

# Deteccción precoz de la enfermedad de Alzheimer mediante clasificación automática de imágenes estructurales de resonancia magnética con SVM

---

**Proyecto de Fin de Máster para la obtención del título de Máster en Física Médica**

## **Autor**

David Álvarez Llorente

## **Directores de Proyecto**

Dra. Cristina Santa Marta Pastrana – Departamento de Física Matemática y de Fluidos (Facultad de Ciencias)

Dr. Mariano Rincón Zamorano – Departamento de Inteligencia Artificial (ETSI Informática)

Diciembre 2012

Universidad Nacional de Educación a Distancia

España



## Tabla de contenido

Índice de figuras .....	5
Índice de tablas .....	5
Resumen.....	7
Introducción .....	9
Enfermedad de Alzheimer.....	9
Máquinas de vectores soporte (SVM) y aprendizaje automático.....	10
Árboles de decisión binarios .....	12
Objetivo.....	15
Materiales .....	17
Método.....	19
Procesado de imágenes .....	19
Extracción de características.....	21
Selección de características .....	23
Entrenamiento de la SVM .....	24
Evaluación del rendimiento .....	27
Resultados .....	29
Discusión de los resultados .....	39
Primera fase de pruebas (80 casos) .....	39
Segunda fase de pruebas (414 casos) .....	40
Conclusiones .....	41
Bibliografía .....	43



## Índice de figuras

<i>Figura 1: Principio de funcionamiento de una SVM lineal.</i> .....	10
<i>Figura 2: Uso de un kernel para obtener una frontera no lineal.</i> .....	11
<i>Figura 3: Ejemplo de árbol de decisión</i> .....	13
<i>Figura 4: Plantilla de parcelación</i> .....	18
<i>Figura 5: Fases de procesado de las imágenes</i> .....	21
<i>Figura 6: Asignación de vóxeles de las imágenes normalizadas.</i> .....	22
<i>Figura 7: Fracción de aciertos de un clasificador entrenado con 80 casos.</i> .....	25
<i>Figura 8: Comparación de sensibilidades obtenidas con 80 casos</i> .....	35
<i>Figura 9: Comparación de especificidades obtenidas con 80 casos</i> .....	35
<i>Figura 10: Comparación de puntuaciones F1-score obtenidas con 80 casos</i> .....	36
<i>Figura 11: Comparación de sensibilidades obtenidas con 414 casos</i> .....	36
<i>Figura 12: Comparación de especificidades obtenidas con 414 casos</i> .....	37
<i>Figura 13: Comparación de puntuaciones F1-score obtenidas con 414 casos</i> .....	37

## Índice de tablas

<i>Tabla 1: Resultados obtenidos empleando como características las cantidades relativas de materia gris de 116 regiones del cerebro con 80 casos demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del número de aciertos obtenido.</i> .....	30
<i>Tabla 2: Resultados obtenidos empleando como características las cantidades relativas de materia gris de 116 regiones del cerebro con 80 casos demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del F1 score obtenido.</i> .....	31
<i>Tabla 3: Resultados obtenidos empleando como características los volúmenes relativos de 116 regiones del cerebro con 80 casos demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del número de aciertos obtenido.</i> .....	32
<i>Tabla 4: Resultados obtenidos empleando como características los volúmenes relativos de 116 regiones del cerebro con 80 casos demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del F1 score obtenido.</i> .....	33
<i>Tabla 5: Resultados obtenidos empleando 414 casos de características demográficas heterogéneas y distintos conjuntos de características y criterios de optimización.</i> .....	34



## **Resumen**

Este trabajo estudia la capacidad de una máquina de vectores soporte para diagnosticar de forma automática si un paciente sufre la enfermedad de Alzheimer en su estado inicial a partir de imágenes estructurales del cerebro obtenidas mediante resonancia magnética.

Para ello se emplearon los estudios de 414 individuos (tanto enfermos como controles sanos) disponibles en el proyecto OASIS. Todas las imágenes fueron segmentadas para distinguir materia blanca y materia gris y normalizadas a un mismo espacio (plantilla MNI) para poder compararlas entre sí. Las características que se utilizaron para describir a cada paciente fueron las cantidades relativas de materia gris y el volumen de las distintas regiones del cerebro (identificadas mediante una plantilla del Atlas de Talairach adaptada a las dimensiones a las que estaban normalizadas las imágenes).

Se comprobó el rendimiento del sistema entrenándolo con distintos conjuntos de características: empleando sólo materia gris, sólo volúmenes relativos, utilización de todas las características o únicamente un subconjunto de ellas elegido en base dos métodos diferentes: árboles de decisión y literatura clínica.

El mejor rendimiento del sistema se obtuvo empleando únicamente los volúmenes de aquellas regiones que, según la bibliografía clínica, sufren un deterioro más rápido debido a la enfermedad. La sensibilidad y especificidad obtenidas en ese caso fueron superiores al 80% en caso de emplear un subconjunto de 80 pacientes de características demográficas similares. Sin embargo, el sistema presentó una sensibilidad muy inferior en caso de ser entrenado con una población heterogénea (414 individuos de un amplio rango de edades y ambos sexos). Las cantidades de materia gris de las distintas regiones del cerebro fueron en esa ocasión las características que ofrecieron un mejor resultado.



## **Introducción**

El diagnóstico asistido por ordenador (CAD) y, en menor medida, el diagnóstico automático por ordenador, son tecnologías que han despertado un importante interés investigador en los últimos años y que progresivamente se están incorporando a los procedimientos de trabajo en los centros sanitarios. Un software de CAD realiza un procesado automático de las imágenes diagnósticas de un paciente, tratando de localizar indicios de una determinada patología. El objetivo de un sistema CAD es ofrecer al médico una segunda opinión que complemente a su propio diagnóstico, de modo que el resultado final sea mejor que el que habría obtenido cada uno de ellos por separado. La utilidad de estos sistemas se pone de manifiesto especialmente en aquellos casos de diagnóstico dudoso para el especialista.

Los primeros sistemas CAD estuvieron orientados a la detección a partir de imágenes radiológicas de lesiones vasculares, nódulos pulmonares y grupos de microcalcificaciones en mamografías [Doi2007]. En la actualidad se investiga la posible aplicación del CAD a otros ámbitos, como la detección de patologías neurológicas a partir de imágenes del cerebro.

Este trabajo estudia la posibilidad de emplear un algoritmo de aprendizaje automático que, una vez entrenado mediante casos de ejemplo, sea capaz de detectar la enfermedad de Alzheimer en sus etapas iniciales a partir de imágenes de resonancia magnética del cerebro del paciente. Entre los trabajos que han explorado anteriormente esta estrategia pueden citarse los siguientes: [Magnin2009], [Savio2010], [Magalhaes2010].

### ***Enfermedad de Alzheimer***

La enfermedad de Alzheimer es un problema sanitario creciente en las sociedades desarrolladas debido al aumento de la esperanza de vida. La enfermedad se manifiesta como un deterioro cognitivo que en un primer momento afecta a la memoria y progresivamente llega a incapacitar totalmente al paciente.

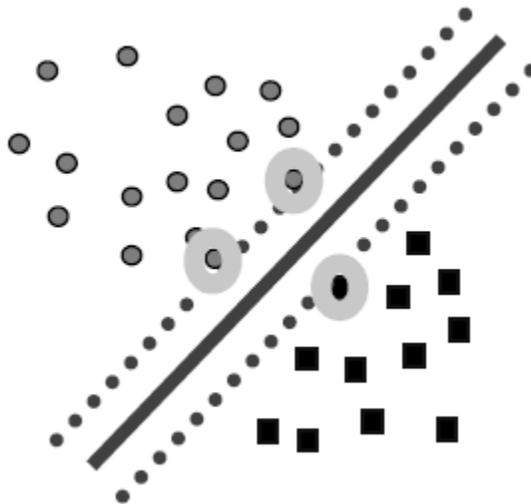
A nivel neurológico la enfermedad produce cambios degenerativos en áreas concretas del cerebro, entre las que se incluyen los lóbulos temporales y parietales y algunas zonas del giro cingulado [Wenk2003], [Busatto2008]. El proceso se inicia con la acumulación de proteína amiloide, que produce inflamación entorno a las placas seniles. A partir de aquí se desencadena un proceso que en último término produce la muerte de un gran número de neuronas.

En la actualidad se aceptan como marcadores diagnósticos de la enfermedad en imagen de resonancia magnética la atrofia de estructuras mediales de los lóbulos temporales en la etapa inicial de deterioro cognitivo leve. Asimismo, puede predecirse el paso de un deterioro cognitivo leve a enfermedad de Alzheimer a partir de las alteraciones en el hipocampo y el córtex entorrinal. La evaluación visual de las imágenes permite al especialista detectar pacientes de Alzheimer con una sensibilidad/especificidad del 80-85% [Frisoni2010]. Estos valores son ligeramente inferiores en la etapa inicial de la enfermedad (déficit cognitivo leve).

## ***Máquinas de vectores soporte (SVM) y aprendizaje automático***

Las *máquinas de vectores soporte* (SVM) son un avance relativamente reciente en el campo de los algoritmos de aprendizaje automático, que constituyen una disciplina en auge dentro de la inteligencia artificial.

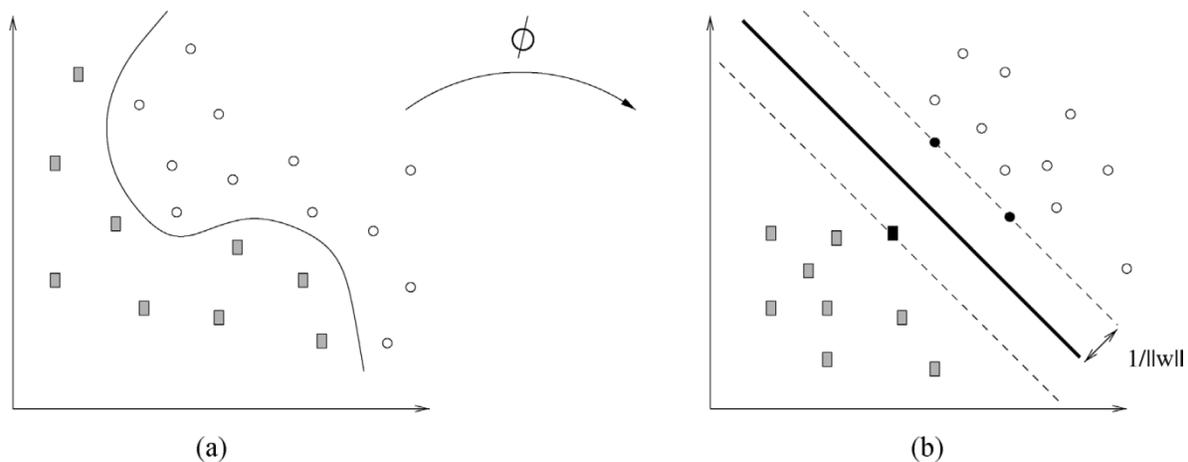
Al igual que otras técnicas más conocidas como las redes neuronales, las SVM son capaces de *aprender* a través de ejemplos una función o relación matemática que no se conoce *a priori* o que no es posible representar explícitamente en forma de ecuación matemática (típicamente una clasificación de muestras en dos o más posibles categorías). Dichos ejemplos, representados por conjuntos de números que codifican sus *características*, son presentados al sistema en una fase inicial de entrenamiento, proporcionándole además la categoría a la que corresponde cada uno de ellos. Idealmente, una vez completada esta fase la SVM debería ser capaz de asignar la categoría correcta a cualquier muestra que le sea presentada, aunque no forme parte del conjunto con el que fue entrenada. En la práctica, la tasa de aciertos de un clasificador depende de múltiples factores, como la cantidad de información útil que contengan las muestras (fundamental en el caso de este proyecto), la idoneidad del conjunto de entrenamiento, o la capacidad del sistema para encontrar una frontera en el espacio de características de las muestras que separe completamente las categorías.



