.2 (izquierda). La precisión observada es levemente mejor para una capa L1 utilizando 80 características, siendo esta precisión similar para los cubos de tamaño $4 \times 4 \times 4$ y $5 \times 5 \times 5$. Aunque gráficamente no se puede apreciar, la presión es levemente mejor para el cubo de tamaño $4 \times 4 \times 4$, por lo que

se fijará este tamaño para el resto de las pruebas realizadas. Por tanto, se tomará como parámetro óptimos para la primera capa de SF, el tamaño de cubo 4x4x4 y una capa L1 de 80 características.

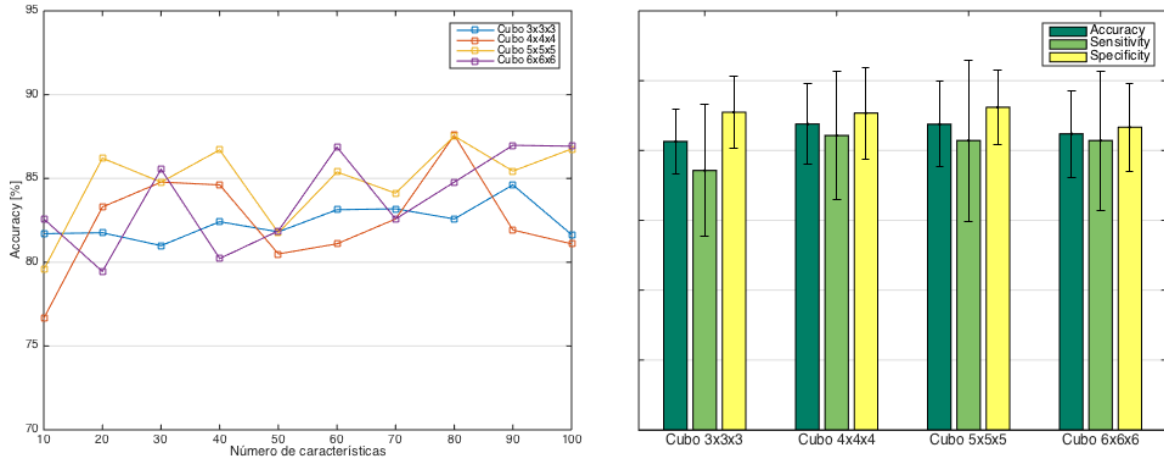


Figura 6.2: Izquierda: Clasificación de accuracy para diferente número de características. Derecha: Precisión, sensibilidad, especificidad para una capa de 80 características y desviación típica entre los valores k de la validación cruzada.

Una vez fijado los parámetros anteriores, se implementará una nueva capa L2 a *sparse filtering*. Para esta segunda capa se evaluará los resultados de la clasificación con los mismo conjuntos de características que antes.

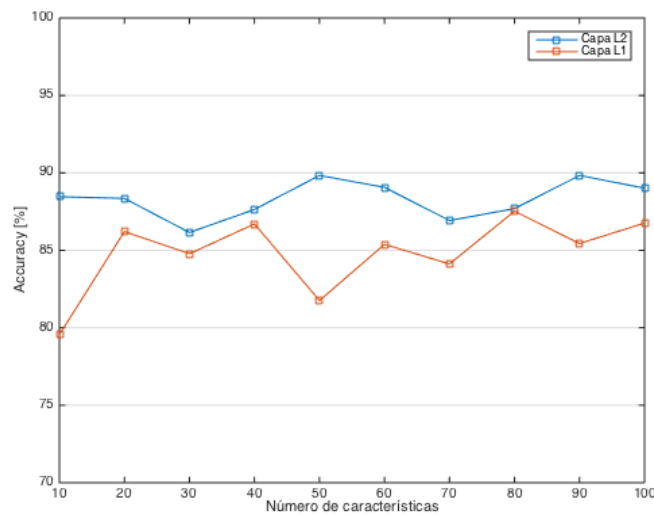


Figura 6.3: Implementación de otra capa L2 a *sparse filtering*.

La implementación de esta segunda capa L2 da mejores resultados que utilizando solamente una capa L1, Figura 6.3. Este resultado era de esperar ya que, como se explico en el capítulo 4, *sparse filtering* en las capas más bajas detectan patrones de comportamiento simples como son las intensidades o los bordes, mientras que en las capas superiores, son capaces de aprender información más compleja. En general siempre es mejor añadir nuevas capas, ya que caracterizará mejor al sistema, aunque esto también significa mayor coste computacional, es decir, si añadimos un número excesivo de capas puede que el tiempo de entrenamiento sea excesivamente grande, por lo que se debe llegar a un compromiso.

