



UNIVERSIDAD DE GRANADA

TESIS DOCTORAL

**STATISTICAL NEUROIMAGE MODELING, PROCESSING
AND SYNTHESIS BASED ON TEXTURE AND
COMPONENT ANALYSIS: TACKLING THE SMALL
SAMPLE SIZE PROBLEM.**

Presented by:
Francisco Jesús Martínez Murcia

Advisors:
Javier Ramírez Pérez de Inestrosa
Juan Manuel Górriz Sáez

To apply for the:
**International PhD Degree in Information and Communication
Technologies.**

Programa Oficial de Doctorado en Tecnologías de la Información
y las Comunicaciones

Junio 2017

Francisco Jesus Martinez Murcia

Statistical Neuroimage Modeling, Processing and Synthesis based on Texture and Component Analysis: Tackling the Small Sample Size Problem.

Tesis Doctoral. ©© 2017

This document was written with \LaTeX on Linux using a modified ArsClassica, a reworking of the ClassicThesis style designed by André Miede.

License



Este obra está bajo una licencia de **Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional**. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Contact

✉ fjesusmartinez@ugr.es

Hazlo o no lo hagas, pero no lo intentes.

— El imperio contraataca

A la memoria de Obdulia Martínez.

DECLARACIÓN

El doctorando Francisco Jesús Martínez Murcia y los directores de la tesis Javier Ramírez Pérez de Inestrosa y Juan Manuel Górriz Sáez garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, Junio 2017

Directores de la Tesis:

Javier Ramírez Pérez de Inestrosa

Doctorando:

Francisco Jesús Martínez Murcia

Juan Manuel Górriz Sáez

ABSTRACT

The rise of neuroimaging in the last years has provided physicians and radiologist with the ability to study the brain with unprecedented ease. This led to a new biological perspective in the study of neurodegenerative diseases, allowing the characterization of different anatomical and functional patterns associated with them. Computer Aided Diagnosis (CAD) systems use statistical techniques for preparing, processing and extracting information from neuroimaging data pursuing a major goal: optimize the process of analysis and diagnosis of neurodegenerative diseases and mental conditions.

With this thesis we focus on three different stages of the CAD pipeline: pre-processing, feature extraction and validation. For preprocessing, we have developed a method that target a relatively recent concern: the confounding effect of false positives due to differences in the acquisition at multiple sites. Our method can effectively merge datasets while reducing the acquisition site effects. Regarding feature extraction, we have studied decomposition algorithms (independent component analysis, factor analysis), texture features and a complete framework called Spherical Brain Mapping, that reduces the 3-dimensional brain images to two-dimensional statistical maps. This allowed us to improve the performance of automatic systems for detecting Alzheimer's and Parkinson's diseases. Finally, we developed a brain simulation technique that can be used to validate new functional datasets as well as for educational purposes.

RESUMEN AMPLIO EN CASTELLANO

Motivación

Durante los últimos años ha habido una explosión en el uso de neuroimagen en la práctica clínica. Gracias a su facilidad para explorar de forma no invasiva el interior del cerebro, se ha mejorado y acelerado sustancialmente el proceso de diagnóstico de un gran número de enfermedades. Asimismo, el uso de neuroimagen en investigación también está muy extendido. Campos tan variados como la psiquiatría, neurología, psicología, ciencias del comportamiento o biología hacen un uso extensivo de la neuroimagen en sus estudios.

Las bases de estos estudios son comunes: un proceso mediante el cual se recluta un conjunto representativo de sujetos, el desarrollo de un procedimiento experimental con cada sujeto y un análisis de los datos obtenidos. Normalmente, cuando se estudia una determinada enfermedad, es normal reclutar tanto sujetos afectados como un set similar de individuos sanos, conocidos como controles. En un ejemplo típico, se adquieren una serie de imágenes de cada una de estas poblaciones, y se analiza la función o la estructura cerebral de estos individuos utilizando diferentes herramientas estadísticas. El resultado de estos análisis es, frecuentemente, una lista de las diferencias más significativas en función o estructura que pueden ser asociadas a una enfermedad.

Existe un nuevo paradigma conocido como Computer Aided Diagnosis (CAD) (diagnóstico asistido por computador), que proporciona diversas herramientas para realizar estos análisis. Actualmente es un área de investigación multidisciplinar que combina informática, matemáticas, inteligencia artificial, medicina o estadísticas, entre otros. [7]. El objetivo principal de un CAD es ayudar al personal médico en el proceso de diagnóstico y estudio de las enfermedades, proporcionando software capaz de reconocer patrones, caracterizar diferencias y realizar predicciones sobre datos médicos.

Un problema al que se enfrentan continuamente estos estudios es el tamaño muestral. El número de sujetos reclutados habitualmente va desde algunas decenas hasta cientos de sujetos. Sin embargo, el número de características utilizadas puede llegar a varios millones (voxels en las imágenes), causando un problema que se conoce como *Small Sample Size Problem* [138] (problema del pequeño tamaño muestral). Este problema afecta el poder estadístico de

cualquier experimento realizado sobre bases de datos que reunen estas características [31].

Objetivos

El objetivo fundamental de esta tesis es superar el Small Sample Size problem en neuromagen. De este modo, los sistemas CAD resultantes serán más precisos y se reducirá el número de falsos positivos, incrementando la fiabilidad de sus resultados.

Para ello, utilizaremos dos enfoques diferentes, como hemos comentado: incrementar el tamaño muestral y reducir el espacio de características. Por todo ello, podemos definir los siguientes objetivos:

- Desarrollar y evaluar algoritmos que reducen el espacio de características, en lo que usualmente se conoce como selección y extracción de características.
- Desarrollar y evaluar nuevas estrategias para incrementar de forma segura el tamaño muestral en estudios de neuroimagen.

La mayor parte de la literatura existente se centra en el primer objetivo. Para ello, se han utilizado extensivamente algoritmos como Principal Component Analysis (PCA) [13, 60] o Partial Least Squares (PLS) [40]. En este sentido, hemos desarrollado tres estrategias diferentes:

- Una combinación de algoritmos de descomposición de imágenes y selección de características. Aquí hemos aplicado tests de hipótesis bajo tres criterios diferentes para seleccionar los voxels más significativos de las imágenes, y luego hemos aplicado Factor Analysis (FA) y Independent Component Analysis (ICA) para descomponer los datos y reducir el espacio de características.
- Una extracción de características basada en análisis de texturas.
- Una estrategia novedosa que hemos llamado Spherical Brain Mapping (SBM). En esta técnica se utiliza el sistema de coordenadas esférico para mapear diferentes medidas estadísticas en un plano bidimensional. Para ello se utilizan caminos que son usados como vectores de selección, a lo largo de los cuales se calculan las diferentes medidas.

Por otro lado, hemos evaluado nuevas formas de incrementar el tamaño muestral en estudios de neuroimagen, para los cuales hemos desarrollado:

- Un sistema llamado Significance Weighted Principal Component Analysis (SWPCA) que reduce varianza no deseada en estudios estructurales multicentro. El propósito de este sistema es reducir la cantidad de falsos positivos en grandes colaboraciones, suministrando imágenes más homogéneas que mejoren el resultado de los análisis estadísticos.
- Un algoritmo que simula imágenes cerebrales funcionales utilizando datos existentes, y que por tanto puede incrementar el número de muestras usados en estos estudios.

Contribuciones

Capítulo 4: Descomposición de Imágenes

En este capítulo proponemos una serie de sistemas CAD que combinan selección de características y descomposición de imágenes, y que fueron publicados en [24, 35, 47, 59].

Estos sistemas utilizan selección de características basada en tests de hipótesis, usando tres criterios diferentes: el test-t de student, la entropía relativa (o divergencia de Kullback-Leibler (KL)) y el test-U de Mann-Whitney-Wilcoxon (MWW). Tras calcular la significancia estadística de cada voxel, utilizamos FA o ICA para desarrollar el set de voxels más significativos como combinación de un número de componentes que varía entre 2 y 25.

Para evaluar esta metodología, propusimos varios experimentos que hemos realizado sobre las bases de datos de Alzheimer (ADNI-PET y VDNL-HMPAO) y Parkinson (VDLN-DAT, VDLV-DAT y PPMI-DAT). En el caso de Alzheimer, nuestro sistema CAD alcanzó valores de precisión superiores a 90 %, en cualquier combinación de selección de características y descomposición. Se detectaron diferencias entre los métodos de selección, que focalizaban su atención en áreas ampliamente relacionadas con la enfermedad en la bibliografía como los lóbulos occipital y temporal, partes del giro angular o incluso el hipocampo [96, 160].

Para el caso de las bases de datos de Parkinson, estas diferencias se localizaron fundamentalmente en el estriado, y aquellos métodos que seleccionaban otras áreas obtuvieron peores resultados. En general, la precisión alcanzada en las bases de datos de Parkinson alcanzó el 90 % en casi todos los casos, e incluso superior a 95 % para la base de datos VDLV-DAT, la única que no contenía sujetos con parkinsonismo, pero sin evidencia de déficit de dopamina (SWEDD).

Capítulo 5: Análisis de Textura

Para el capítulo 5, probamos los métodos de análisis de textura de Haralick aplicados a la construcción de matrices de co-ocurrencia de niveles de gris (GLCM) tridimensionales, tal y como se detalla en [25, 34].

Esta metodología está basada en el cálculo de una GLCM a lo largo de la imagen médica, usando una distancia y dirección determinada, a partir de la cual se pueden derivar diferentes características de textura, conocidas como características de Haralick (el investigador que las propuso en [172]). Estas características son capaces de detectar patrones como la tendencia a formar clusters, energía, correlación homogeneidad, o entropía, entre otros.

Para su evaluación, propusimos varios experimentos, aplicados a las imágenes DaTSCAN de las tres bases de datos VDLN-DAT, VDLV-DAT y PPMI-DAT. En primer lugar comprobamos qué características eran capaces de discriminar mejor entre afectados por la enfermedad de Parkinson y no afectados (esta vez eliminando a los sujetos SWEDD). Obtuvimos que las mejores características eran la tendencia a formar clusters y la homogeneidad. Dado que la distribución de intensidades en DaTSCAN está muy concentrada en el estriado, diferencias de forma (tendencia a formar clusters) y distribución del radiofármaco (homogeneidad) explicaban mejor que otras las diferencias entre sujetos. Una combinación de estas y otras características, evaluando su significancia utilizando los criterios estadísticos mencionados en el capítulo anterior, logró alcanzar más del 97 % de precisión, cuando se aplicó a la base de datos PPMI-DAT, junto a la divergencia de KL.