**Figura 1:** Principio de funcionamiento de una SVM lineal. La figura muestra un conjunto de muestras pertenecientes a dos categorías diferentes (círculos y cuadrados) en un espacio de 2 dimensiones. La línea azul representa la frontera óptima de separación entre ambas categorías hallada durante el entrenamiento de la SVM, dado que ofrece el máximo margen entre ellas (distancia entre las dos líneas de puntos). Los vectores de soporte que determinan dicho margen son las características rodeadas por círculos. Fuente: [Bennet2000]

El proceso de entrenamiento consiste en optimizar una función de coste de modo que se maximice el *margen* que separa ambas categorías, determinando cuáles son las muestras (vectores de características) que limitan dicho margen. Estos vectores críticos reciben el nombre de *vectores soporte* o *de apoyo* porque el margen de separación “se apoya” en ellos (ver figura 1) y dan nombre a esta técnica.

En su forma básica, la SVM es un clasificador lineal, por lo que únicamente puede discriminar categorías linealmente separables en el espacio de características. La frontera entre las categorías será por lo tanto un hiperplano en dicho espacio. Sin embargo, mediante una técnica conocida como *kernel trick* es posible obtener fronteras no lineales entre categorías. Esta técnica consiste en realizar una transformación no lineal de los vectores de características (que no son linealmente separables en el espacio de características original) y trasladarlos a un espacio de características de dimensionalidad más elevada en el que sí puedan ser separados linealmente.



**Figura 2:** Uso de un kernel para obtener una frontera no lineal. Muestras de 2 categorías no separables linealmente, con una posible frontera no lineal (a); Transformando las características a un espacio diferente mediante un kernel no lineal, ambos conjuntos son linealmente separables en el espacio transformado. En este nuevo espacio puede emplearse una SVM para determinar la frontera de separación entre ambas categorías (b). Fuente: [Moguerza2006]

Existen múltiples *kernels* (funciones de transformación) posibles, siendo el más habitual el de *funciones base radiales* (RBF). Este kernel tiene dos parámetros ajustables cuyos valores deben elegirse para optimizar su funcionamiento en cada problema en particular.

Es posible construir clasificadores de categorías múltiples con SVMs recurriendo a distintas estrategias: conectar SVMs *en cascada* que distingan conjuntos de categorías cada vez más pequeños [Guo2007], o entrenar una SVM para distinguir cada una de las categorías de todas las demás.

Las SVM presentan una serie de ventajas frente a otros algoritmos de aprendizaje automático, entre las que destacan:

- **Inexistencia de mínimos locales en la función de coste** que se optimiza durante el entrenamiento, lo cual impide que el sistema termine su entrenamiento *aprendiendo* una frontera no óptima para clasificar el conjunto de entrenamiento utilizado. El proceso de optimización de las SVM equivale a optimizar un problema cuadrático (convexo), por lo que el único mínimo es el mínimo absoluto de la función de coste.
- **Las SVM son robustas frente al problema del *sobreentrenamiento***. Este fenómeno se produce cuando el sistema no es capaz de generalizar adecuadamente a partir de los ejemplos que le son presentados, y se limita a memorizar el conjunto de entrenamiento. De este modo, el sistema parece clasificar de forma muy eficiente los ejemplos con los que fue entrenado pero presenta una elevada tasa de fallos con el resto de casos. Este problema se presenta frecuentemente cuando se emplean redes neuronales.

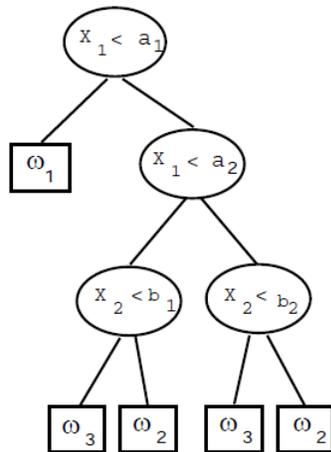
Para una descripción rigurosa de los fundamentos teóricos de las SVM y de los algoritmos que permiten su implementación puede recurrirse a las siguientes referencias bibliográficas: [Vapnik1998], [Burges1998], [Moguerza2006], [Fletcher2009], [Platt1998].

### ***Árboles de decisión binarios***

En este trabajo se hace uso de los denominados *árboles de decisión binarios* para la selección de características, por lo que se incluye a continuación una breve descripción de los mismos.

Un árbol de decisión binario es un algoritmo de clasificación jerárquico. Se basa en la idea de descomponer una decisión compleja (la pertenencia de una muestra a una determinada categoría) en una serie de decisiones más simples. En concreto, el árbol se compone de una serie de nodos interiores (de decisión), cada uno de los cuales representa una decisión tomada en base el valor de una variable, y tiene exactamente dos nodos hijos. Los nodos finales (denominados *hojas*) representan las categorías que el árbol asigna a las muestras en función de las decisiones tomadas en los nodos que llevan hasta cada una de ellas.

El paso previo a la utilización de este clasificador es la creación del árbol de decisión a partir de un conjunto de muestras de entrenamiento. A través de sucesivas iteraciones, se busca aquella variable cuyo valor permita dividir el conjunto de muestras en dos subconjuntos lo más homogéneos posible. Como medidas de dicha homogeneidad pueden utilizarse el *índice de Gini* o la *entropía* ([Breiman1984], [Bittencourt2004]). Este proceso se repite con los sucesivos subconjuntos hasta que se logra una clasificación correcta de todas las muestras o se alcanza un umbral de separación que se considere aceptable.



**Figura 3:** Ejemplo de árbol de decisión. Los valores de las variables  $X_1$  y  $X_2$  asociados a una muestra particular son utilizados en los nodos interiores (óvalos) para decidir en qué categoría ( $\omega_1$ ,  $\omega_2$  o  $\omega_3$ ) se clasifica dicha muestra. Las posibles decisiones son representadas por los nodos hoja (cuadrados).

Fuente: [Safavian1991]

Los árboles de decisión presentan las siguientes ventajas

- Representan las razones que han llevado a una determinada decisión de clasificación de una forma fácilmente comprensible, al contrario que otras técnicas como las redes neuronales o las máquinas de vectores soporte.
- Son una técnica sencilla para reducir el número de variables iniciales a aquellas más relevantes para la clasificación. Además muestran la importancia relativa de cada variable, ya que aquellas que tengan mayor poder de discriminación estarán más cerca del nodo raíz.

Esta técnica también presenta algunas desventajas, entre las que cabe destacar las siguientes [Safavian1991]:

- En aquellos conjuntos de muestras en los que exista una notable superposición de características entre categorías, el número de nodos hoja puede llegar a ser mucho mayor que el número de categorías posibles, aumentando el tiempo de búsqueda y los requisitos de memoria.
- Encontrar el árbol óptimo para un problema dado puede ser difícil, ya que existen multitud de árboles posibles para cada problema, con el agravante de que la eficiencia de clasificación del sistema depende críticamente del diseño de dicho árbol.



## Objetivo

La finalidad de este trabajo es evaluar la eficacia de las *máquinas de vectores soporte* (SVM) para el diagnóstico automatizado de la enfermedad de Alzheimer en estadios leve o muy leve a partir de imágenes del cerebro del paciente. En concreto se desea evaluar el resultado empleando imágenes estructurales 3D de resonancia magnética ponderadas en T1.

Para lograr el fin propuesto, se han de conseguir dos objetivos diferentes:

- Descubrir qué conjuntos de características extraídas de las imágenes son óptimos de cara a la clasificación de un sujeto como sano o enfermo. Para ello se compara el resultado obtenido empleando características correspondientes a todas las regiones del cerebro, características elegidas en base a los conocimientos actuales sobre la evolución de la enfermedad, o características elegidas en base a *árboles de decisión*.
- Encontrar cuál es la mejor estrategia de entrenamiento de la SVM para este problema en particular. Esto incluye tanto el algoritmo de búsqueda de los parámetros óptimos de la SVM como la elección de las características demográficas de los casos y controles que se utilicen para el entrenamiento.



## Materiales

Para la realización del estudio se emplearon los siguientes medios:

- **Imágenes anatómicas de cráneo del proyecto OASIS<sup>1</sup> [Marcus2007] obtenidas mediante resonancia magnética con ponderación en T1.** Cada uno de los estudios consiste en varias exploraciones (3 ó 4) realizadas en una única sesión en idénticas condiciones, a fin de obtener una mejor relación señal-ruido. Además de las imágenes adquiridas se incluye para cada paciente una imagen promediada de todas las exploraciones, debidamente corregistradas y corregidas por artefactos de movimiento [Marcus2007a]. Todas las imágenes tienen unas dimensiones de 256x256x160 vóxeles de 1x1x1mm. Estas imágenes promediadas fueron tomadas como punto de partida para la realización del estudio. La homogeneidad de la técnica de adquisición de los estudios supone una ventaja frente a otros proyectos como ADNI<sup>2</sup>, que está compuesto por imágenes adquiridas en diferentes equipos y empleando diferentes secuencias.

Se emplearon un total de 414 estudios, de los cuales 314 corresponden a individuos sanos y 100 a individuos diagnosticados clínicamente de enfermedad de Alzheimer en grado muy leve o leve según la escala CDR [Morris1993]. Todos los pacientes diagnosticados tenían 60 años o más, mientras que los individuos sanos pertenecían a todos los rangos de edad por encima de 18 años.

- **Entorno de programación Matlab<sup>3</sup>** para el desarrollo de las rutinas de software necesarias.
- **Librería de programación de SVM *libSVM*<sup>4</sup>**[Chang2011]. Este paquete de software se encuentra disponible en Internet de forma gratuita y es de uso común en trabajos de investigación con SVM. Las rutinas del paquete pueden ser utilizadas directamente desde el entorno de Matlab, y permiten tanto entrenar una SVM a partir de un conjunto de muestras como clasificar nuevas muestras con la SVM una vez que se ha completado dicho entrenamiento.
- **Paquete de software estadístico SPM8<sup>5</sup>.** Aunque la aplicación está orientada a la realización de mapas estadísticos paramétricos (que no se utilizan en este estudio) el paquete incluye rutinas de normalización espacial y segmentación de tejidos que fueron utilizadas como paso previo a la extracción de características de la imagen.