6.2.2. Segundo Diseño

Para el segundo diseño se ha entrenado SF con las clases AD y NC de forma independiente como se describía en la Figura 5.4. Esto proporciona dos conjuntos distintos de características, uno para cada clase. Para realizar la clasificación se han concatenado ambos conjuntos de características, esto nos proporciona el doble de características para cada paciente por región.

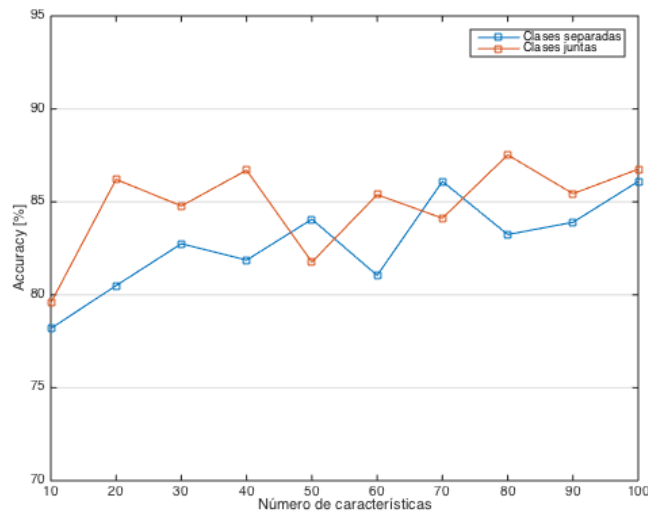


Figura 6.4: Accuracy obtenido para el entrenamiento con clases separadas y juntas utilizando cubos de $4 \times 4 \times 4$.

Dado que se entrena con el doble de características es de esperar que los resultados de la clasificación sean mejores. Sin embargo, aunque para algunos conjuntos de características los resultados de la clasificación, entrenado con ambos conjuntos independientemente, sobresalen, por lo general, la precisión del clasificador es mucho mejor para los conjuntos de entrenamiento de ambas clases conjuntas (Figura 6.4). Esto puede ser debido a que cuando se realiza el entrenamiento entre ambas clases de forma con-

junta, SF “aprende” características que permitirán diferenciar una de la otra, mientras que cuando entrena con las clases separadas “aprende” características que definen a esa clase solamente.

6.2.3. Curva ROC

Para saber como de bueno es el diagnóstico calculamos la curva ROC expresada como la sensibilidad y especificidad del diagnóstico, donde la sensibilidad será la probabilidad de que un paciente AD sea clasificado como AD mientras que la especificidad será la probabilidad de que un paciente NOR sea clasificado como NOR, por tanto $1 - \text{Especificidad}$ será la probabilidad de que un paciente NOR sea clasificado como AD. La curva ROC estará en función de la sensibilidad y $1 - \text{especificidad}$. En la Tabla 5.3 queda patente la efectividad del método para clasificar correctamente a partir de las características extraídas por SF, volviendo a constatar que para el tamaño de $4 \times 4 \times 4$ se obtiene una clasificación más precisa.

La curva ROC ha sido representada para evaluar la clasificación de los pacientes NC/AD tomando como conjunto de características una capa L1 de 80 características, Figura 6.5. El punto óptimo de funcionamiento (cut-off point), varía dependiendo de nuestros intereses. En nuestro caso el umbral del modelo a alcanzar, trata de maximizar los AD que son clasificados como AD, minimizando los AD clasificados como NOR.

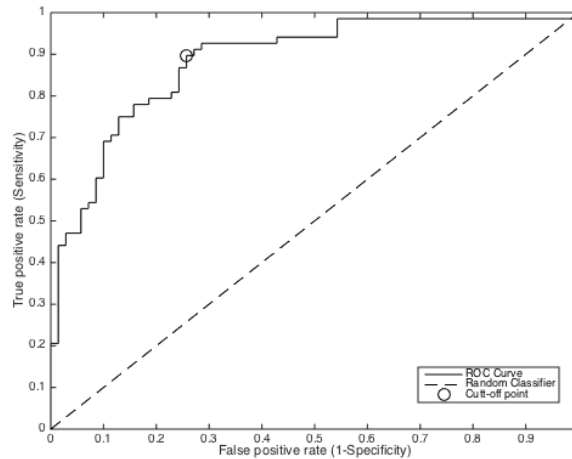


Figura 6.5: Curva ROC para clasificación NC/AD utilizando capa L1

| Cubo | Accuracy | Sensitivity | Specificity | AUC |
|-------|---------------|---------------|---------------|--------|
| 3x3x3 | 0.8258±0.0930 | 0.7429±0.1881 | 0.9095±0.1034 | 0.8397 |
| 4x4x4 | 0.8758±0.1154 | 0.8429±0.1838 | 0.9071±0.1311 | 0.8845 |
| 5x5x5 | 0.8753±0.1229 | 0.8286±0.2313 | 0.9238±0.1052 | 0.7660 |
| 6x6x6 | 0.8478±0.1242 | 0.8286±0.1998 | 0.8667±0.1261 | 0.6655 |