Capítulo 6: Spherical Brain Mapping (SBM)

El Spherical Brain Mapping (SBM) es una de las contribuciones más novedosas de esta tesis, y es un marco que ha sido desarrollado en [6, 7, 14, 23]. La base de SBM es el establecimiento de un origen de coordenadas (el centro de la imagen o la comisura anterior, por ejemplo), a partir del cual se crean diferentes vectores a diferentes ángulos de elevación y azimuth. Este vector de mapeo se utiliza para muestrear todos los voxels que son cruzados por él, usando los cuales se pueden obtener diferentes medidas estadísticas y morfológicas.

También, en la llamada Volumetric Radial LBP (VRLBP) se establece un muestreo helicoidal alrededor del vector de mapeo, y es usado para calcular características de textura en la dirección en la cual se ha generado el vector. Por último, en los caminos basados en modelos ocultos de Markov (HMM), se sustituye el vector de mapeo por un camino sujeto a dos restricciones: mínima variación de intensidad y orientado en un azimuth y elevación determinada.

En este capítulo desarrollado cinco experimentos para comprobar la habilidad de los mapas generados mediante SBM en el diagnóstico del Alzheimer. En primer lugar hemos probado las diferentes medidas (superficie, grosor cortical, número de pliegues, promedio, entropía, kurtosis y VRLBP), de las cuales obtuvimos hasta un 90 % de precisión en el diagnóstico diferencial del Alzheimer. Posteriormente se probó un enfoque por capas de estos mapas, en los cuales el vector de mapeo se dividía en cuatro partes iguales, y se obtenía un mapa por cada una de estas partes. Estos mapas no consiguieron mejorar la precisión alcanzada anteriormente, pero eran capaces de localizar mejor las diferencias entre individuos, al poseer mayor resolución de profundidad.

Por último, los tres últimos experimentos probaban la habilidad de los caminos basados en HMM como selección de características y como cómputo de mapas bidimensionales de textura, usando una adaptación de la GLCM. Gracias a estos experimentos, vimos que el desempeño de estos sistemas era peor que el del SBM original, aunque gracias a adaptarse a la anatomía interna del cerebro, podrían ser mejor utilizados en otros procesos como segmentación, o incluso para extraer características morfológicas.

Capítulo 7: Significance Weighted Principal Component Analysis (SWPCA)

Propusimos el algoritmo Significance Weighted Principal Component Analysis (SWPCA) como solución al problema de mezclar imágenes procedentes de diferentes lugares o equipamiento de adquisición en [1]. El proyecto surgió cuando, mediante la evaluación de la base de datos de autismo del Medical Research Council Autism Imaging Multicentre Study (MRC-AIMS) nos dimos cuenta de que las diferencias entre sitios de adquisición eran mayores que las diferencias entre sujetos afectados y no afectados por el trastorno del espectro autista.

SWPCA se basa en una descomposición de las imágenes usando PCA, un modelado de relaciones de cada componente con una variable categórica, en este caso el sitio de adquisición, usando Analysis Of Variance (ANOVA), y una reconstrucción pesada de las imágenes utilizando una función que definimos en dicho artículo. Gracias a esta reconstrucción, todas las componentes de variabilidad relacionadas con el sitio de adquisición se reducen, y por tanto, las imágenes son más homogéneas y en un posterior análisis, tendremos una mayor certeza de que los resultados son debidos a la clase de la imagen, y no al sitio de adquisición.

Hemos probado el algoritmo en la base de datos AIMS-MRI, donde comprobamos que las diferencias entre sitios de adquisición se reducían sustan-

cialmente después de reconstruir la base de datos mediante SWPCA. Cuando analizamos las imágenes “limpias” mediante SPM, apenas encontramos diferencias estructurales entre sujetos afectados y controles. Esto puede ser debido a una multitud de factores, entre ellas que existe una gran variabilidad entre patrones estructurales en el trastorno del espectro autista, o que directamente no existen diferencias entre afectados y controles. Esto nos llevó a plantear la necesidad de establecer diferentes subgrupos para realizar un mejor análisis, ya que en trabajos recientes se había demostrado que los resultados de diferencias estructurales en autismo son generalmente inconsistentes [21].

Capítulo 8: Simulación de Imágenes Funcionales

Finalmente, en el capítulo 8 propusimos un sistema capaz de simular nuevas imágenes funcionales, similares a una base de datos dada. El objetivo de este sistema era generar cientos de nuevas imágenes que comparten características con las imágenes originales, pero siendo independientes de ellas.

El procedimiento consta de un análisis de las imágenes, previamente normalizadas espacialmente y en intensidad, utilizando PCA. Esto genera unas coordenadas de cada sujeto en el espacio ‘eigenbrain’, la base espacial definida por las componentes principales. Gracias a ello, podemos analizar la distribución estadística de estas coordenadas asociadas a la clase a la que pertenecen, y utilizar procedimientos de estimación de densidad de probabilidad para modelarlos. En este trabajo hemos utilizado la distribución normal multivariable y Kernel Density Estimation (KDE) como herramienta, pero se puede utilizar cualquier otro modelo de estimación de la función de densidad de probabilidad (PDF) que se desee. Una vez caracterizada esta PDF, podemos generar nuevos elementos en el espacio ‘eigenbrain’ a partir de la misma, cuya transformación inversa será una imagen en el espacio MNI estándar.

Las imágenes sintéticas fueron analizadas visualmente, y se comprobó que son muy parecidas a las originales, e incluso expertos en el campo fueron incapaces de distinguir entre imágenes sintéticas y reales en unos tests preliminares. Los experimentos de análisis de clasificación y SPM que desarrollamos en el capítulo probaron que las imágenes pueden predecir ejemplos del mundo real, y a la vez, ser independientes de la base de datos originales. El análisis mediante SPM demostró que las diferencias más significativas entre clases estaban situadas en los mismos lugares en las imágenes reales y las simuladas, de modo que los análisis realizados sobre ambas bases de datos son equivalentes.

Conclusiones

Las diversas contribuciones de esta tesis han sido probadas en una serie de experimentos, de las cuales se obtienen las siguientes conclusiones:

- Hemos propuesto dos grandes enfoques para solucionar el *Small Sample Size problem* en bases de datos de neuroimagen: mediante extracción de características y mediante el incremento del tamaño muestral. Estos dos enfoques son complementarios, y se pueden utilizar en conjunto para construir nuevos sistemas CAD más fiables.
- Los algoritmos de extracción de características propuestos en los capítulos 4, 5 y 6 alcanzan un alto rendimiento en el diagnóstico diferencial de enfermos y controles, particularmente cuando se aplican a las bases de datos de Alzheimer y Parkinson. El mayor reto es utilizar estos algoritmos para estudiar la progresión de enfermedades neurodegenerativas, y en particular, la conversión de Mild Cognitive Impairment (MCI) a demencia.
- Las técnicas de descomposición de imagen utilizadas en el capítulo 4 son capaces de reducir significativamente la carga computacional cuando analizamos imágenes médicas, a la vez que mantienen o incluso incrementan el rendimiento de los sistemas CAD propuestos, que han probado sobradamente su capacidad de discriminación en modalidades funcionales como SPECT and PET.
- El análisis de texturas se ha revelado como una herramienta muy útil para analizar imágenes DaTSCAN en el capítulo 5. Características de textura como la tendencia a formar *clusters* o la homogeneidad han modelado con gran precisión la gran variedad de formas e intensidades que se encuentran en el cuerpo estriado en estas modalidades, y que pueden ser relacionados con el Parkinson.
- El marco SBM que proponemos en el capítulo 6 puede considerarse una herramienta muy poderosa (y además, extensible) para el análisis de imágenes estructurales. Los mapas estadísticos y de textura que se derivan de su aplicación facilitan la inspección visual para detectar los cambios relacionados con el Alzheimer, a la vez que realizan una reducción significativa del espacio de características. Utilizando SBM hemos podido diagnosticar el Alzheimer a través de sus patrones estructurales, principalmente debidos a la textura o cambios en la densidad de tejidos, alcanzando una gran precisión.

- Tras encontrar una gran heterogeneidad en bases de datos de imagen estructural adquiridas en múltiples centros, pudimos reducir este efecto mediante la aplicación del algoritmo SWPCA (capítulo 7). Este hecho fue detectado en la base de datos de autismo AIMS-MRI, donde un clasificador era capaz de detectar mejor diferencias entre sitios que entre clases. Tras aplicar SWPCA, las diferencias entre sitios se redujeron casi al completo, mientras la precisión del diagnóstico diferencial se mantuvo similar.
- El análisis de las diferencias más significativas entre individuos autistas y controles no fue capaz de detectar cambios significativos, o en todo caso, que dichos cambios eran tan heterogéneos que no se podían establecer patrones comunes. Esto apunta a que gran parte de los resultados obtenidos cuando se analiza la estructura de cerebros autistas puede ser debida a diferencias en la adquisición, tal y como se ha revelado en publicaciones recientes [21].
- Finalmente, el procedimiento de síntesis de imágenes funcionales cerebrales propuesto en el capítulo 8 fue capaz de incrementar el número de sujetos de una base de datos de forma segura, a la vez que mantenía las características individuales. Hemos probado aquí que las imágenes simuladas pertenecen a la misma distribución estadística de las imágenes originales y pueden predecir imágenes del mundo real, e incluso ser usadas para entrenar a futuros profesionales.

Todos los algoritmos y marcos propuestos son complementarios entre sí. Se puede utilizar el algoritmo de síntesis para doblar el tamaño de una base de datos existente, y luego utilizar los métodos de descomposición para reducir el espacio de características. O podemos “limpiar” una base de datos multicentro con SWPCA y luego visualizar las diferencias más significativas con SBM. Todo lo que proponemos con esta tesis son solo gotas de agua que contribuyen al gran océano que es el procesado de neuroimagen, en un humilde intento de mejorar la velocidad, fiabilidad e información suministrada por las técnicas de diagnóstico actuales.