---

<sup>1</sup> OASIS: Open Access Series of Imaging Studies

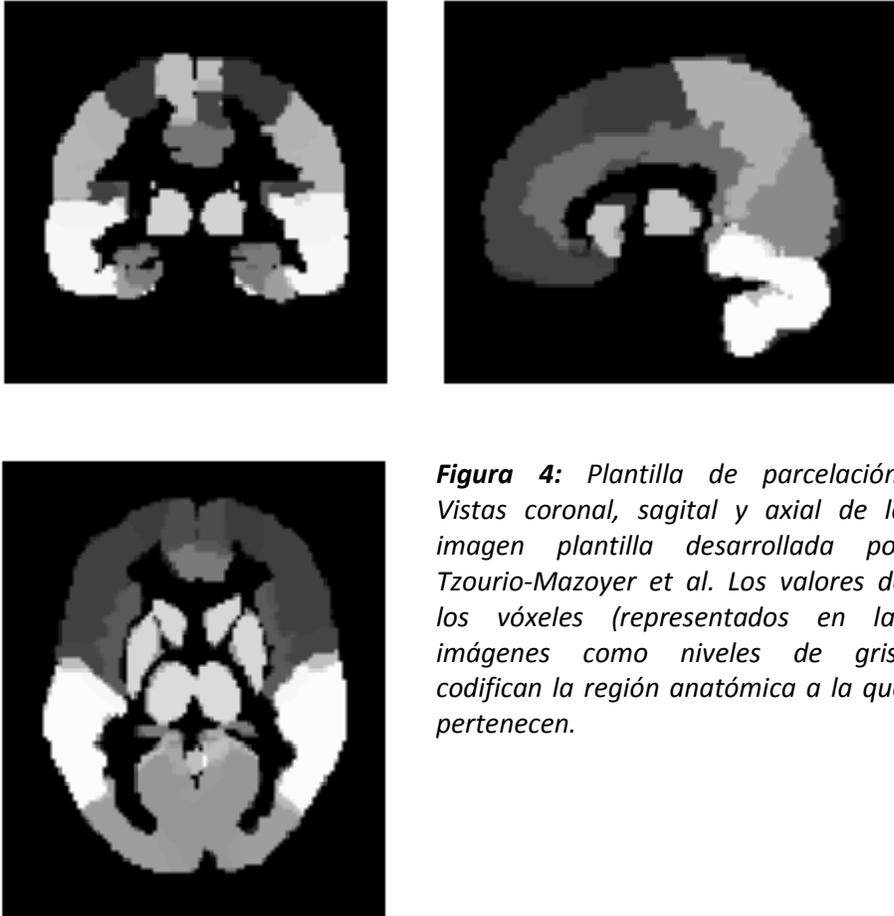
<sup>2</sup> ADNI: Alzheimer's Disease Neuroimaging Initiative. <http://adni.loni.ucla.edu/>

<sup>3</sup> <http://www.mathworks.com>

<sup>4</sup> <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

<sup>5</sup> <http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>

- **Plantilla de parcelación del cerebro en áreas anatómicas** desarrollada por Tzourio-Mazoyer *et al* [Tzourio2002]. Esta plantilla fue creada tomando como punto de partida el Atlas de Talairach [Talairach1988], y se encuentra disponible como parte de la aplicación *Anatomical Automatic Labeling*<sup>6</sup>. A partir de ella, es posible asignar cada voxel de la plantilla MNI a una región anatómica del cerebro.



**Figura 4:** *Plantilla de parcelación. Vistas coronal, sagital y axial de la imagen plantilla desarrollada por Tzourio-Mazoyer et al. Los valores de los vóxeles (representados en las imágenes como niveles de gris) codifican la región anatómica a la que pertenecen.*

---

<sup>6</sup> [http://www.cyceron.fr/web/aal\\_\\_anatomical\\_automatic\\_labeling.html](http://www.cyceron.fr/web/aal__anatomical_automatic_labeling.html)

## Método

La metodología seguida en el estudio se divide en las siguientes etapas:

- **Procesado inicial de las imágenes.** Los estudios de todos los pacientes fueron sometidos a un proceso de normalización espacial para permitir su intercomparación. Así mismo, las imágenes fueron segmentadas para identificar los vóxeles correspondientes a materia gris, materia blanca y líquido cefalorraquídeo.
- **Extracción de características.** Los vóxeles de cada una de las imágenes segmentadas fueron asignados a distintas regiones del cerebro en base a una plantilla preestablecida, de modo que se contabilizase la cantidad de materia gris y el volumen (suma de materia gris y materia blanca) de un total de 116 regiones anatómicas del cerebro de cada paciente. Los vóxeles clasificados como líquido cefalorraquídeo sólo fueron utilizados en el cálculo del volumen intracraneal total de cada sujeto.
- **Selección de características y entrenamiento de la SVM.** Se comprobó la tasa de aciertos obtenida al entrenar la SVM con distintos conjuntos de características y distintas estrategias de optimización. En primer lugar se emplearon un total de 80 casos demográficamente homogéneos, utilizando todas las características o seleccionando subconjunto de ellas en base a distintos criterios. Posteriormente se amplió el estudio a 414 casos demográficamente heterogéneos, empleando sólo aquellas características que ofrecieron mejor resultado en la etapa anterior.
- **Evaluación del rendimiento.** La capacidad predictiva con cada uno de los conjuntos de características se evaluó entrenando el algoritmo de forma sucesiva con todos los casos excepto uno, y comprobando el resultado obtenido al presentarle el caso no utilizado (técnica conocida como *leave-one-out*). Para cada conjunto de características se calculó la sensibilidad y especificidad obtenidas, así como su *F1-score*<sup>7</sup>.

A continuación se detalla cada una de las etapas indicadas.

### ***Procesado de imágenes***

Las imágenes fueron procesadas mediante la aplicación SPM8 en la plataforma Matlab. En primer lugar se modificó su orientación original para hacerla coincidir con las plantillas de normalización que utiliza SPM8. A continuación se ejecutó la rutina de segmentación de imágenes cerebrales en materia gris, materia blanca y líquido cefalorraquídeo que incorpora este software. Esta rutina es una implementación del algoritmo descrito en [Ashburner2005].

---

<sup>7</sup> El *F1-score* es una valoración simultánea de la sensibilidad y especificidad en una única cifra y viene dado por la expresión

$$F1 - score = \frac{2 \cdot sensibilidad \cdot precisión}{sensibilidad + precisión}$$

El algoritmo parte de una plantilla con la distribución de materia gris promedio de un conjunto de individuos, que se utiliza como probabilidad *a priori* de que el mismo vóxel en el sujeto de estudio corresponda a materia gris. La probabilidad *a posteriori* se obtiene mediante la *regla de Bayes* de probabilidad condicional, empleando los valores de la plantilla y los de la imagen del sujeto corregistrada con la plantilla. Por lo tanto, en la determinación del tipo (materia gris, materia blanca o líquido cefalorraquídeo) que corresponde a cada voxel se tiene en cuenta tanto los valores de la plantilla como los de la imagen del paciente

Dado que el cálculo de la transformación necesaria hace uso de la segmentación y viceversa, el algoritmo obtiene ambas cosas simultáneamente en un proceso iterativo de optimización. El hecho de emplear únicamente la materia gris para el proceso de normalización se debe a que de este modo se reduce la variabilidad entre sujetos debida a estructuras ajenas al cerebro.

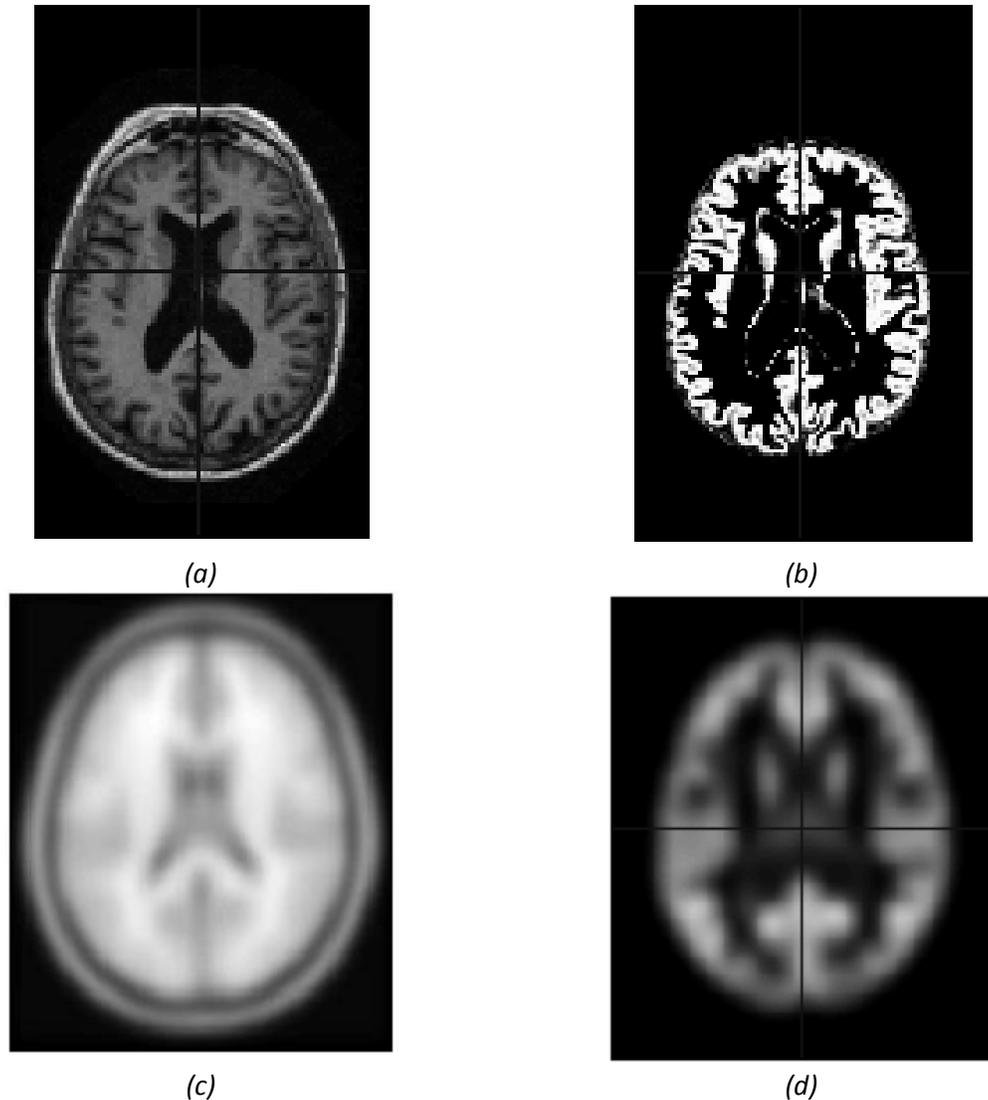
Una vez segmentadas las imágenes se procedió a normalizarlas espacialmente, de forma que cada uno de los vóxeles de las imágenes finales se correspondiese con la misma región del cerebro en todos los casos. El proceso de normalización espacial consiste en un registro en dos pasos (rígido y elástico) de las imágenes de cada paciente con una imagen de referencia que hace la función de plantilla. Esto se consigue mediante el cálculo de una serie de transformaciones afines (rotaciones, desplazamientos y escalados) y de un campo de deformaciones para cada una de las imágenes que permitan maximizar su coincidencia con la citada plantilla. El cálculo se realiza de forma iterativa siguiendo el algoritmo DARTEL<sup>8</sup> descrito en [Ashburner2007] al mismo tiempo que la segmentación, por lo que la transformación necesaria para la normalización ya se obtuvo en el paso anterior. Como plantilla de normalización espacial se utilizó una imagen del *cerebro estándar* del Instituto de Neurología de Montreal, conocido como *plantilla MNI*. Esta plantilla es un promedio de 152 estudios corregistrados entre sí realizados a individuos sanos, y se incluye en SPM como una imagen de 91x109x91 vóxeles de dimensiones 2x2x2 mm. Dado que las imágenes proporcionadas por OASIS tienen un tamaño de voxel de 1x1x1 mm, éstas fueron remuestreadas a las dimensiones de la plantilla durante el proceso de normalización.