Tabla 6.1: Resultados de clasificación. Valor medio ± Desviación típica

El mejor resultado de AUC se obtiene para los cubos de tamaño 4x4x4, Tabla 6.1, por tanto ese será el tamaño elegido para implementar una segunda capa, L2. La implementación de esta segunda capa da la misma precisión para dos conjuntos de características, L2=50 y L2=90, Figura 6.3. Para calcular la curva ROC, Figura 6.6, hemos elegido el conjunto de características 50 porque daba mejores resultados. Los resultados obtenidos para una segunda capa son mucho mejores que para una sola capa, Tabla 6.2, alcanzando una precisión de hasta un 90% y un AUC de 0.9384, siendo AUC=1 un sistema perfecto.

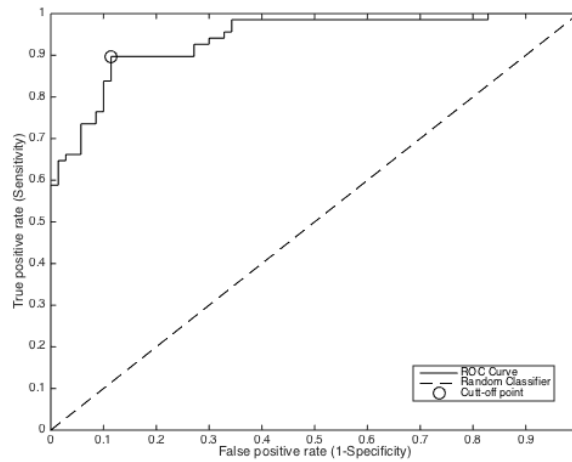


Figura 6.6: Curva ROC para clasificación NC/AD utilizando capa L2

| Accuracy | Sensitivity | Specificity | AUC |
|---------------|---------------|---------------|--------|
| 0.8984±0.1130 | 0.8714±0.1622 | 0.9262±0.1543 | 0.9384 |

Tabla 6.2: Resultados de clasificación. Valor medio ± Desviación típica capa L2

Basándonos en los resultados obtenidos podemos concluir que los hiperparámetros óptimos que nos ofrecen una clasificación más precisa serán los cubos de tamaño 4x4x4, una capa L1 de 80 características y otra capa L2 de 50 características para *sparse filtering*.

Los resultados de la clasificación evaluados mediante validación cruzada demuestran que el método propuesto es válido para diferenciar entre los sujetos controles (NC) y los sujetos AD, incluso podría dar buenos resultados con sujetos MCI (deterioro cognitivo leve).

Capítulo 7. Conclusiones y trabajos futuros

El desarrollo de este proyecto ha tenido como finalidad la obtención de unos parámetros óptimos que permitan obtener las características de mayor relevancia para la enfermedad de Alzheimer y que por tanto ayuden a la clasificación de los pacientes que sufre dicha enfermedad. Para ello se hicieron uso de las imágenes de topografía por emisión de positrones (PET) de las regiones más relevantes para la enfermedad. El método se ha aplicado sobre una base de datos de 138 sujetos compuestas por dos clases distintas: Pacientes diagnosticados de la enfermedad de Alzheimer y pacientes sanos. Después del procesado y registrado de las imágenes, se realiza la extracción de las características más relevantes para la enfermedad mediando un algoritmo de aprendizaje no supervisado, *sparse filtering*. Esta técnica se combina con un método de clasificación supervisada basado en *support vector machine*, alcanzando valores de precisión de hasta el 87% para una sola capa y de casi un 90% implementando una segunda capa. Mientras que las técnicas que usan los voxels como características (por ejemplo, la técnica VAF [27][23], *Voxels as features*) no consiguen más de un 87% de accuracy ni valores de AUC superiores a 0.89. Por lo tanto, las características que calculamos, son más discriminantes que los voxels directamente. Otra de las pruebas realizada ha sido la clasificación utilizando ambas clases (pacientes con Alzheimer y sanos) separadas en el entrenamiento de las capas. En la clasificación de esta segunda prueba se obtienen resultados en torno al 86% para un número de cubos y de características concretas.