AGRADECIMIENTOS

El trabajo que aquí se presenta se ha financiado mediante un contrato de investigación asociado al proyecto de excelencia P11-TIC- 7103 de la Consejería de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía, Spain). Para la estancia en la Universidad de Cambridge se ha contado con una financiación complementaria del Vicerrectorado de Relaciones Internacionales de la Universidad de Granada y el CEI BioTic, bajo el programa Convocatoria de Movilidad Internacional de Estudiantes de Doctorado Curso 2013/2014. Asimismo, a lo largo de este periodo predoctoral se ha contado con financiación complementaria de los proyectos P09-TIC-4530 de la Junta de Andalucía y TEC2008-02113, TEC2012-34306 y TEC2015-64718-R del MINECO/FEDER.

Esta tesis no sería posible de no haber contado con Juanma y Javi como directores del proyecto. A vosotros os debo la máxima gratitud por vuestra calidad personal y profesional. Porque me habéis guiado desde el primer día hasta el último, a la vez que me habéis dejado una grandísima libertad en mi trabajo. Hemos recorrido todo este camino juntos, y me llevo eso que llaman una formación ‘integral’, intelectual, laboral y en valores. No podría imaginar mejores directores para esta tesis. Gracias por todo este camino, y todo lo que nos queda.

Al resto del grupo SiPBA y colaboradores; a Diego Salas, que fue el que me trajo a este grupo, por esas cosas de conocerse con una guitarra delante; Fermín, Ignacio, Diego Castillo (novatillo!), Andrés, Brahim, Laila, Miriam, Rosa, Alberto... Gracias por dar vida a este grupo de investigación, donde se respira ciencia y buen hacer.

A John Suckling, por acogerme como uno más -y, si hay suerte, en próximas ocasiones- y a todos los compañeros del departamento de Psiquiatría y el departamento del Regal en Cambridge. Gracias porque allí se demuestra que multidisciplinariedad es un palabra que solo significa ‘tenemos mucho que aprender unos de otros’.

Al departamento de Teoría de la Señal, Telemática y Comunicaciones, donde he aprendido muchas cosas y donde he tenido el gusto de descubrirme en la docencia, un trabajo que al final me ha encantado. Y a la Universidad de Granada, donde me he formado todos estos años, y donde miles de personas trabajan para que siga siendo una de las referencias a nivel español e internacional.

A todos los compis del CITIC, muy especialmente a mis compañeros de despacho, donde al final hemos hecho piña y estamos como en casa: Fermín, Pablo, Rober, Marta, Gabri, Diego y Samuel. Por todos esas encuestas y consejos de colores, charlas mañaneras y cafeses varios.

A mis padres, que han estado apoyándome desde el primer día, incluso cuando no había muchas perspectivas de financiación. Sois un ejemplo de trabajo y dedicación, y gracias a vuestro tesón estoy escribiendo estas líneas. Siempre recordaré vuestro “nosotros no tenemos tierras, ni negocios, lo único que podemos daros es una educación”. No es tangible, y quizá por eso, es el mejor patrimonio que se puede heredar. Gracias por todo lo que me habéis transmitido, cuidado y dado a lo largo de toda mi vida.

Gracias a mi familia, y a todos los amigos que me habéis aguantado en esos: “no puedo, que mañana hay que levantarse temprano para la tesis”; que me habéis preguntado, os habéis preocupado por qué tal andamos. Sin duda, esto también es mérito vuestro. A Jiménez y Rocío, que este es vuestro año, a Elena, compi Teleco.. A todos los que participasteis en la encuesta para decidir el mapa de color de las imágenes médicas, sin duda el impacto de vuestros votos os garantiza una presencia en este documento.

A la banda sonora de mi vida: los muy grandes Adversos, el Coro de Ciencias, el de Consolación y los maravillosos componentes de Musicalarte. Y a la música en general, ese virus que me infecta día tras día, el lenguaje de lo inexpresable. El árbol de la vida cuyo fruto fue, casualmente, mi interés por las señales, sin el cual no habría trabajado en esta tesis.

A los compañeros del Three-minute thesis, por todos esos momentos que hemos vivido juntos, formación, entrevistas, ensayos y en definitiva, todo lo que hemos crecido gracias a esta aventura. Y a Carlos, Susana y Emilio, por guiarnos en ese camino.

A tí, que estás leyendo estas líneas, por interesarte por nuestra humilde aportación a la ciencia. Esperamos que sea de tu agrado y que la disfrutes una mínima parte de todo lo que la hemos disfrutado nosotros.

Gracias a tí, Inma. Por todo lo que hemos pasado juntos, por querer compartir una vida conmigo. Porque día tras día me demuestras que el tesón y la fuerza que tienes es única. Porque pese a tu dulzura y tu aparente fragilidad, sigues resistiendo a todas las adversidades, luchando por lo que crees. Porque eres muy grande, mucho más de lo que crees, y de lo que algunos ciegos son incapaces de percibir. Tú eres la luz que veo cada mañana y cada noche. El potencial de mi energía. Mi hogar, al que siempre quiero volver. Gracias.

CONTENTS

1	Introduction	1
1	Introduction	3
1.1	Motivation	3
1.2	The Small Sample Size Problem	4
1.3	Aims and Objectives	5
1.4	Organization of this Thesis	7
1.5	Contributions	7
1.5.1	Articles	7
1.5.2	Conferences	8
1.5.3	Books	9
2	State of the Art	11
2.1	Introduction to Neuroimaging	11
2.1.1	Magnetic Resonance Imaging (MRI)	11
2.1.2	Single Photon Emission Computed Tomography (SPECT)	13
2.1.3	Positron Emission Tomography (PET)	14
2.2	Medical Background	14
2.2.1	Alzheimer's Disease (AD)	14
2.2.2	Parkinsonism (PKS)	16
2.2.3	Autism Spectrum Disorder (ASD)	17
2.3	Voxelwise Analyses	18
2.3.1	Statistical Parametric Mapping (SPM)	18
2.3.2	Voxel Based Morphometry (VBM)	20
2.3.3	The Multiple Comparisons Problem	20
2.4	Machine Learning in Neuroimaging	23
2.4.1	Voxels As Features (VAF)	24
2.4.2	Multivariate Analyses	24
3	General Methodology	27
3.1	Spatial Preprocessing	27
3.1.1	Spatial Normalization or Registration	28
3.1.2	Segmentation	31
3.2	Intensity Normalization	31
3.3	Evaluation Parameters and Methodology	33
3.3.1	Cross-validation	33
3.3.2	Classification Performance	34

II	Reducing the Feature Space	37
4	Image Decomposition	39
4.1	Feature Selection	40
4.1.1	t-test	41
4.1.2	Relative Entropy	41
4.1.3	Mann-Whitney-Wilcoxon (MWW) U-test	41
4.2	Decomposition Algorithms	42
4.2.1	Factor Analysis (FA)	42
4.2.2	Independent Component Analysis (ICA)	45
4.3	Results	47
4.3.1	Alzheimer's Disease	48
4.3.2	Parkinson's Disease	53
4.4	Discussion	61
5	Texture Features	67
5.1	Introduction	67
5.2	Methodology	68
5.2.1	Volume selection	68
5.2.2	Haralick Texture Analysis	69
5.2.3	Experiments	71
5.3	Results	72
5.3.1	Experiment 1: Individual Texture Features	72
5.3.2	Experiment 2: Selected Texture Features	76
5.4	Discussion	80
6	Spherical Brain Mapping	85
6.1	Introduction	85
6.2	Spherical Brain Mapping (SBM)	86
6.2.1	Layered Extension	88
6.2.2	Volumetric Radial LBP (VRLBP)	88
6.2.3	Anatomical Reference	92
6.3	Sampling Paths via Hidden Markov Models (HMMs)	92
6.3.1	Radial Texture Features	96
6.4	Evaluation	99
6.5	Results	100
6.5.1	Experiment 1: Original and VRLBP Spherical Brain Mapping	100
6.5.2	Experiment 2: Layered Extension	104
6.5.3	Experiment 3: HMM on Synthetic Datasets	108
6.5.4	Experiment 4: Feature Selection using HMM Paths	110
6.5.5	Experiment 5: Texture SBM Maps based on HMM Paths	112

6.6	Discussion	114
6.6.1	Spherical Brain Mapping	114
6.6.2	Paths via HMM	116
6.6.3	SBM Framework Comparison	119
III	Increasing the Sample Size	123
7	Significance Weighted Principal Component Analysis	125
7.1	Significance Weighted Principal Component Analysis (SWPCA)	126
7.1.1	Principal Component Analysis (PCA)	127
7.1.2	One-Way Analysis Of Variance (ANOVA)	129
7.1.3	Weighting Function	129
7.2	Evaluation of SWPCA	130
7.3	Results for AIMS-MRI Dataset	132
7.3.1	Experiment 1: Effect of Acquisition Site	132
7.3.2	Experiment 2: Within-site Between-Group Differences	133
7.3.3	Experiment 3: Effect of SWPCA on Group Differences	133
7.4	Discussion	141
8	Simulation of Functional Brain Images	147
8.1	Simulation Methodology	148
8.1.1	PCA Decomposition	148
8.1.2	Density Estimation	148
8.1.3	Brain Image Synthesis	149
8.2	Experimental Setup	150
8.3	Results	151
8.3.1	Baseline	151
8.3.2	Experiment 1: Predictive Power of the Simulated Images	153
8.3.3	Experiment 2: Independence of the Simulated Images	154
8.4	Discussion	154
IV	General Discussion and Conclusions	165
9	General Discussion and Conclusions	167
9.1	General Discussion	167
9.1.1	Discussion on the algorithms	167
9.1.2	Discussion on the Disorders	169
9.2	Conclusions	172
V	Appendix	175
A	Datasets	177
A.1	MRI	177

A.1.1	ADNI-MRI and the Alzheimer's Disease Neuroimaging Initiative	177
A.1.2	AIMS-MRI, MRC-AIMS Consortium	178
A.2	PET	180
A.2.1	ADNI-PET, Alzheimer's Disease Neuroimaging Initiative	180
A.3	SPECT	181
A.3.1	VDLN-HMPAO, Virgen de las Nieves	181
A.3.2	VDLN-DAT, Virgen de las Nieves	182
A.3.3	VDLV-DAT, Virgen de la Victoria Hospital	183
A.3.4	PPMI-DAT, Parkinson's Progression Markers Initiative	184
B	Background on Support Vector Machines (SVMs)	187
	Bibliography	189