El resultado final de la normalización es un conjunto de imágenes segmentadas en el mismo sistema de coordenadas que la plantilla MNI. Dado que se deseaba conocer las cantidades de materia blanca y materia gris de las distintas regiones del cerebro la transformación se realizó con modulación, es decir conservando en el proceso de normalización de cada imagen la cantidad total del tejido correspondiente. De este modo, si por ejemplo un voxel de la imagen segmentada de materia gris (que originalmente tendría valor 1) correspondiese a 2 vóxeles en la imagen normalizada, cada uno de estos vóxeles tendría un valor igual a 0.5.

En todos los casos se emplearon los valores de los parámetros recomendados por defecto por los desarrolladores del algoritmo [Ashburner2010].

---

<sup>8</sup> Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra

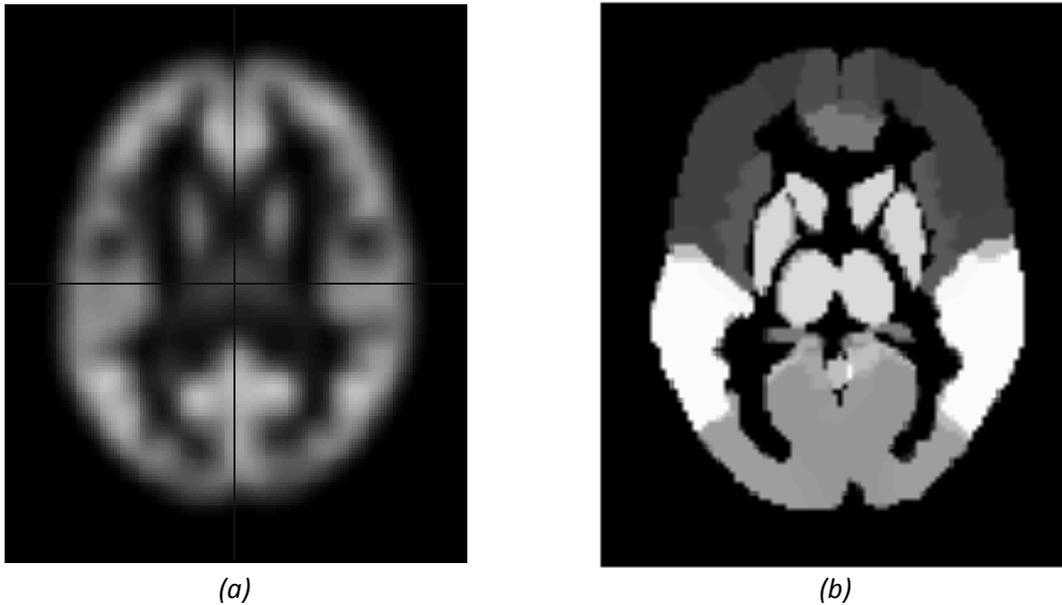


**Figura 5:** Fases de procesado de las imágenes. Imagen original del sujeto (a); Imagen segmentada para identificar las áreas correspondientes a materia gris (b); Plantilla MNI utilizada para la normalización espacial de todas las imágenes (c); Imagen segmentada normalizada espacialmente mediante la citada plantilla (d).

### ***Extracción de características***

Una vez obtenidas todas las imágenes normalizadas (una imagen de materia gris y otra de materia blanca por individuo) se procedió a calcular una serie de valores numéricos sobre ellas. Estos valores serán después utilizados por el algoritmo de clasificación como características descriptivas de cada uno de los cerebros estudiados. Los valores calculados fueron las sumas de materia gris y materia blanca en un total de 116 regiones del cerebro, que se identificaron mediante una plantilla en el espacio MNI realizada por Tzourio et al. [Tzourio2002] en base al Atlas de Talairach [Talairach1988]. Estas regiones están anatómicamente bien determinadas, lo que permite comparar los resultados obtenidos a partir de ellas con la información histológica disponible sobre la evolución de la enfermedad. Se conoce, por ejemplo, que determinadas regiones del cerebro

presentan una alteración más temprana que otras. Las dimensiones de la plantilla utilizada son idénticas a las de las imágenes normalizadas en el paso anterior.



**Figura 6:** Asignación de vóxeles de las imágenes normalizadas. El valor de cada voxel de la imagen normalizada (a) es asignado a la región anatómica indicada por la plantilla de parcelación (b); debe tenerse en cuenta que en la imagen muestra un único corte, por sólo se aprecian algunas de las 116 regiones en las que la plantilla parcela el cerebro completo.

Se implementó un programa en Matlab que calculó para cada individuo un array de 116 valores correspondientes a los volúmenes de materia gris en cada una de las regiones, y otro array de igual tamaño correspondiente a los volúmenes de materia blanca. La suma de ambos arrays elemento a elemento permite conocer el volumen de cada una de las 116 regiones.

Asimismo, para compensar las diferencias debidas a los diferentes tamaños de los cerebros, se dividieron todos los valores de cada individuo por una constante proporcional a su volumen intracraneal particular. Este valor se obtiene sumando en las imágenes segmentadas sin normalizar el número de vóxeles correspondientes a materia gris, materia blanca y líquido cefalorraquídeo del individuo correspondiente. De este modo se consigue una *normalización global* de toda la población estudiada [Ashburner2010]. Esta normalización anula las posibles diferencias de tamaño de estructuras debidas a volúmenes de cerebro diferentes, de forma que destaquen las diferencias causadas por la presencia de la enfermedad.

Finalmente, los datos fueron normalizados de modo que el valor de cada característica sobre el conjunto de la población tuviese media 0 y desviación típica 1. Esto permite que las características con un valor típico más grande no anulen la influencia de características de menor valor en el entrenamiento del clasificador.

## ***Selección de características***

El funcionamiento del clasificador empleando dos criterios de optimización y distintas características fue estudiado en dos etapas. En una primera etapa se seleccionó un subconjunto de 80 casos correspondientes a mujeres con características demográficas similares, de las cuales 20 presentaban síntomas muy leves de la enfermedad, 20 síntomas leves o moderados y 40 eran controles sanos. El sistema fue entrenado con las siguientes combinaciones de características:

- Todas las características (116 valores).
- Suma de características izquierda y derecha en el caso de regiones presentes en ambos hemisferios. La suma de los valores de un mayor número de vóxeles intenta reducir el *ruido* de las características, asumiendo que las posibles alteraciones se producirán de forma simétrica en ambos hemisferios (62 valores).
- Características correspondientes a aquellas regiones que, según la bibliografía clínica (véanse por ejemplo [Fennema2009], [Frisoni2010]), son las primeras en mostrar signos de alteración debida a la enfermedad de Alzheimer. Se seleccionaron las características correspondientes a la ínsula, cíngulo del cuerpo calloso, hipocampo, parahipocampo, amígdala, circunvolución fusiforme, putamen y lóbulos temporales (26 valores).
- Partiendo de las mismas características que en el caso anterior, suma de aquellos pares de características correspondientes a regiones simétricas del cerebro (13 valores).
- Selección de las características mediante un *árbol de decisión*. Se utilizó el algoritmo de entrenamiento de árboles de decisión CART descrito en [Breiman1984] que forma parte de la distribución de Matlab (función *classregtree*).

En una segunda etapa se comprobó el funcionamiento del clasificador empleando 414 casos con características demográficas muy diversas, limitando las pruebas a aquellas combinaciones de características que ofrecieron un mejor resultado en la etapa anterior.

En general un menor número de características reduce la posibilidad de que el sistema sea *sobreentrenado*, es decir que mediante el ajuste de un gran número de parámetros el sistema se limite a memorizar los casos que le son presentados pero no sea capaz de generalizar. Un entrenamiento correcto debe lograr que el sistema infiera qué combinaciones de características son las que determinan que las muestras pertenezcan a una u otra categoría. Esto le permitirá clasificar correctamente nuevas muestras que no le fueron presentadas durante el entrenamiento.

Las pruebas se realizaron tanto con las cantidades de materia gris como con los volúmenes de las distintas regiones. La razón de emplear los dos tipos de características es que la evolución de la enfermedad afecta a ambos factores, por lo que se quiso comprobar la utilidad de cada uno de ellos.