En líneas futuras, se podrían implementar nuevas capas a *sparse filtering* que permitan lograr características aun más representativas, así como hacer uso de otros clasificadores más discriminantes. Además, este método utiliza el aprendizaje supervisado para el entrenando del clasificador, lo cual podría ser un problema cuando se encuentran muestras de entrenamiento sin etiquetadas o las etiquetas no son lo suficientemente fiables para usarlas como verdad fundamental. También se podría utilizar una fusión de las imágenes MRI y PET, ya que hay estudios en los que queda demostrado que la fusión de ambas da mejores resultados que con el estudio de cada una por separado [22].

Bibliografía

- [1] Martín Carrasco M., Bulbena Vilarrasa A., Diagnóstico precoz de la enfermedad de Alzheimer. *Psiquiatría Biológica* 2003; 10(4): 119-132.
- [2] Khachaturian ZS. Diagnosis of Alzheimer's disease. *Arch Neurol* 1985; 42:1097–1105.
- [3] Petrella J.R, Coleman R.E., Doraiswamy P.M., Neuroimaging and Early Diagnosis of Alzheimer Disease: A Look to the Future. *Radiology* 2003; 226:315–336.
- [4] Ortiz, A.; Fajardo, D.; Górriz, J.M.; Ramírez, J.; Martínez-Murcia, F.J.: "Multimodal image data fusion for Alzheimer's Disease diagnosis by sparse representation". *International Conference on Innovation in Medicine and Healthcare (InMed)*, 2014.
- [5] Alzheimer's Association. 2016 Alzheimer's Disease Facts and Figures, Alzheimer's & Dementia
- [6] Folstein, M.F., Folstein, S.E., McHugh, P.R. "mini-mental state". a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res.*, 12:189-198, 1975
- [7] Gauthier S., *Clinical Diagnosis and Management of Alzheimer's Disease*, 1996.
- [8] www.hipocampo.org
- [9] Christensen DD., *Practical Principles for the Management of Alzheimer's Disease, Primary Care Companion J Clin Psychiatry* 2002; 4(2).
- [10] Alzheimer's Association www.alz.org
- [11] ADNI initiative www.adni-info.org
- [12] Chaves Rodríguez, R., 2013. Un nuevo modelo de la interconectividad funcional cerebro mediante reglas de asociación aplicado a la detección de alteraciones neurológicas. Universidad de Granada.
- [13] Martí-Climent J.M., Prieto E., López Lafuente J. y Arvizu J. Neuroimagen: Fundamentos técnicos y prácticos. *Rev Esp Med Nucl.* 29(4):189–210, 2010.
- [14] Tarantola, G., Zito, F., Gerundini, P., 2003. Pet instrumentation and reconstruction algorithms in whole-body applications. *Journal of Nuclear Medicine* 44 (5), 756–769.
- [15] www.nifti.nimh.nih.gov
- [16] Ngiam J, Chen Z, Bhaskar SA, Koh PW, Ng AY. Sparse filtering. In: *Advances in neural information processing systems*, pp 1125–1133, 2011.

- [17] Massimo Zennaro F., Chen K., Towards Understanding Sparse Filtering: A Theoretical Perspective. School of Computer Science. The University of Manchester, M13 9PL, UK.
- [18] C. Cortes and V. Vapnik. Support Vector networks. *Machine Learning*, 20:273–297, 1995.
- [19] Williams C., Support Vector Machines, School of Informatics, University of Edinburgh, 2008.
- [20] Zhu X., Support Vector Machines. *Advance Natural Language Processing*, University of Wisconsin, 2010.
- [21] Hatie T., Tibshirani R., Friedman J., The elements of statistical learning, data mining, inferencen, and prediction. Springer, 2008.
- [22] Ortiz, A.; Munilla, J., Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer’s Disease. Universidad de Málaga, 2016.
- [23] Ortiz, A.; Munilla, J.; Illán, I.; Górriz, J.M. Ramírez, J.: "Exploratory graphical models of functional and structural connectivity patterns for Alzheimer's Disease diagnosis". *Frontiers in Computational Neuroscience*, 9:132. doi: 10.3389/fncom.2015.00132.
- [24] Wen Zhu, Nancy Zeng, Ning Wang. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations, 1K&L consulting services, Inc, Fort Washington, PA Octagon Research Solutions, Wayne, PA, 2010.
- [25] Hwee Bee Wong, Gek Hsiang Lim. *Measures of Diagnostic Accuracy: Sensitivity, Specificity, PPV and NPV*, Health Services Research and Evaluation Division, Ministry of Health, Singapore.
- [26] Hajian-Tilaki K., Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*. 2013 Spring; 4(2): 627–635.
- [27] Illán, I.; Gorrina J.M.; López M.M.; Ramírez J.; Salas-Gonzalez D.; Segovia F.; Chaves R.; Puntonet C.G.: “Computer aided diagnosis of Alzheimer’s disease using component based SVM”. *Applied Soft Computing* 11 (2011) 2376–2382.