LIST OF FIGURES

Figure 1.1	Illustration of one and two-dimensional spaces.	4
Figure 1.2	Structure of the thesis.	6
Figure 2.1	Example of T ₁ and T ₂ -weighted MRI images.	12
Figure 2.2	Example of SPECT images.	15
Figure 2.3	Example of a PET-FDG image.	15
Figure 2.4	Example of a SPM analysis on a PET dataset.	19
Figure 2.5	Illustration of a typical neuroimaging CAD system.	24
Figure 2.6	Example of the cortical thickness of a subject.	25
Figure 3.1	Typical pre-processing pipeline in MRI	28
Figure 3.2	Comparison of the affine registration and the application of non-linear transformations to the images	30
Figure 3.3	Comparison between different types of intensity normalization.	33
Figure 3.4	Evolution of bias and variance in CV.	34
Figure 4.1	Illustration of how decomposition algorithms work.	39
Figure 4.2	Illustration of the system used in chapter 4 .	40
Figure 4.3	Original PET image and its reconstruction using FA or ICA.	43
Figure 4.4	Variance of reconstruction error in FA.	44
Figure 4.5	Average performance of the AD datasets in FA.	49
Figure 4.6	Average performance of the AD datasets in ICA.	50
Figure 4.7	Performance at the operation point for the AD datasets, over the number of selected voxels.	51
Figure 4.8	Performance at the operation point for the AD datasets, over the number of components.	52
Figure 4.9	Average performance of the PKS datasets in FA.	55
Figure 4.10	Average performance of the PKS datasets in ICA.	56
Figure 4.11	Performance at the operation point for the PKS datasets, over the number of selected voxels.	58
Figure 4.12	Performance at the operation point for the PKS datasets, over the number of components.	59
Figure 4.13	Comparison between the different filtering methods in ADNI-PET.	61

Figure 4.14	Comparison between the different filtering methods in VDLN-HMPAO.	62
Figure 4.15	Comparison between the different filtering methods in PPMI-DAT.	65
Figure 5.1	Schema of the proposed Texture-based CAD system.	67
Figure 5.2	Comparison of the different I_{th} values.	69
Figure 5.3	Evolution of the average accuracy with the intensity threshold.	73
Figure 5.4	Violin plot of all accuracy values, grouped by database.	75
Figure 5.5	Accuracy obtained by averaging all accuracy values using a given volume selection threshold I_{th} .	77
Figure 5.6	Average accuracy computed for each selection criteria over N in Experiment 2.	78
Figure 5.7	Distribution of the 100 first selected features by means of the different selection methods.	82
Figure 6.1	Flow diagram of the procedure used in the textural analysis of projected MR brain images.	85
Figure 6.2	Illustration of the computation of the mapping vector $v_{\theta,\varphi}$, the angles θ and φ and the r-neighbourhood of v (see Section 6.2.2).	86
Figure 6.3	Resulting GM and WM maps of the same control subject using the six proposed measures: surface, thickness, number of folds, average, entropy and kurtosis.	89
Figure 6.4	Example of the layered approach using the average measure on GM maps.	90
Figure 6.5	Example of how the basic LBP is computed.	90
Figure 6.6	An example of the VRLBP projection for GM and WM Tissues.	91
Figure 6.7	SBM mapping of different cortical regions.	93
Figure 6.8	SBM mapping of some subcortical regions and organs.	93
Figure 6.9	Set of HMM based paths over the MRI DARTEL template.	97
Figure 6.10	SBM t-maps under the AD vs CTL for GM and WM images.	101
Figure 6.11	t-maps for the VRLBP-SBM in the AD vs CTL scenario.	102
Figure 6.12	Performance for the different SBM approaches over the: a) GM and b) WM.	105
Figure 6.13	Performance of the four-layered mappings.	106
Figure 6.14	t-maps that present the level of statistical relevance in the AD vs CTL paradigm.	107

Figure 6.15	Path traced over a gaussian mixture distribution of 4 isotropic gaussian kernels.	109
Figure 6.16	HMM path computed inside a density distribution defined by an helix.	109
Figure 6.17	Simulation of the HMM-based path tracing over an Iberian Peninsula height map, interconecting different cities.	110
Figure 6.18	canonical paths computed in each direction (θ, φ). Each path's colour represent the accuracy in a differential diagnosis. Only one in every five paths are shown for clarity purposes.	111
Figure 6.19	Example of the resulting hybrid HMM-SBM texture maps.	113
Figure 6.20	Performance at the operation point for the different mappings over the GM and WM, compared with the performance of VAF.	114
Figure 6.21	Paths that obtain more than 75% accuracy, and a three-dimensional representation of the structures crossed by them.	118
Figure 7.1	Summary of the SWPCA algorithm, along with its context in the pipeline used in this article.	127
Figure 7.2	Weighting function $\Lambda_c(p_c, p_{th})$ used in SWPCA.	130
Figure 7.3	Box-plot of the distribution of the component scores at each site of the AIMS-MRI dataset (see Sections 7.3 and A.1.2) in the four first components.	131
Figure 7.4	Brain t-map (VBM) of significant ($p < 0.01, t > 2.57$) GM and WM between-group differences using $qT_1, qT_2, synT_1$, GM and WM modalities after applying SWPCA to remove site effects.	134
Figure 7.5	Brain t-map (VBM) of significant ($p < 0.01, t > 2.57$) GM and WM differences in ASD using $qT_1, qT_2, synT_1$, GM and WM maps before and after applying SWPCA to remove site effects.	138
Figure 7.6	Brain Z-map (CBM) of significant ($p < 0.01, t > 2.57$) GM and WM differences using $qT_1, qT_2, synT_1$, GM and WM maps before and after applying SWPCA to remove site effects.	139
Figure 7.7	Location of the significant region labelled D (posterior part of the superior temporal gyrus) within the MNI template.	143

- Figure 7.8 The template used in this work compared to two of the participants with abnormal ventricle size (21016 and 21018). 144
- Figure 8.1 Schema of the brain image synthesis algorithm. 147
- Figure 8.2 Illustration of the four first eigenbrains in PCA. 156
- Figure 8.3 Comparison between the MVN and KDE PDF estimation methods for the first component and the three groups, setting the histogram as reference. 156
- Figure 8.4 Pairplot plotting the coordinates of the subjects in the ADNI-PET database over the four first eigenbrains. In the diagonal, distribution of the classes estimated using KDE for each eigenbrain. 158
- Figure 8.5 Examples of some original and simulated subjects from the ADNI-PET dataset. 159
- Figure 8.6 Examples of some original and simulated subjects from the PPMI dataset. 160
- Figure 8.7 SPM Analysis of the ADNI dataset (AD vs CTL). 161
- Figure 8.8 SPM Analysis of the ADNI dataset (CTL vs MCI). 162
- Figure 8.9 SPM Analysis of the ADNI dataset (MCI vs AD). 163
- Figure 8.10 SPM Analysis of the PPMI dataset. 164

LIST OF TABLES

Table 3.1	Confusion matrix and its parts	35
Table 4.1	Performance values for the AD datasets	53
Table 4.2	Performance values for the PKS datasets	60
Table 4.3	Percentage of overlap between the selected areas by each method and the AAL atlas regions.	63
Table 5.1	Maximum scoring feature (single approach).	74
Table 5.2	Maximum scoring feature (cumulative approach).	76
Table 5.3	Best results of the experiment 2 per database, normalization and selection criteria.	79
Table 5.4	Comparison between our proposed system and other PD diagnosis systems in the literature.	83
Table 6.1	Performance values (Average \pm Standard Deviation) for the different SBM approaches.	103
Table 6.2	Performnace values for using the HMM paths in feature selection.	112
Table 6.3	Performance values (\pm SD) for the hybrid HMM-SBM maps, using some statistical measures.	112
Table 6.4	Performance values (\pm SD) for each of the 10 texture features.	113
Table 6.5	Comparison between our algorithm performance values (best values for selected voxels in all paths and texture features) (\pm SD) and other methods in the bibliography	120
Table 7.1	Between-site classification accuracy (\pm standard deviation) for different modalities and masks without and with SWPCA correction.	135
Table 7.2	Classification accuracy (acc), sensitivity (sens) and specificity (spec) \pm standard deviation for each modality and mask using the participants acquired at the LON and CAM sites.	136
Table 7.3	Classification accuracy (acc), sensitivity (sens), and specificity (spec) \pm STD for the different modalities and masks using ALL, before and after applying SWPCA.	140

Table 8.1	Baseline experiment, in which we evaluate the performance of a VAF system under the different scenarios of the two datasets.	152
Table 8.2	Performance of Experiment 1, demonstrating the predictive ability of the simulated images over the real dataset.	153
Table 8.3	Performance of Experiment 2, proving the independence of the simulated images with respect to the originals.	155
Table A.1	Summary of the datasets used in this thesis.	177
Table A.2	Demographics of the ADNI-MRI dataset.	178
Table A.3	Demographics of the AIMS-MRI dataset.	179
Table A.4	Demographics of the ADNI-MRI dataset.	181
Table A.5	Demographic of the VDNL-HMPAO dataset.	182
Table A.6	Demographics of the VDNL-DAT dataset.	183
Table A.7	Demographics of the VDLV-DAT dataset.	184
Table A.8	Demographics of the PPMI-DAT dataset.	185

ACRONYMS

AAL	Automated Anatomical Labeling
AC	Anterior Commissure
AD	Alzheimer's Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANOVA	Analysis Of Variance
ASD	Autism Spectrum Disorder
CAD	Computer Aided Diagnosis
CBM	Component Based Morphometry
CDF	Cumulative Density Function
CSF	cerebro-spinal fluid
CT	Computed Tomography
CTL	Control Subject
CV	Cross-validation
DAT	Dopamine Transporters
DTI	Diffusion Tensor Imaging
EM	Expectation-Maximization
EVD	eigen-value decomposition
FA	Factor Analysis
GLCM	Grey Level Co-occurrence Matrix
GLM	General Linear Model
GM	grey matter
FBP	Filtered Back Projection