## Entrenamiento de la SVM

En la primera fase de pruebas se entrenó una SVM con cada uno de los conjuntos de características citados para un subconjunto de 80 pacientes seleccionados de la base de datos OASIS, tratándose en todos los casos de mujeres diestras mayores de 60 años. El proceso de entrenamiento fue el siguiente:

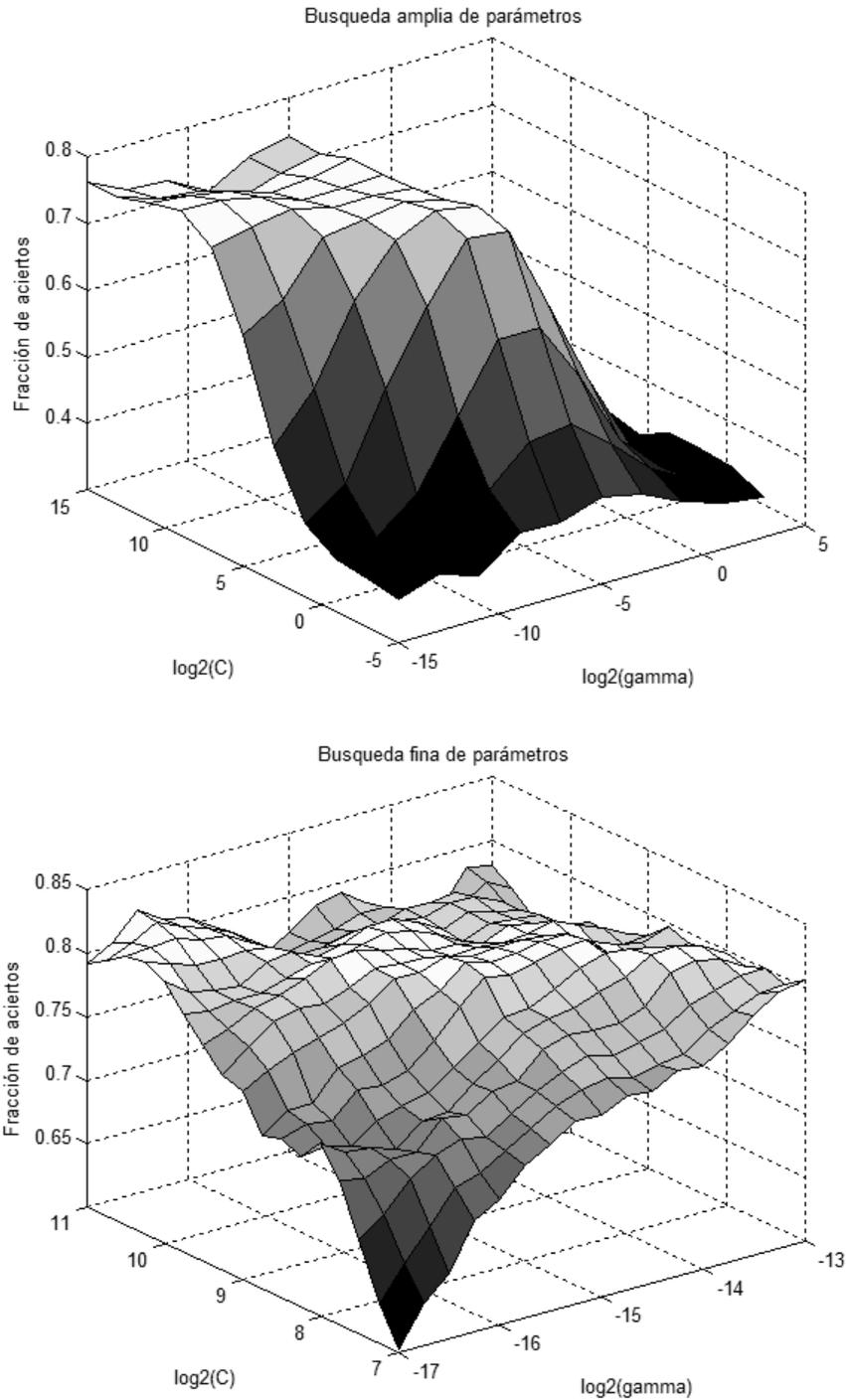
- De forma sucesiva se selecciona uno de los casos y el sistema es entrenado con los casos restantes. A continuación el sistema se utiliza para clasificar la muestra no utilizada, comprobándose si el resultado es correcto. El proceso se repite con cada una de las muestras del conjunto, obteniéndose una buena estimación del rendimiento del sistema, ya que se utiliza el máximo número de casos posible para el entrenamiento y se evalúa el rendimiento con un caso que no ha sido presentado antes al sistema. Esta técnica se conoce como *leave one out*.
- Se empleó un *kernel de funciones base radiales* (RBF) gaussianas, debido a su versatilidad [Hsu2010], y a que en anteriores estudios se demostró su buen funcionamiento en la resolución de problemas semejantes ([Savio2010], [Magnin2009]). La expresión de dicho kernel es

$$K(\vec{x}, \vec{y}) = e^{-\gamma \|\vec{x} - \vec{y}\|^2}$$

Siendo  $\vec{x}$  e  $\vec{y}$  los vectores de características de 2 muestras y  $\gamma$  un parámetro a determinar. Para emplear kernels RBF tan sólo es necesario optimizar el valor de dos parámetros:

- **C**: establece un nivel de compromiso entre el número de muestras clasificadas erróneamente tras el entrenamiento y la complejidad de la superficie de frontera entre categorías. Este parámetro debe optimizarse en cualquier SVM, independientemente del kernel utilizado.
- **$\gamma$** : Parámetro del kernel que controla la influencia en la decisión de cada muestra de entrenamiento en función de la *distancia* a la muestra que se desea clasificar (es decir, de cuán diferentes sean los valores de sus características).
- Para determinar los valores óptimos de dichos parámetros, se utilizó una validación cruzada *5-fold*. Esta estrategia consiste en dividir el conjunto de entrenamiento en 5 grupos y, de manera sucesiva, entrenar el clasificador con 4 de los grupos y comprobar su rendimiento clasificando el grupo restante. El proceso se repite para todos los posibles valores de los parámetros, eligiéndose aquellos que proporcionen el mejor resultado. En concreto, siguiendo las recomendaciones de [Hsu2010] se realizó una primera búsqueda en cuadrícula para los siguientes valores:

$$C = 2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}, 2^{13}, 2^{15}$$
$$\gamma = 2^{-15}, 2^{-13}, 2^{-11}, 2^{-9}, 2^{-7}, 2^{-5}, 2^{-3}, 2^{-1}, 2^1, 2^3, 2^5$$



**Figura 7:** Fracción de aciertos de un clasificador entrenado con 80 casos. Se emplearon 40 controles y 40 casos patológicos y distintos valores de los parámetros ( $C$ ,  $\gamma$ ): Búsqueda amplia en el rango completo de valores de los parámetros (arriba); Búsqueda fina entorno a los valores óptimos encontrados en la búsqueda amplia:  $C=2^9$ ,  $\gamma=2^{-15}$  (abajo).  
(Nota: ambas superficies fueron suavizadas para facilitar su visualización)

A continuación se realiza una búsqueda más fina entorno al mejor par de valores ( $C$ ,  $\gamma$ ) obtenido en la primera búsqueda, cubriendo dos órdenes de magnitud por encima y por debajo de los valores obtenidos en incrementos de  $2^{0.25}$ . La figura 7 muestra el rendimiento del clasificador en función de los parámetros ( $C$ ,  $\gamma$ ) en las búsquedas “amplia” y “fina”.

Es importante destacar que la estrategia de validación cruzada para la determinación de parámetros ya viene incorporada en la librería *libSVM*, y el criterio de valoración que emplea es la fracción de aciertos al clasificar el grupo restante. Este sistema fue el utilizado en un principio, pero se comprobó que los valores de los parámetros obtenidos no eran los idóneos cuando el número de muestras de las dos categorías era muy diferente. En esos casos, el sistema tendía a clasificar las muestras en la categoría más frecuente. Al ser más numerosos los casos de control que los patológicos, esto se traducía en un sistema de alta especificidad pero baja sensibilidad. Para paliar este problema, se programó un algoritmo de valoración alternativo que evaluaba cada conjunto de parámetros según la *F1-score* (definida a continuación) obtenida en cada caso.

- Todo el proceso se repitió 10 veces variando aleatoriamente el orden de las muestras, ya que se comprobó que dicho orden afecta a la selección de parámetros en la fase de validación cruzada, obteniéndose resultados ligeramente diferentes. Se tomó como resultado final de clasificación el promedio de los diez resultados.

La mayoría de las pruebas realizadas en esta fase requirió un tiempo de cómputo inferior a una hora por cada repetición en un PC dotado de un procesador Intel Core Duo de 2 GHz y 2 gigabytes de memoria RAM.

En una segunda fase de pruebas se repitió el proceso empleando un total de 414 casos disponibles en OASIS, a fin de comprobar el funcionamiento del clasificador sobre una población de características demográficas heterogéneas. El procedimiento seguido fue igual al anterior, pero únicamente se emplearon los conjuntos de características que ofrecieron un mejor rendimiento en la primera fase (véase el apartado de resultados). En concreto, se emplearon las cantidades de materia gris y los volúmenes de las regiones que, según la bibliografía clínica, sufren alteraciones en las etapas tempranas de la enfermedad. A efectos de comparación, también se comprobó el resultado al utilizar todas las características disponibles.

En esta segunda fase cada entrenamiento se repitió sólo 5 veces en lugar de 10 veces como en la primera debido a que el tiempo de computación necesario era muy elevado. Cada repetición requirió desde aproximadamente 6 horas (en caso de optimizar en función del número de aciertos con la rutina propia de *libSVM*) hasta 12 horas (en caso de optimizar en función del *F1-score* con la rutina programada al efecto).

### ***Evaluación del rendimiento***

El rendimiento del clasificador entrenado con cada conjunto de características se evaluó contando en cada caso el número de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN) obtenidos en el proceso de clasificación. A partir de estos valores se calculó la sensibilidad, especificidad y *F1-score* correspondientes a cada prueba según las expresiones:

$$\text{sensibilidad} = \frac{VP}{VP + FN}$$

$$\text{especificidad} = \frac{VN}{VN + FP}$$

$$F1 - score = \frac{2 \cdot \text{sensibilidad} \cdot \text{especificidad}}{\text{sensibilidad} + \text{especificidad}}$$



## Resultados

En las siguientes tablas se muestran los resultados obtenidos en las distintas pruebas. Las tablas 1 a 4 corresponden a las pruebas realizadas con 80 casos demográficamente homogéneos, utilizando como características las cantidades relativas de materia gris y los volúmenes relativos de 116 regiones del cerebro. Los casos se corresponden con 40 individuos sanos de control, 20 con síntomas muy leves de la enfermedad de Alzheimer y 20 con síntomas leves a moderados. En cada prueba se indica qué grupos de casos se intentó discriminar: controles frente a todos los casos diagnosticados, (indicado como “todos” en la tabla), controles frente a casos muy leves o controles frente a casos leves. También se indica qué subconjunto de las 116 características extraídas de las imágenes se empleó para entrenar el clasificador en cada caso. Todas las pruebas se repitieron empleando dos criterios de optimización diferentes en la etapa de entrenamiento: números absoluto de aciertos y F1-score.

La tabla 5 muestra los resultados obtenidos en la segunda fase de pruebas. En este caso se emplearon 414 casos de individuos de diferentes características demográficas: hombres y mujeres de edades comprendidas entre 18 y 96 años, de los cuales 314 eran controles sanos, 70 presentaban síntomas muy leves de la enfermedad de Alzheimer y 30 presentaban síntomas leves o moderados. En esta fase se emplearon en primer lugar las características escogidas en base a la bibliografía clínica, ya que fueron las que proporcionaron un mejor resultado en la primera fase de pruebas. La prueba se repitió utilizando todas las características, tanto de materia gris como de volumen, y empleando como criterios de optimización tanto el número de aciertos como el F1-score.

Por último, se incluyen una serie de gráficas que muestran los valores de sensibilidad, especificidad y F1-score obtenidos en las dos fases de pruebas, tanto empleando todas las características como únicamente el subconjunto elegido en base a la bibliografía clínica.