FDR	False Discovery Rate
fMRI	functional MRI
FN	False Negative
FP	False Positive
FWE	Family Wise Error rate
HMM	Hidden Markov Model
ICA	Independent Component Analysis
KDE	Kernel Density Estimation
KL	Kullback-Leibler
LBP	Local Binary Patterns
LOO	Leave-One-Out
MCI	Mild Cognitive Impairment
MNI	Montreal Neurological Institute
MRC-AIMS	Medical Research Council Autism Imaging Multicentre Study
MRI	Magnetic Resonance Imaging
MVN	Multivariate Normal distribution
MWW	Mann-Whitney-Wilcoxon
PCA	Principal Component Analysis
PD	Parkinson's Disease
PET	Positron Emission Tomography
PDF	Probability Density Function
PKS	Parkinsonism
PLS	Partial Least Squares
PPMI	Parkinson's Progression Markers Initiative
qT ₁	quantitative T ₁ - weighted

qT ₂	quantitative T ₂ - weighted
RBF	Radial Basis Function
rCBF	regional Cerebral Blood Flow
RF	radiofrequency
ROI	Region of Interest
SBM	Spherical Brain Mapping
SNR	Signal-To-Noise Ratio
SPM	Statistical Parametric Mapping
SPM8	Statistical Parametric Mapping Software, version 8
SPECT	Single Photon Emission Computed Tomography
SVA	Surrogate Variable Analysis
SVC	Support Vector Classifier
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SWEDD	subjects without evidence of dopaminergic deficit
SWPCA	Significance Weighted Principal Component Analysis
synT ₁	simulated T ₁ - weighted Inversion Recovery
TN	True Negative
TP	True Positive
VAF	Voxels As Features
VBM	Voxel Based Morphometry
VDLN	Virgen de las Nieves Hospital
VDLV	Virgen de la Victoria Hospital
VRLBP	Volumetric Radial LBP
WM	white matter

Part I

INTRODUCTION

1

INTRODUCTION

1.1 Motivation

In recent years, there has been a rise in the use of neuroimaging in the clinical practice. It has improved and speeded the procedure of diagnostic, providing unprecedented insight into the brain. Neuroimaging is very extended in research as well. Different fields such as psychiatry, neurology, psychology, behavioural science or biology make extensive use of brain imaging in their studies.

The basis of most of these studies is common: a procedure by which a representative set of subjects is recruited, the fulfilment of an experiment on (or by) each subject and a statistical analysis of the acquired data. Particularly, when studying a certain disease, it is common to recruit subjects affected by the disease and a matched population of non-affected, healthy subjects, usually known as Control Subjects (CTLs). Then, in this typical example, both affected and CTLs are scanned, and brain anatomy or function is analysed using statistical tools. The result of this analysis is frequently a list of significant differences between structure or function that could be linked to the disease.

Computer Aided Diagnosis (CAD) systems provide a set of tools to help setting up and performing these studies. It is currently a thriving area of research involving multidisciplinary teams, combining computer science, mathematics, medicine, artificial intelligence, statistics, machine learning, and many others [7]. The main aim is to assist clinicians in the procedure of diagnosis and study of the diseases by providing software that can effectively recognize disease patterns, characterize differences and make predictions.

One fundamental issue often found in this studies is the sample size. The number of subjects frequently ranges from tens to hundreds, whereas the number of features (namely voxels) to be analysed can add up to millions. This causes the so-called *Small Sample Size Problem* [138] which negatively affects the statistical power of any experiment performed using these datasets [31].

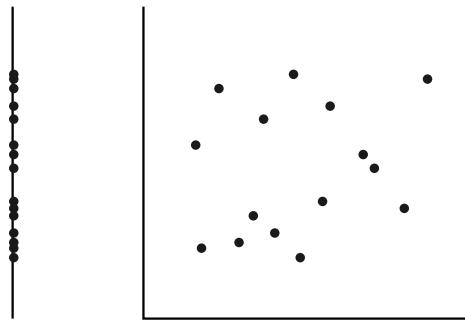


Figure 1.1: Illustration of the separation between points in one-dimensional and two-dimensional spaces.

1.2 The Small Sample Size Problem

The *Small Sample Size Problem* refers to a problem that arises when the ratio between the number of subjects and the number of features is large. Think of, e.g., 15 points in a one-dimensional line, as in Figure 1.1. If we think of a subject as a vector, we would have 15 subjects in a one-dimensional space. Now, imagine that we add a second dimension. It is easy to see that our subjects would be more distant than in the one-dimensional line. And the same would happen if we move to three, ten or thousands of dimensions. The farther our points are, the more difficult is for a statistical tool to extract information. That is what we call *almost empty spaces* [120, 138], in contrast to *dense spaces*, where points are closer.

Neuroimaging provides hundreds of thousands, or even millions of voxels, in what could mean millions of features. That implies that any calculation performed in those almost empty spaces will eventually lack information. This implies a loss of statistical power of the methods used, usually producing false negatives (the system is unable to detect real signal) and false positives (the system detects signal where there is not). These are known in statistics as Type I and Type II errors respectively.

In differential diagnosis studies, the small sample size problem leads to wrong conclusions about where real differences are located. This, in addition to untracked confounding variables are some of the fundamental sources of non-reproducibility in current neuroimaging studies [31].

The solution might seem straightforward: increase sample size. But this is not always possible, since neuroimaging studies do their best at recruiting as

many people as they can with a limited budget. Many efforts have been put into establishing multi-centre collaborations that allow the recruitment of a larger population, but despite offering a higher statistical power, these studies still suffer from a number of confounding variables such as population bias or scanner differences [21]. In chapters 7 and 8 we explore different approaches to this solution.

Another option involves reducing the number of features, via feature selection or feature extraction. This has been widely used in computed-aided methodology for neuroimaging very successfully [7, 54, 56, 81, 95], and solutions using this approach will be presented in chapters 4, 5 and 6.

The Small Sample Size problem is directly related to the *Curse of Dimensionality* [170], which proves that, in contrast to what might be expected, once a certain classifier performance has been achieved, it holds or even decreases when feeding more features to the classifier. The problem also affects statistical hypothesis testing, a tool widely used for inference in neuroimaging, in what is known as the *Multiple Comparisons* problem [62], a particular field that is still being studied.

1.3 Aims and Objectives

This thesis aims to contribute new approaches to overcome the small sample size problem in neuroimaging. This can lead to more accurate CAD systems by reducing the number of false positives, increasing the reliability of their results.

We will take two different approaches, as commented in the previous section: increasing the sample size and reducing the feature space. Therefore, we can define the following objectives:

- Develop and evaluate algorithms that reduce the feature space, in which is usually known in the field as feature extraction and feature selection strategies.
- Develop and evaluate new strategies to increase the sample size in neuroimaging studies.

Most of the studies in the literature focus on the first objective. Feature extraction algorithms that use Principal Component Analysis (PCA) [13, 60] or Partial Least Squares (PLS) [40], among others, are widely studied. We have developed three different approaches to those:

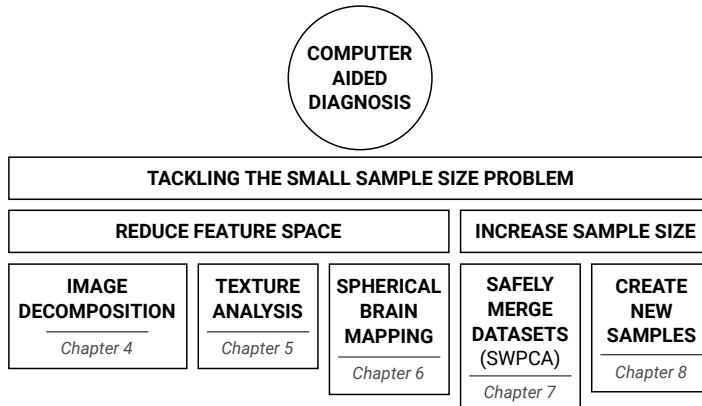


Figure 1.2: Structure and connexions between the different strategies proposed in this thesis, organized by chapters and parts.

- A combination of image decomposition algorithms and feature selection. In this approach we used three statistical tests to select the most significant voxels from the images, and then applied Factor Analysis (FA) and Independent Component Analysis (ICA) to decompose the data and significantly reduce their feature space.
- A feature extraction based on texture analysis.
- A novel strategy called Spherical Brain Mapping (SBM). This feature extraction technique uses a spherical coordinate system to map statistical measures to a bidimensional plane. It builds paths used as feature selection vectors where several measures are computed.

On the other hand, we have evaluated newer ways to safely increase sample size in neuroimaging studies. We have developed:

- A system to reduce undesired variance in structural multicentre studies, called Significance Weighted Principal Component Analysis (SWPCA). This system is intended to reduce the amount of false positives in large collaborations, providing more homogeneous images and improving their statistical power.
- An algorithm to simulate functional brain images using existing data, and therefore, increase sample size.

1.4 Organization of this Thesis

This thesis work is organized in four parts plus appendices, each of which is subdivided in several chapters. In the first part, we introduce the motivations and main aims of this work (chapter 1), examine the state of the art in medicine, neuroimaging and CAD systems (chapter 2) and present a general methodology that will be followed throughout this thesis, including pre-processing and evaluation (chapter 3).

Parts ii and iii refers to each of the solutions outlined in the previous section, and disaggregated in Figure 1.2. In part ii we focus on the feature reduction techniques, including decomposition methods (chapter 4), texture analysis (chapter 5), and the novel algorithm Spherical Brain Mapping (chapter 6). On the other hand, Part iii is focused on two different strategies used to increase the sample size: the Significance Weighted Principal Component Analysis algorithm (chapter 7), used to safely merge structural images acquired at different centres, and a neuroimage simulation algorithm (chapter 8) that can be used to extend existing functional datasets.

Finally, in Part iv we provide a general discussion of the results presented in this thesis, conclusions about the methods and prospective work that could be performed with this basis.

1.5 Contributions

Some ideas and figures have appeared previously in the following publications, that we divide here in articles, conference presentations and book chapters.

1.5.1 Articles

F.J. Martínez-Murcia et al. "On the Brain Structure Heterogeneity of Autism: Parsing out Acquisition Site Effects With Significance-Weighted Principal Component Analysis". In: *Human Brain Mapping* 38.3 (Mar. 2017), pp. 1208–1223.