Características seleccionadas	Casos utilizados	Nº de características	Verdaderos positivos	Verdaderos negativos	Falsos positivos	Falsos negativos	Aciertos (%)	Sensibilidad	Especificidad	F1 score
<b>Todas</b>	Control/Muy leve	116	10.5	36.8	3.2	9.5	78.8	0.525	0.920	0.669
	Control/Leve	116	8.1	34.3	5.7	11.9	70.7	0.405	0.858	0.550
	Todos	116	26.7	27.5	12.5	13.3	67.8	0.668	0.688	0.677
<b>Suma regiones izda/dcha</b>	Control/Muy leve	62	10.0	36.1	3.9	10.0	76.8	0.500	0.903	0.643
	Control/Leve	62	8.3	36.3	3.7	11.7	74.3	0.415	0.908	0.570
	Todos	62	23.7	27.8	12.2	16.3	64.4	0.593	0.695	0.640
<b>Regiones bibliografía clínica</b>	Control/Muy leve	26	6.8	35.8	4.2	13.2	71.0	0.340	0.895	0.493
	Control/Leve	26	9.2	34.9	5.1	10.8	73.5	0.460	0.873	0.602
	Todos	26	28.2	27.5	12.5	11.8	69.6	0.705	0.688	0.696
<b>Regiones bib. clínica (suma izda/dcha)</b>	Control/Muy leve	13	6.9	37.4	2.6	13.1	73.8	0.345	0.935	0.504
	Control/Leve	13	10.7	33.6	6.4	9.3	73.8	0.535	0.840	0.654
	Todos	13	26.9	27.9	12.1	13.1	68.5	0.673	0.698	0.685
<b>Árbol – 1 nivel</b>	Todos	1	27.7	25.5	14.5	12.3	66.5	0.693	0.638	0.664
<b>Árbol – 2 niveles</b>	Todos	3	21.5	30.5	9.5	18.5	65.0	0.538	0.763	0.631
<b>Árbol – 3 niveles</b>	Todos	6	24.1	32.9	7.1	15.9	71.3	0.603	0.823	0.696
<b>Árbol – 4 niveles</b>	Todos	8	25.4	31.8	8.2	14.6	71.5	0.635	0.795	0.706

**Tabla 1:** Resultados obtenidos empleando como características las cantidades de **materia gris** de 116 regiones del cerebro con **80 casos** demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del **número de aciertos** obtenido.

Características seleccionadas	Casos utilizados	Nº de características	Verdaderos positivos	Verdaderos negativos	Falsos positivos	Falsos negativos	Aciertos (%)	Sensibilidad	Especificidad	F1 score
Todas	Control/Muy leve	116	12.3	35.6	4.4	7.7	79.8	0.615	0.890	0.727
	Control/Leve	116	10.6	34.6	5.4	9.4	75.3	0.530	0.865	0.657
	Todos	116	28.3	26.5	13.5	11.7	68.5	0.708	0.663	0.684
Suma regiones izda/dcha	Control/Muy leve	62	11.4	31.8	8.2	8.6	72.0	0.570	0.795	0.664
	Control/Leve	62	12.0	34.7	5.3	8.0	77.8	0.600	0.868	0.709
	Todos	62	27.8	28.5	11.5	12.2	70.4	0.695	0.713	0.704
Regiones bibliografía clínica	Control/Muy leve	26	10.4	31.8	8.2	9.6	70.3	0.520	0.795	0.629
	Control/Leve	26	12.1	32.3	7.7	7.9	74.0	0.605	0.808	0.692
	Todos	26	28.6	27.4	12.6	11.4	70.0	0.715	0.685	0.700
Regiones bib. clínica (suma izda/dcha)	Control/Muy leve	13	8.7	32.1	7.9	11.3	68.0	0.435	0.803	0.564
	Control/Leve	13	13.5	32.9	7.1	6.5	77.3	0.675	0.823	0.741
	Todos	13	26.4	26.2	13.8	13.6	65.8	0.660	0.655	0.657
Árbol – 1 nivel	Todos	1	28.2	25.9	14.1	11.8	67.6	0.705	0.648	0.675
Árbol – 2 niveles	Todos	3	22.4	28.2	11.8	17.6	63.3	0.560	0.705	0.624
Árbol – 3 niveles	Todos	6	24.6	32.2	7.8	15.4	71.0	0.615	0.805	0.697
Árbol – 4 niveles	Todos	8	26.5	30.9	9.1	13.5	71.8	0.663	0.773	0.713

**Tabla 2:** Resultados obtenidos empleando como características las cantidades de **materia gris** de 116 regiones del cerebro con **80 casos** demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del **F1 score** obtenido.

Características seleccionadas	Casos utilizados	Nº de características	Verdaderos positivos	Verdaderos negativos	Falsos positivos	Falsos negativos	Aciertos (%)	Sensibilidad	Especificidad	F1 score
<b>Todas</b>	Normal/Muy leve	116	7.7	37.1	2.9	12.3	74.7	0.385	0.928	0.544
	Normal/Leve	116	9.1	36.4	3.6	10.9	75.8	0.455	0.910	0.607
	Todos	116	26.1	28.0	12.0	13.9	67.6	0.653	0.700	0.675
<b>Suma regiones izda/dcha</b>	Normal/Muy leve	62	8.2	35.7	4.3	11.8	73.2	0.410	0.893	0.562
	Normal/Leve	62	9.0	36.4	3.6	11.0	75.7	0.450	0.910	0.602
	Todos	62	24.8	27.5	12.5	15.2	65.4	0.620	0.688	0.652
<b>Regiones bibliografía clínica</b>	Normal/Muy leve	26	9.6	33.5	6.5	10.4	71.8	0.480	0.838	0.610
	Normal/Leve	26	14.8	35.9	4.1	5.2	84.5	0.740	0.898	0.811
	Todos	26	33.3	33.1	6.9	6.7	83.0	0.833	0.828	0.830
<b>Regiones bib. clínica (suma izda/dcha)</b>	Normal/Muy leve	13	6.1	33.5	6.5	13.9	66.0	0.305	0.838	0.447
	Normal/Leve	13	14.8	35.1	4.9	5.2	83.2	0.740	0.878	0.803
	Todos	13	32.9	31.6	8.4	7.1	80.6	0.823	0.790	0.806
<b>Árbol – 1 nivel</b>	Todos	1	26.9	32.8	7.2	13.1	74.6	0.673	0.820	0.739
<b>Árbol – 2 niveles</b>	Todos	3	28.6	31.4	8.6	11.4	75.0	0.715	0.785	0.748
<b>Árbol – 3 niveles</b>	Todos	5	29.3	31.4	8.6	10.7	75.9	0.733	0.785	0.758
<b>Árbol – 4 niveles</b>	Todos	8	28.7	31.8	8.2	11.3	75.6	0.718	0.795	0.754

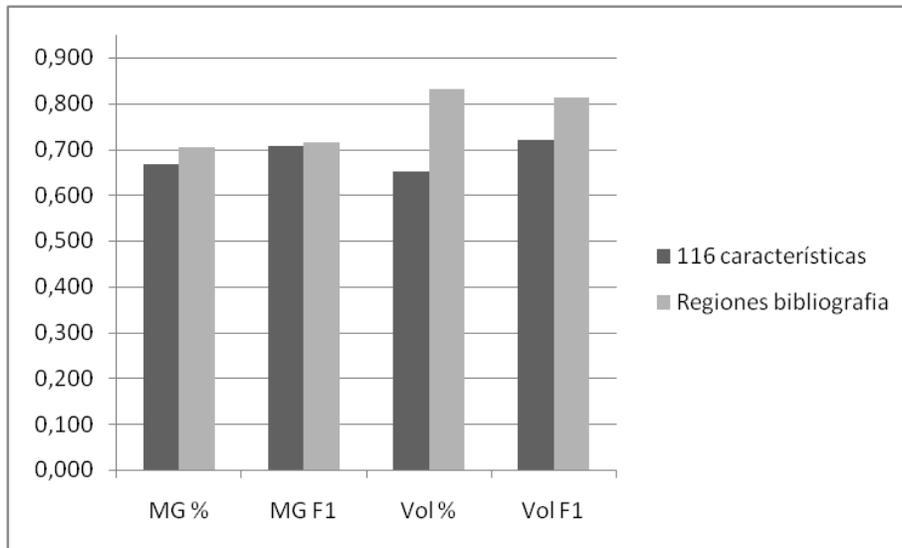
**Tabla 3:** Resultados obtenidos empleando como características los **volúmenes** de 116 regiones del cerebro con **80 casos** demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del **número de aciertos** obtenido.

Características seleccionadas	Casos utilizados	Nº de características	Verdaderos positivos	Verdaderos negativos	Falsos positivos	Falsos negativos	Aciertos (%)	Sensibilidad	Especificidad	F1 score
<b>Todas</b>	Normal/Muy leve	116	10.5	32.6	7.4	9.5	71.8	0.525	0.815	0.639
	Normal/Leve	116	11.2	32.6	7.4	8.8	73.0	0.560	0.815	0.664
	Todos	116	28.8	27.9	12.1	11.2	70.9	0.720	0.698	0.709
<b>Suma regiones izda/dcha</b>	Normal/Muy leve	62	10.4	31.6	8.4	9.6	70.0	0.520	0.790	0.627
	Normal/Leve	62	10.3	33.3	6.7	9.7	72.7	0.515	0.833	0.636
	Todos	62	26.2	26.9	13.1	13.8	66.4	0.655	0.673	0.664
<b>Regiones bibliografía clínica</b>	Normal/Muy leve	26	11.3	31.7	8.3	8.7	71.7	0.565	0.793	0.660
	Normal/Leve	26	14.9	35.4	4.6	5.1	83.8	0.745	0.885	0.809
	Todos	26	32.5	32.2	7.8	7.5	80.9	0.813	0.805	0.809
<b>Regiones bib. clínica (suma izda/dcha)</b>	Normal/Muy leve	13	9.1	31.1	8.9	10.9	67.0	0.455	0.778	0.574
	Normal/Leve	13	14.6	33.9	6.1	5.4	80.8	0.730	0.848	0.784
	Todos	13	31.3	31.1	8.9	8.7	78.0	0.783	0.778	0.780
<b>Árbol – 1 nivel</b>	Todos	1	26.6	32.5	7.5	13.4	73.9	0.665	0.813	0.731
<b>Árbol – 2 niveles</b>	Todos	3	28.5	31.2	8.8	11.5	74.6	0.713	0.780	0.745
<b>Árbol – 3 niveles</b>	Todos	5	29.9	30.5	9.5	10.1	75.5	0.748	0.763	0.755
<b>Árbol – 4 niveles</b>	Todos	8	29.0	32.5	7.5	11.0	76.9	0.725	0.813	0.766

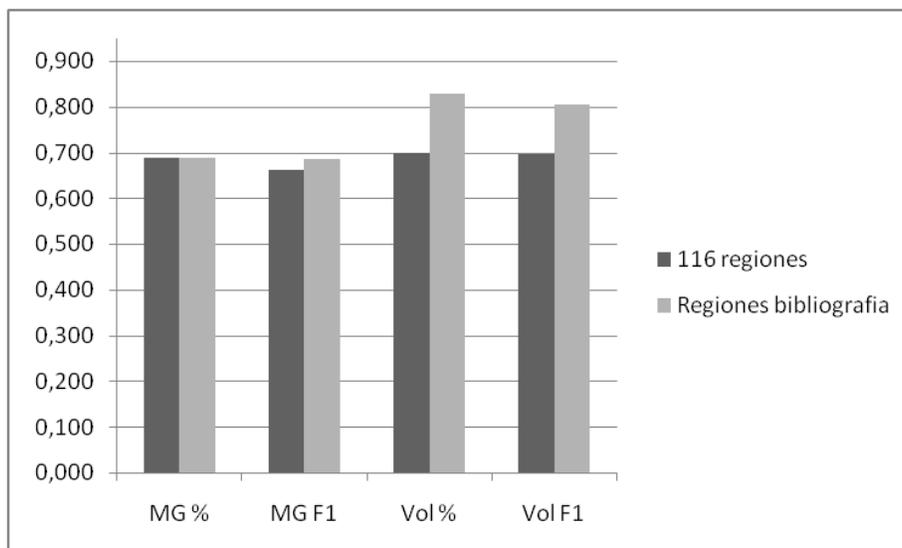
**Tabla 4:** Resultados obtenidos empleando como características los **volúmenes** de 116 regiones del cerebro con **80 casos** demográficamente homogéneos, y seleccionando los parámetros óptimos de la SVM en función del **F1 score** obtenido.