DOI: [10.1002/hbm.23449](https://doi.org/10.1002/hbm.23449).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, and A. Ortiz. "A Spherical Brain Mapping of MR Images for the Detection of Alzheimer's Disease". In: *Current Alzheimer Research* 13.5 (Mar. 2016), pp. 575–588.

DOI: [10.2174/1567205013666160314145158](https://doi.org/10.2174/1567205013666160314145158).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, and A. Ortiz. "A Structural Parametrization of the Brain Using Hidden Markov Models-Based Paths in Alzheimer's Disease". In: *International Journal of Neural Systems* 26.7 (Nov. 2016), p. 1650024.

DOI: [10.1142/S0129065716500246](https://doi.org/10.1142/S0129065716500246).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, and A. Ortiz. "Automatic detection of Parkinsonism using significance measures and component analysis in DaTSCAN imaging". In: *Neurocomputing* 126 (Feb. 2014), pp. 58–70.

DOI: [10.1016/j.neucom.2013.01.054](https://doi.org/10.1016/j.neucom.2013.01.054).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, M. Moreno-Caballero, and M. Gómez-Río. "Parametrization of textural patterns in ^{123}I -ioflupane imaging for the automatic detection of Parkinsonism". In: *Medical Physics* 41.1 (2014), p. 012502.

DOI: [10.1118/1.4845115](https://doi.org/10.1118/1.4845115).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, C.G. Puntonet, and I.A. Illán. "Functional activity maps based on significance measures and Independent Component Analysis". In: *Computer Methods and Programs in Biomedicine* 111.1 (July 2013), pp. 255–268.

DOI: [10.1016/j.cmpb.2013.03.015](https://doi.org/10.1016/j.cmpb.2013.03.015).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, C.G. Puntonet, and D. Salas-González. "Computer Aided Diagnosis tool for Alzheimer's Disease based on Mann-Whitney-Wilcoxon U-Test". In: *Expert Systems with Applications* 39.10 (Aug. 2012), pp. 9676–9685.

DOI: [10.1016/j.eswa.2012.02.153](https://doi.org/10.1016/j.eswa.2012.02.153).

1.5.2 Conferences

F.J. Martínez-Murcia, A. Ortiz, J. Manuel Górriz, J. Ramírez, and I.A. Illán. "A volumetric radial LBP projection of MRI brain images for the diagnosis of alzheimer's disease". In: *Artificial Computation in Biology and Medicine*. Vol. 9107. Springer Science + Business Media, 2015, pp. 19–28.

DOI: [10.1007/978-3-319-18914-7_3](https://doi.org/10.1007/978-3-319-18914-7_3).

F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I. Álvarez Illán, D. Salas-González, and F. Segovia. "Projecting MRI brain images for the detection of Alzheimer's Disease". In: *Studies in Health Technology and Informatics*. Vol. 207. 2014, pp. 225–233.

DOI: [10.3233/978-1-61499-474-9-225](https://doi.org/10.3233/978-1-61499-474-9-225).

- F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, and C.G. Puntonet. "Texture Features Based Detection of Parkinson's Disease on DaTSCAN Images". In: *Natural and Artificial Computation in Engineering and Medical Applications*. Vol. 7931 LNCS. PART 2. Springer Science + Business Media, 2013, pp. 266–277.
 DOI: [10.1007/978-3-642-38622-0_28](https://doi.org/10.1007/978-3-642-38622-0_28).
- F.J. Martínez, D. Salas-González, J.M. Górriz, J. Ramírez, C.G. Puntonet, and M. Gómez-Río. "Analysis of Spect Brain Images Using Wilcoxon and Relative Entropy Criteria and Quadratic Multivariate Classifiers for the Diagnosis of Alzheimer's Disease". In: *New Challenges on Bioinspired Applications*. Vol. 6687. Lecture Notes in Computer Science PART 2. Springer Science + Business Media, 2011, pp. 41–48.
 DOI: [10.1007/978-3-642-21326-7_5](https://doi.org/10.1007/978-3-642-21326-7_5).

1.5.3 Books

- F.J. Martinez-Murcia, J. Górriz, and J. Ramírez. "Computer Aided Diagnosis in Neuroimaging". In: *Computer-aided Technologies - Applications in Engineering and Medicine*. Ed. by Razvan Udroiu. 1st ed. InTech, Dec. 2016. Chap. 7, pp. 137–160. ISBN: 978-953-51-4895-1.
 DOI: [10.5772/64980](https://doi.org/10.5772/64980).
- F.J. Martinez-Murcia, J.M. Górriz, and J. Ramírez. "Feature Extraction". In: *Wiley Encyclopedia of Electrical and Electronics Engineering*. Wiley, 2016, Accepted.
 DOI: [10.1002/047134608X](https://doi.org/10.1002/047134608X).

2

STATE OF THE ART

We have already stated the motivation and objectives of this thesis. Now, we will describe in detail some of the more relevant issues for the state of the art. First, in section 2.1, we will make an introduction to the different neuroimaging modalities used in our experiments. Afterwards, we will provide some insights into the neurological and psychiatric disorders treated here in section 2.2 or the most extended voxel-wise analyses used in the neuroimaging community at section 2.3. Finally, at section 2.4 we will explore recent contributions to the field that use Machine Learning.

2.1 Introduction to Neuroimaging

Medical imaging refers to all types of 2D, 3D and 4D images used in clinical practice. These involve many different modalities, among them X-rays, ultrasound, endoscopy, microscopy, etc. In neuroimaging, the most extended is by far MRI, which provides intensity maps that represent the internal structure of the brain. Other modalities are aimed at studying the function of the brain, by injecting radioactive ligands that, bound to a receptor, can measure its distribution. This is the case of PET and SPECT.

2.1.1 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is perhaps the most widespread modality in neuroimaging, given its ability to visualize both structural and functional (in functional MRI) properties of the brain, and, in contrast to other imaging modalities that use radiotracers, is considered non-invasive. MRI uses strong magnetic fields to excite certain atomic nuclei, that can absorb and emit this energy.

MRI combines a constant magnetic field with a radiofrequency (RF) emission to excite the atomic nuclei present in corporal structure, resulting in a image of the distribution of certain atoms in the body. Most MRI systems use hydrogen atoms, since they are present in water (which adds up to around 70% of

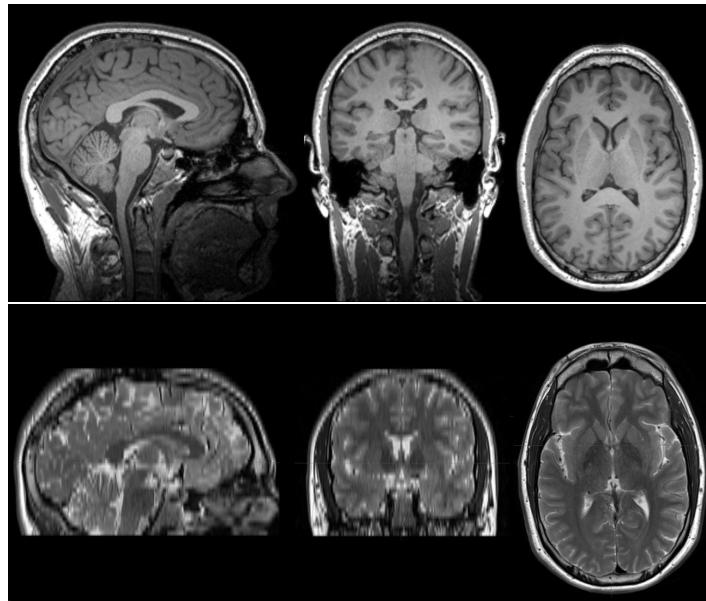


Figure 2.1: Example of T₁ and T₂-weighted MRI images of the same subject (me).

body mass) and the signal derived is stronger than other atoms, increasing the Signal-To-Noise Ratio (SNR), and therefore, the image quality.

The procedure uses a strong magnetic field B_0 to align the magnetic moment of the hydrogen nuclei in parallel or anti-parallel (depending of their initial spin). This way, the magnetic moment of all nuclei will increase up to a stable state, in contrast to their null value in absence of B_0 . Within this magnetic field, the hydrogen atoms precess around an axis along the direction of the field.

A given nuclei has a resonance frequency which is proportional to the intensity of B_0 , which, by using strong fields, allow us to resonate hydrogen far below potentially damaging frequencies. The precession frequency is determined by the Larmor equation (2.1):

$$f_0 = \frac{\gamma}{2\pi} B_0 \quad (2.1)$$

where γ depends on the nuclei, which in the case of hydrogen, $\gamma = 42.6$ MHz/T. When a subject is introduced in the MRI scanner, it is submitted to the magnetic field B_0 , so that the hydrogen nuclei are aligned to the field, with a precession frequency f_0 . Then, a RF pulse of the same frequency is generated, which is then absorbed by the nuclei, forcing them to place perpendicular to the field. Once the RF emission is interrupted, the nuclei return to its equilibrium state by means of a procedure called relaxation. In this procedure,

they emit part of the absorbed energy, which is then captured by a RF receptor. Usually, position information is encoded in the RF signal by varying B_0 using gradient coils.

The RF signal is measured during the relaxation time, and two different relaxation times are set: the T₁ (spin-lattice) relaxation time and the T₂ (spin-spin) relaxation time. The T₁ time is the time during which nuclei emit energy to the adjacent tissue and realign to the longitudinal plane (z axis), whereas the T₂ time refers to the time when nuclei realign to the transversal plane (y axis). These times are used to create T₁-weighted and T₂-weighted images (see Figure 2.1). T₁-weighted images allow to distinguish between grey matter (GM) and white matter (WM) in the cerebral cortex, to identify fatty tissue, and generally, obtain structural information. Conversely, T₂-weighted images are used to assess cerebro-spinal fluid (CSF) or to visualize and identify WM lesions.

2.1.2 Single Photon Emission Computed Tomography (SPECT)

The Single Photon Emission Computed Tomography (SPECT) is based on the principles of Computed Tomography (CT), by which a series of signal acquisitions at different angles can be reconstructed back into a bidimensional distribution of the signal. In SPECT, a gamma photon emitting radioisotope is linked to a pharmaceutical that binds to a given biomarker, generating a radiopharmaceutical or agent. This agent is injected into the patient, and after a certain time in which the radiopharmaceutical is distributed, the patient is introduced into the SPECT-CT scan.