Características seleccionadas	Criterio de optimización de parámetros	Nº de características	Verdaderos positivos	Verdaderos negativos	Falsos positivos	Falsos negativos	Aciertos (%)	Sensibilidad	Especificidad	F1 score
<b>Materia gris de todas las regiones</b>	Nº de aciertos	116	58.0	285.8	28.2	42.0	83.0	0.580	0.910	<b>0.709</b>
	F1 score	116	62.2	281.8	32.2	37.8	83.1	0.622	0.897	<b>0.735</b>
<b>Volumen de todas las regiones</b>	Nº de aciertos	116	59.2	290.2	23.8	40.8	84.4	0.592	0.924	<b>0.722</b>
	F1 score	116	59.8	283.2	30.8	40.2	82.9	0.598	0.902	<b>0.719</b>
<b>Materia gris de regiones bib. clínica</b>	Nº de aciertos	26	64.8	287.4	26.6	35.2	85.1	0.648	0.915	<b>0.759</b>
	F1 score	26	65.6	286.4	27.6	34.4	85.0	0.656	0.912	<b>0.763</b>
<b>Volumen de regiones bib. clínica</b>	Nº de aciertos	26	46.6	289.2	24.8	53.4	81.1	0.466	0.921	<b>0.619</b>
	F1 score	26	51.4	281.0	33.0	48.6	80.3	0.514	0.895	<b>0.653</b>

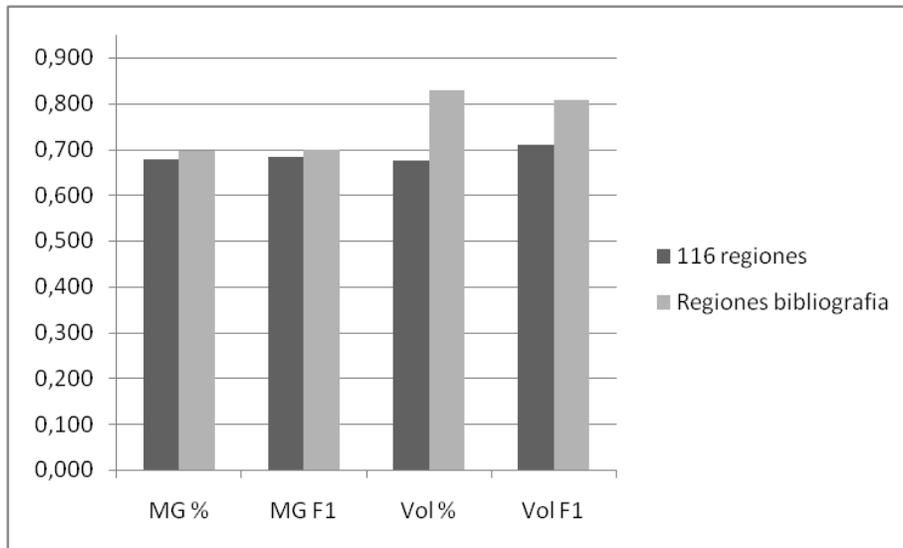
**Tabla 5:** Resultados obtenidos empleando **414 casos** de características demográficas heterogéneas y distintos conjuntos de características y criterios de optimización.



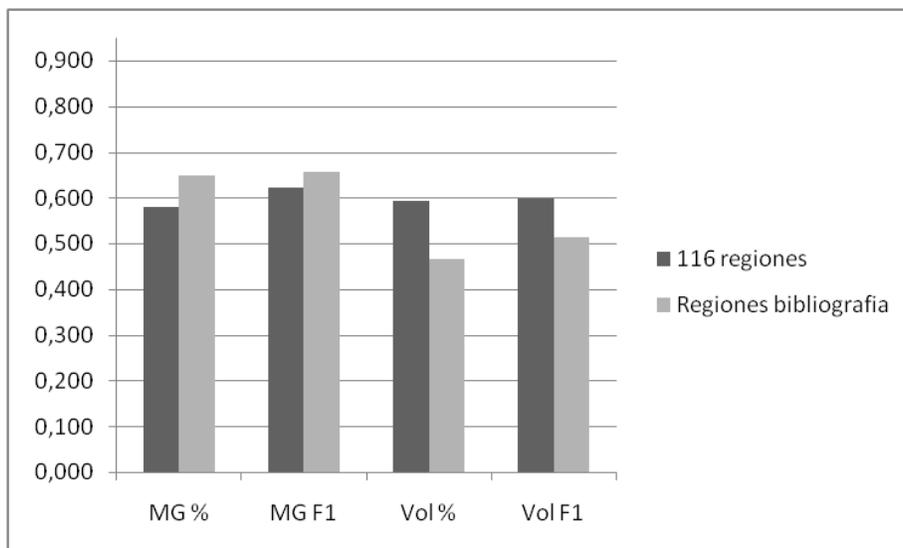
**Figura 8:** Comparación de **sensibilidades** obtenidas con **80 casos** empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



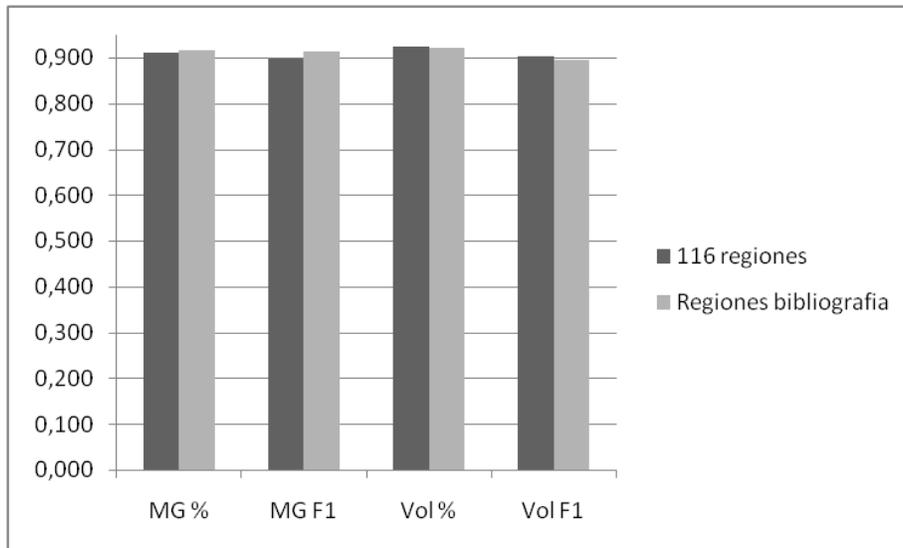
**Figura 9:** Comparación de **especificidades** obtenidas con **80 casos** empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



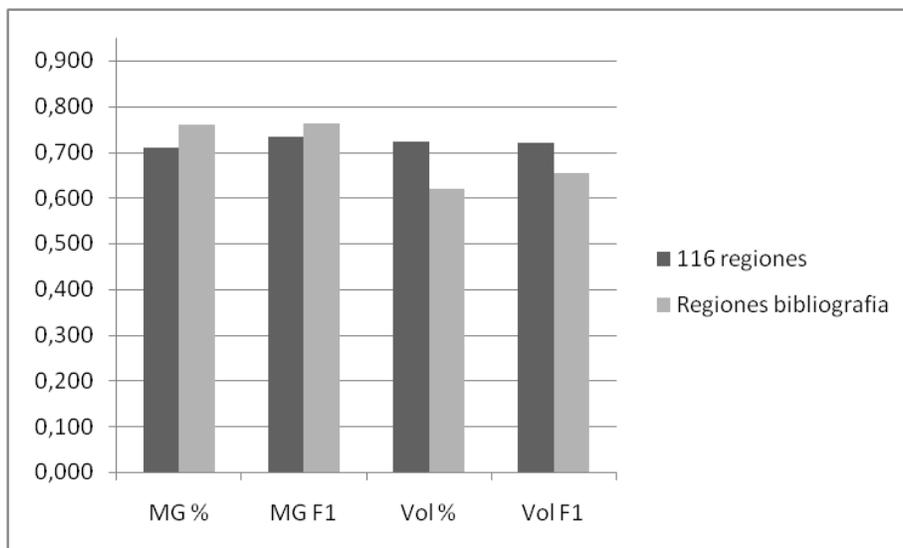
**Figura 10:** Comparación de puntuaciones **F1-score** obtenidas con **80 casos** empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



**Figura 11:** Comparación de **sensibilidades** obtenidas con **414 casos** empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



**Figura 12:** Comparación de *especificidades* obtenidas con **414 casos** empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



**Figura 13:** Comparación de puntuaciones F1-score obtenidas con 414 casos empleando todas las características y un subconjunto elegido en base a la bibliografía. Se muestran los resultados obtenidos empleando como características las cantidades de materia gris (MG) y los volúmenes de las distintas regiones (Vol), así como dos criterios de optimización diferentes: porcentaje de aciertos (%) y F1-score (F1).



## Discusión de los resultados

### *Primera fase de pruebas (80 casos)*

Los resultados obtenidos empleando como características los volúmenes fueron mejores que con las cantidades de materia gris. La mejoría se aprecia especialmente al reducir el número de características a las que se suponían más significativas (bien fuese basándose en la bibliografía clínica o en árboles de decisión). De este modo se reduce el riesgo de sobreentrenamiento. Como excepción, se observó un peor resultado al reducir el número de características a la mitad sumando los volúmenes de las regiones equivalentes de ambos hemisferios. Este resultado concuerda con algunos resultados publicados [Fennema2009] que señalan una degeneración asimétrica de algunas estructuras en ambos hemisferios. De ser esto cierto, el hecho de sumar los valores de ambos hemisferios estaría restando información útil.