Afterwards, the scanner performs a series of acquisitions at different planes and angles from the body, from which the gamma signal is measured. For each plane, all acquisitions at each angle are pooled and a single two-dimensional image is reconstructed using a Filtered Back Projection (FBP) algorithm, or Radon inversion formula [72], which derives from the Fourier's Theorem. A total of 180 projections per plane, using an angular resolution of 2 degrees, are usually taken.

There exist a wide variety of radiopharmaceuticals used in clinical practice, and therefore, we will only focus on the two varieties used in this thesis. First, we use an agent called ^{99m}Tc-HMPAO, which consists of two stereoisomers of hexametazime (HMPAO) linked to the radioisotope technetium 99-metastable. This agent is usually used to assess regional Cerebral Blood Flow (rCBF), which can be used to diagnose neurological diseases or cancer.

Additionally, we use images generated using the agent Ioflupane (^{123}I), a cocaine analog with high binding affinity for Dopamine Transporters (DAT). It is commonly known for its tradename, DaTSCAN, and is used fundamentally in the assessment of Parkinson's Disease (PD), given that the disease is associated with a loss of dopaminergic neurons in the striatal region.

2.1.3 Positron Emission Tomography (PET)

The Positron Emission Tomography (PET) is a technique similar to SPECT, but in this case, the agent used and the equipment is designed to deal with a pair of gamma photons resulting of the annihilation of a positron with its corresponding antiparticle, the electron. The pair of photons are generated in opposite directions, and the detection depends on them being simultaneously or coincidently detected at the receptor. The receptor comprises a scintillator which emits light when the gamma photon incides, and a detector, usually a photomultiplier tube or silicon avalanche photodiodes.

It uses the same FBP algorithm as SPECT in the reconstruction of the images, and a similar strategy for acquiring the signal at different angles. However, the amount of data is smaller than in SPECT, and therefore, the reconstruction procedure is harder. As a result, PET scanner operation is considered more costly than SPECT [3].

The agent used in the images that we have processed is Fludeoxyglucose (^{18}F or FDG). It is a glucose analogue that allows us to measure the glucose metabolism in the brain. It is widely used in neurology [129] and cancer detection [111], since it can be correlated with cellular activity.

2.2 Medical Background

2.2.1 Alzheimer's Disease (AD)

Alzheimer's Disease (AD) is the most common cause of dementia in the world, with more than 46 million people affected, and it is likely to increase up to 135.5 million by 2050 [2]. Its causes are still not clear, but it is characterized by deposits of high amounts of structures such as Amyloid- β (A β) plaques or neurofibrillary tangles accompanied by synaptic dysfunction and neurodegeneration that eventually lead to cell death [9, 52].

Diagnosis of AD is often based on the clinical history of the patient, using cognitive tests along with medical imaging and blood tests. A definite diagno-

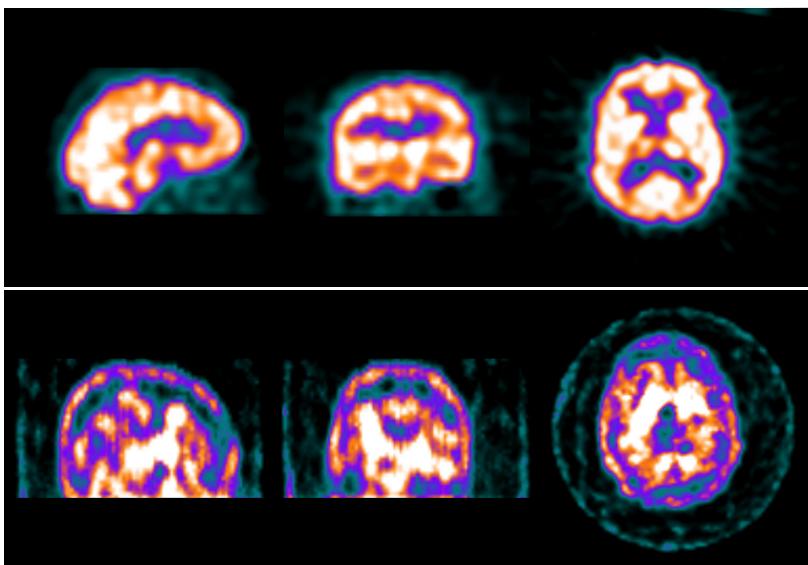


Figure 2.2: Example of SPECT images, a SPECT-HMPAO and a SPECT-DaTSCAN.

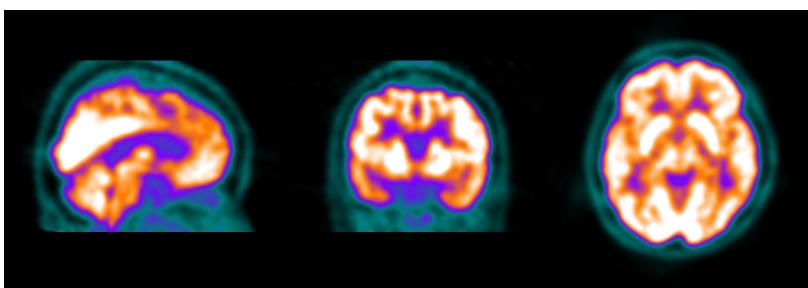


Figure 2.3: Example of a PET-FDG image.

sis can only be addressed post-mortem, via a direct examination of the brain tissue [52]. Cognitive tests such as *Mini Mental State Exam* or *Clinical Dementia Ratio* are widely used in clinical practice.

Initial symptoms of AD are often mistaken for normal ageing, in what is known as Mild Cognitive Impairment (MCI). Not all MCI-affected subjects develop AD. In fact, the prediction of MCI conversion is the most urging challenge in AD research, since an early diagnosis can lead to an improvement in life expectancy and quality of life of the patients.

MRI brain images have been extensively used in the diagnosis of AD by assessing neurodegeneration on GM and WM tissues. Research has shown in [37, 74, 96, 133] that neurodegeneration in Alzheimer's Disease mainly occurs in the GM tissue. Particularly grey matter loss has been described in the hippocampus and parahippocampal gyrus, according to the NINCDS-ADRDA criteria for AD diagnosis [96], with further atrophy described in the medial temporal structures, the posterior cingulate gyrus and adjacent precuneus [133]. Moreover, significantly lower volumes of certain regions in GM and WM have been considered a promising biomarker and predictor of the progression of AD in a longitudinal study involving MCI patients [74], and some structures in the striatum (putamen and caudate nucleus) have shown important volume abnormalities [37]. All these data suggest that many of the symptoms of AD can be observed in anatomical MRI images even in early stages of the disease, which could be of great help in its successful diagnosis and treatment.

Nuclear imaging, such as PET or SPECT, have also been used in clinical practice, especially to discard other diseases. In typical PET-FDG or SPECT-HMPAO, AD is characterized by reduced brain activity in bilateral regions, such as the posterior cingulate gyri and precunei, as well as the temporo-parietal region [160]. It also affects the frontal cortex and the whole brain in severe cases [140, 168].

Recently, new more specific radiopharmaceuticals have been developed, among them the Pittsburgh compound B (PiB). This drug binds to fibrillar A β allowing *in vivo* visualization of A β plaques in the brain [84]. However, due to their technical requirements -a relatively small half-life of the radioactive element-, they are unusual in clinical practice.

2.2.2 Parkinsonism (PKS)

Parkinsonism (PKS) or Parkinsonian Syndrome is the second most common neurodegenerative disease in the world, with a prevalence of 1-3% in the elder population (over 65 years)[97]. It is characterized by hypokinesia, rigidity,

tremor and postural instability [97]. It is not a single disease itself, but a wide range of conditions that share similar symptoms. The most common cause of PKS is Parkinson's Disease (PD), but other possible causes include toxins, a few metabolic diseases, and other extrapyramidal syndromes such as Multiple System Atrophy, Progressive Supranuclear Palsy or Cortico-Basal Degeneration [93, 115].

2.2.2.1 *Parkinson's Disease (PD)*

The etiology of Parkinson's Disease (PD) involves the progressive loss of Dopamine Transporters (DAT) of the nigrostriatal pathway. This causes a decrease in the dopamine content of the striatum, since the pathway connects the substantia nigra to the striatum.

In PD, structural imaging such as MRI has limited value, since structural abnormalities can be seen only in the latter stages. Nevertheless, in molecular imaging, a number of radiopharmaceuticals have been proposed to assess the levels of pre and post-synaptic DAT at the striatum. Among them, DaTSCAN (a.k.a. ioflupane) is perhaps the most popular. DaTSCAN binds to the DATs at the striatum [97, 101, 126], allowing the estimation of DAT density by means of a SPECT scanner. In DaTSCAN images, a reduced uptake at one or both striata is a clear indication of DAT loss, and therefore, of PD [101].

2.2.2.2 *Extrapyramidal Syndroms*

Among the extrapyramidal syndroms of PKS, the most relevant diagnoses are Multiple System Atrophy, Progressive Supranuclear Palsy and Cortico-Basal Degeneration [93]. An accurate diagnosis can positively impact in the health of these patients, avoiding wrong treatment decisions.

Structural imaging, as in the previous case, has little value here. To establish a differential diagnosis with PD, different drugs have been proposed. DaTSCAN is widely used to assess pre-synaptic dopaminergic loss, but in a post-synaptic level, one of the most extended is ^{18}F -DMFP-PET [8]. However, in this thesis we have focused only on the pre-synaptic level, and therefore, in the diagnosis of PD.

2.2.3 *Autism Spectrum Disorder (ASD)*

Autism Spectrum Disorder (ASD) is a neurodevelopmental syndrome characterized by social and communication impairment as well as restricted, repet-

itive patterns of behaviour, interests or activities. Its origins are still unknown, although research suggest [146] that there exist genetic risk factors.

The evidence of either functionally or structurally affected areas in the brain is a major concern [20, 21]. In the latter years, many strategies have been explored to recruit large samples in order to detect significant differences between affected and non-affected subjects. This has been addressed by initiatives such as the MRC-AIMS [32, 44] or the Autism Brain Imaging Data Exchange (ABIDE) [19].