El mejor resultado en esta fase de pruebas se obtuvo empleando únicamente los volúmenes de las áreas más significativas según los estudios clínicos. El resultado en este caso en particular fue ligeramente mejor empleando la optimización por fracción de aciertos que por F1-score.

En general, la optimización basada en F1-score ofrece un mejor resultado que la basada en la fracción de aciertos, lográndose valores más altos de sensibilidad. Esto se aprecia muy especialmente en aquellos casos en que el número de muestras de las dos categorías es diferente, lo que da más generalidad al resultado.

En la mayoría de los casos, la sensibilidad fue más baja que la especificidad. Esto significa que una SVM entrenada en estas condiciones es mucho más fiable para descartar que para confirmar la presencia de la enfermedad. Este fenómeno ya había sido observado en otros estudios [Magnin2009], [Savio2010].

Cabe señalar que todos los casos excepto uno el sistema obtuvo un mejor resultado distinguiendo pacientes leves frente a controles que en el caso de pacientes muy leves frente a controles. Este es el resultado que cabría esperar, dado que la degradación progresiva de los tejidos debe hacer que la detección de un paciente sea más fácil cuanto más avanzada se encuentre su enfermedad. La única excepción se produjo en el caso de emplear la totalidad de los valores de materia gris (tanto de manera independiente como sumando ambos hemisferios) para entrenar el sistema. Este resultado anómalo, si bien pudo deberse a una anomalía estadística (que podría desaparecer empleando un número mayor de casos) sugiere que los volúmenes de las distintas regiones constituyen características más fiables para la detección de la patología que las cantidades de materia gris, en el caso de un conjunto de muestras homogéneo.

## ***Segunda fase de pruebas (414 casos)***

En la segunda fase del estudio se entrenó una SVM con 414 casos de diversas características demográficas. En este caso únicamente se comprobó el resultado obtenido empleando las cantidades de materia gris y los volúmenes de las regiones seleccionadas a partir de estudios clínicos, y se comparó con el obtenido empleando la totalidad de las características. Al igual que en la primera fase, la prueba se repitió optimizando los parámetros de la SVM tanto en función del número de aciertos como de la puntuación *F1-score* obtenida en el entrenamiento.

En esta fase también se observó una mejora general de los resultados al basar la optimización en el *F1-score* en lugar del número de aciertos. Igualmente, la sensibilidad fue muy inferior a la especificidad en todos los casos. La especificidad estuvo siempre entorno a un 90%. La máxima sensibilidad obtenida fue de aproximadamente un 66%.

El resultado más llamativo fue que, al contrario que en la primera fase, los resultados obtenidos empleando como características los valores de materia gris fueron superiores a los obtenidos empleando los volúmenes de las regiones. En concreto, el mejor resultado se obtuvo en el caso de emplear como características las cantidades de materia gris de las regiones seleccionadas en base a estudios clínicos. De todos modos no debe olvidarse que ambas magnitudes (volúmenes y materia gris) están relacionadas, ya que la reducción de volumen asociada a la enfermedad se debe a pérdida de tejido neuronal y por lo tanto también supone una pérdida de materia gris en esa región.

Una posible explicación a los diferentes resultados obtenidos en las dos fases es que la mayor diversidad anatómica de las muestras empleadas (individuos de un amplio rango de edades y ambos sexos) enmascare las diferencias debidas a los estadios iniciales de la enfermedad. Debe recordarse que el proceso de normalización espacial de las imágenes previo a la extracción de características sirve únicamente para identificar las distintas regiones del cerebro en todos los pacientes mediante una única plantilla, pero no altera las cantidades de materia gris o volúmenes de cada una de ellas (en concreto, el volumen de cada región se calcula como la suma de sus cantidades de materia gris y materia blanca). Por lo tanto, cualquier diferencia de estos valores debida a las variaciones anatómicas de los pacientes o a su edad se conserva tras la normalización, al igual que las posibles alteraciones debidas a la enfermedad.

En cualquier caso los resultados sugieren que el uso de las cantidades de materia gris en las distintas regiones ofrece un resultado mucho más estable frente a la diversidad de las características demográficas de las muestras que el uso de los volúmenes de dichas regiones.

## Conclusiones

De los resultados de las pruebas realizadas cabe extraer las siguientes conclusiones:

- **Cuando la SVM es entrenada en las condiciones óptimas es capaz de realizar una clasificación de pacientes con una fiabilidad similar a la obtenida por radiólogos a partir de imágenes clínicas** en estadios tempranos de la enfermedad [Frisoni2010]. Este rendimiento se alcanza cuando el sistema es **entrenado con un conjunto de casos homogéneo** (individuos de características demográficas similares). Además, **deben emplearse como características los volúmenes relativos de las regiones indicadas por la bibliografía clínica** como más sensibles a sufrir alteraciones debido a la enfermedad. El resultado por lo tanto es coherente con los hallazgos clínicos.
- **La utilización de características adicionales empeora los resultados del sistema.** Esto probablemente se deba a que no aporten una cantidad significativa de información útil para el problema que se estudia favoreciendo además el *sobreentrenamiento* del sistema. El resultado es una disminución de su capacidad para clasificar correctamente casos que no formen parte de su conjunto de entrenamiento.
- **La fiabilidad del método disminuye cuando se aplica a un conjunto de estudios heterogéneo.** Este resultado es razonable dado que factores como la edad o el sexo introducen una mayor variabilidad en la anatomía del cerebro que no está directamente relacionada con la patología que se pretende detectar. Por lo tanto **para un funcionamiento óptimo el sistema debe entrenarse y aplicarse a un conjunto de casos lo más homogéneos posible desde el punto de vista demográfico.**
- **En caso de que la única opción sea entrenar la SVM con una población heterogénea** (amplio rango de edades y/o ambos sexos) **será preferible emplear como características las cantidades relativas de materia gris de las regiones de interés clínico**, ya que en esas condiciones ofrecen un mejor resultado que sus volúmenes



## Bibliografía

- Ashburner, J. (2010), 'VBM Tutorial'. <http://www.fil.ion.ucl.ac.uk/~john/misc/VBMclass10.pdf>
- Ashburner, J. (2007), 'A fast diffeomorphic image registration algorithm.', *Neuroimage* 38(1), 95--113.
- Ashburner, J. & Friston, K. J. (2005), 'Unified segmentation.', *Neuroimage* 26(3), 839--851.
- Bennet, K. & Campbell, C. (2000), 'Support Vector Machines: Hype or Hallelujah?', *SIGKDD Explorations* 2(2).
- Bittencourt, H. & Clarke, R. (2004), 'Feature Selection by Using Classification and Regression Trees (CART)'. 'The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences'. Estambul.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984) *Classification and Regression Trees*. CRC Press.
- Burges, C. (1998), 'A Tutorial on Support Vector Machines for Pattern Recognition', *Data Mining and Knowledge Discovery* 2, 121-167.
- Busatto, G. F.; Diniz, B. S. & Zanetti, M. V. (2008), 'Voxel-based morphometry in Alzheimer's disease.', *Expert Rev Neurother* 8(11), 1691--1702.
- Chang, C.-C. & Lin, C.-J. (2011), 'LIBSVM: A library for support vector machines', *ACM Transactions on Intelligent Systems and Technology* 2, 27:1--27:27.
- Doi, K. (2007), 'Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential', *Comput Med Imaging Graph* 31(4-5), 198-211.
- Fennema-Notestine, C.; McEvoy, L. K.; Hagler, Jr, D. J.; Jacobson, M. W.; Dale, A. M. & Initiative, T. A. D. N. (2009), 'Structural neuroimaging in the detection and prognosis of pre-clinical and early AD.', *Behav Neurol* 21(1), 3--12.
- Fletcher, T. (2009), 'Support Vector Machines Explained'.
- Frisoni, G. B.; Fox, N. C.; Jack, Jr, C. R.; Scheltens, P. & Thompson, P. M. (2010), 'The clinical use of structural MRI in Alzheimer disease.', *Nat Rev Neurol* 6(2), 67--77.
- Guo, L.; Liu, X.; Wu, Y.; Yan, W. & Shen, X. (2007), 'Research on the segmentation of MRI image based on multi-classification support vector machine.', *Conf Proc IEEE Eng Med Biol Soc* 2007, 6020--6023.
- Hsu, W.; Chang, C. & Lin, C.-J. (2010), 'A Practical Guide to Support Vector Classification'.

de Magalhães Oliveira, P. P.; Nitrini, R.; Busatto, G.; Buchpiguel, C.; Sato, J. R. & Amaro, E. (2010), 'Use of SVM methods with surface-based cortical and volumetric subcortical measurements to detect Alzheimer's disease.', *J Alzheimers Dis* 19(4), 1263--1272.

Magnin, B.; Mesrob, L.; Kinkingnéhun, S.; Péligrini-Issac, M.; Colliot, O.; Sarazin, M.; Dubois, B.; Lehéricy, S. & Benali, H. (2009), 'Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI.', *Neuroradiology* 51(2), 73--83.

Marcus (2007), 'OASIS Fact Sheet (rev. 2007-8-20)'.

Marcus, D. S.; Wang, T. H.; Parker, J.; Csernansky, J. G.; Morris, J. C. & Buckner, R. L. (2007), 'Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults.', *J Cogn Neurosci* 19(9), 1498--1507.

Moguerza, J. M. (2006), 'Support Vector Machines with Applications', *Statistical Science* 21(3), 322-336.

Morris, J. C. (1993), 'The Clinical Dementia Rating (CDR): current version and scoring rules.', *Neurology* 43(11), 2412--2414.

Platt, J. (1998), 'Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines', Technical report, Microsoft Research.

Safavian, S. R. & Landgrebe, D. (1991), 'A Survey of Decision Tree Classifier Methodology', *IEEE Transactions on Systems, Man and Cybernetics* 21, 660-674.

Savio, A.; García-Sebastián, M.; Chyzyk, D.; Hernandez, C.; Graña, M.; Sistiaga, A.; de Munain, A. L. & Villanúa, J. (2010), 'Neurocognitive disorder detection based on Feature Vectors extracted from VBM analysis of structural MRI', .

Talairach, J. & Tournoux, P. (1988), *Co-planar Stereotaxic Atlas of the Human Brain 3-Dimensional Proportional System: An Approach to Cerebral Imaging*, Stuttgart, New York.

Tzourio-Mazoyer, N.; Landeau, B.; Papathanassiou, D.; Crivello, F.; Etard, O.; Delcroix, N.; Mazoyer, B. & Joliot, M. (2002), 'Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.', *Neuroimage* 15(1), 273--289.

Vapnik, V. N. (1998), *Statistical Learning Theory*, Wiley-Interscience.

Wenk, G. L. (2003), 'Neuropathologic changes in Alzheimer's disease.', *J Clin Psychiatry* 64 Suppl 9, 7--10.