2.3 Voxelwise Analyses

Traditional analysis of neuroimaging involves visual analysis by experts clinicians, or semi-quantitative analysis of Regions of Interest (ROIs). With the rise of neuroimaging in the mid-nineties, some computer-aided solutions appeared, of which the most extended are the widely known Statistical Parametric Mapping (SPM) [161] and its extension to structural imaging Voxel Based Morphometry (VBM) [137].

2.3.1 Statistical Parametric Mapping (SPM)

Statistical Parametric Mapping (SPM) is a new methodology to automatically examine differences in brain activity in functional imaging studies involving functional MRI (fMRI) or PET, firstly proposed by Friston in [161]. The technique can be applied either to static images (e.g., PET) or timeseries (fMRI), using inference techniques based on hypothesis testing, in order to construct the General Linear Model (GLM) that better describes the variability in the data.

Statistical hypothesis testing involves constructing a pair of hypotheses: H_0 , or the null hypothesis, that states no relationship between variables; and H_1 , the alternative hypothesis. In neuroimaging, H_0 usually means that there are no relevant differences between classes (for example, between patients affected by AD and CTls), and H_1 implies that there is a significant difference. Many different tests such as massive univariate t-Test or ANOVA (see chapter 4 for more information on these techniques) can be used in the SPM software [98], by using a design matrix that describes a t or F based contrast (for t-Test and ANOVA respectively). These terms are generally referred to as Z-values, namely the signed number of standard deviations an observation is above the mean.

The tests are computed voxel-wise, from which a p-value can be obtained, nominally the probability of obtaining equally or more extreme Z values than the one actually found. In many studies $p < 0.05$ is used for measuring statistical significance, which means that only a 5% of the times a experiment is repeated we would obtain that result or a more extreme one. The use of the significance threshold $\alpha = 0.05$ implies that any voxel with a p-value smaller than 0.05 is considered sufficient to reject the null hypothesis.

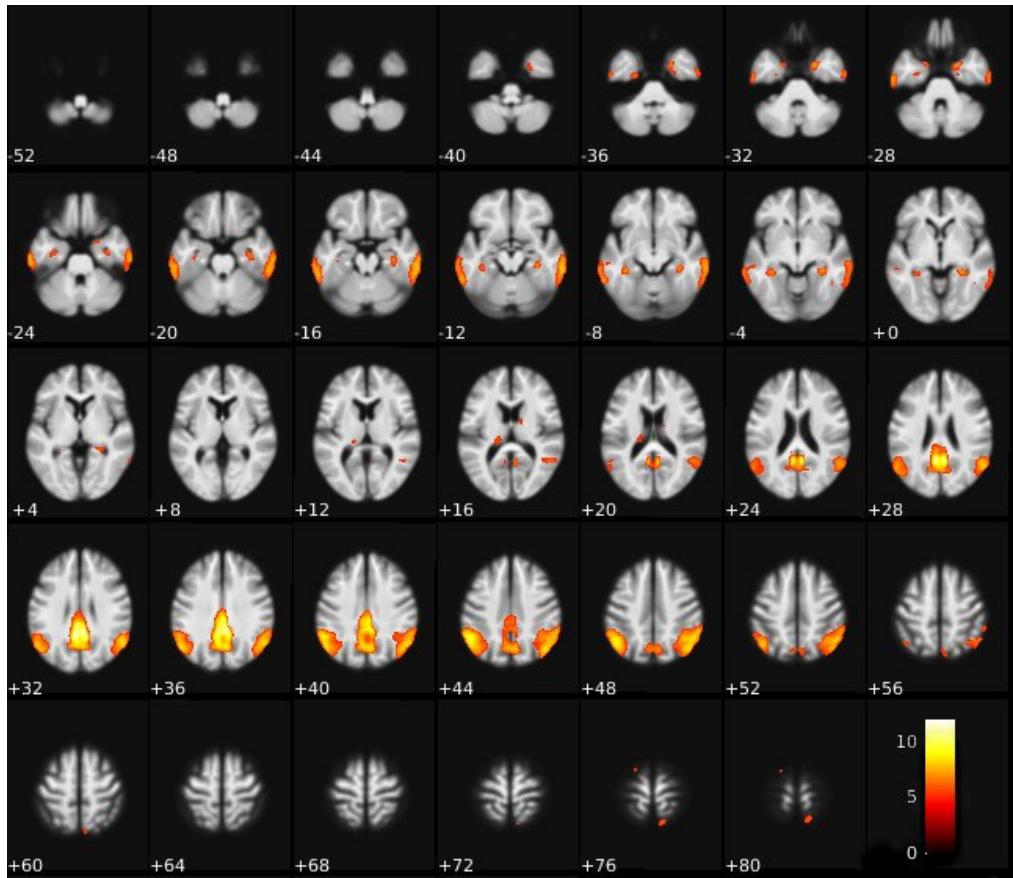


Figure 2.4: Example of a SPM analysis on a PET dataset displaying the differences between AD and CTL, using $p < 0.05$ and FWE correction.

SPM outputs maps like the one shown in Figure 2.4. There, significant Z-values according to a given threshold (FWE uncorrected or corrected, see section 2.3.3) are displayed over an anatomical reference. The resulting maps allow a visual inspection of the significant brain areas, which can later be related to a certain disease or task.

Although SPM's main feature is the estimation of differences, the term has been extended to cover the whole process performed by the SPM software. That is, it generally involves registration to a template, intensity normalization, smoothing, the proper SPM difference estimation and the display of the results. An overview of these procedures is provided at chapter 3.

2.3.2 Voxel Based Morphometry (VBM)

Voxel Based Morphometry (VBM) can be considered an extension of SPM applied to structural MRI images [137]. The procedure involves preprocessing (see chapter 3), where smoothing is applied to account for smaller anatomical differences. Afterwards, a GLM is applied to each voxel in the images, and a Z-score map similar to Figure 2.4 is produced.

Smoothing is more important in VBM than in regular SPM, since MRI images have higher resolution and are less noisy than functional images. Larger smoothing kernels will miss out smaller regions, while smaller kernel can lead to artifacts in the generated Z-maps, including misalignment of brain structures, differences in folding patterns or misclassification of tissue types [4]. Therefore, the kernel size must be carefully chosen, usually using a-priori knowledge about the regions affected, and always double checking for artifacts and reproducibility.

The idea behind VBM has been extended in a number of papers, using multivariate approaches that takes into account all voxels at once, and not their individual differences. Some of them include ICA decomposition of the dataset and a posterior conversion to Z-scores in what was called Source Based Morphometry [81], or multidimensional Tensor Based Morphometry [63].

2.3.3 The Multiple Comparisons Problem

The Multiple Comparisons problem arises when using hypothesis testing to assess statistical significance. This is widely used in neuroimaging, where statistical tests such as the t-Test or ANOVA are used to quantify voxel-wise differences, and state their statistical significance, or p-value. The p-value, as described above, is the probability of any value being more extreme than a certain threshold under a given hypothesis. In our problem, given the t-value T_i for the i^{th} voxel ($i = 1, \dots, N$) of the images, and a threshold T_{th} under the hypothesis H_0 , the significance can be assessed by checking:

$$P(T_i > T_{\text{th}} | H_0) < \alpha \quad (2.2)$$

where α is the significance level.

Choosing α is not trivial in neuroimaging. The use of the significance level $\alpha = 0.05$ implies that any voxel with a p-value smaller than 0.05 is considered sufficient to reject the null hypothesis. This does not directly imply the necessity of accepting the alternative hypothesis H_1 , although it is often thought so. Neither it yields the probability of the null hypothesis [123].

If we apply $p < 0.05$ directly to a medical image of, for example, 300,000 voxels, that could mean the possibility of almost 15,000 voxels being false positives. Controlling the apparition of false positives when applying a massive univariate test is not trivial. It implies a balance between the true positive rate (sensitivity) or true negative rate (specificity), given that, for example, controlling the amount of false negatives will result in many false positives and vice-versa.

Usually, two options for controlling the amount of false positives are given: the Family Wise Error rate (FWE) and the False Discovery Rate (FDR). The FWE is the probability of obtaining at least one type I error. Mathematically, the null hypothesis for the i^{th} voxel H_{0i} states that there is no activation in that voxel. Therefore, the family-wise null hypothesis for our problem is:

$$H_0 = \bigcap_i H_{0i} \quad (2.3)$$

If we reject a single null hypothesis ($T_i > T_{th}$), we reject H_0 . Therefore, we want to control the probability of a single voxel being significant if the family-wise null hypothesis is valid:

$$P \left(\bigcup_i [T_i > T_{th}] | H_0 \right) < \alpha \quad (2.4)$$

In this case, we must obtain the critical value T_{th} , which is the higher t value that matches that expression. Many options have been proposed to this problem, among them the conservative Bonferroni correction, methods that use random field theory or permutation tests.

2.3.3.1 The Bonferroni Correction

The Bonferroni correction [158] rewrites eq. 2.2 setting $\alpha = \frac{\alpha}{N}$ so that:

$$P(T_i > T_{th} | H_0) < \frac{\alpha}{N} \quad (2.5)$$

That way, using the Boole's inequality:

$$FWE \leq \sum_i^N \frac{\alpha}{N} = \alpha \quad (2.6)$$

Therefore, we can comply with the imposed restriction for a maximum FWE, or in our case, a maximum rate α of false positives. This is considered a rather conservative approach. In the example cited above, if we want to keep the FWE below 0.05, we should divide it by N , therefore obtaining a T_{th} that makes $\alpha = 0.05/N = 1.67 \times 10^{-7}$.

Other less conservative options try to compute a critical value T_{th} that minimizes the FWE using spatial information. This is the case of using an approximation of the distribution of the maximum statistic over the image, or the spatial correlation, including elements from random field theory (the approach used in SPM [98]).

2.3.3.2 Random Field Theory

In the random field approach, the maps of the statistic are treated, under the null hypothesis, as a lattice representation of smooth isotropic three dimensional random fields of test statistics. This approximation to the problem allow us to approximate the upper tail of the maximum distribution, the part needed for defining an event that occurs when the map exceeds the critical value T_{th} . Further information about random field theory and how it is applied to the SPM software can be found at [98].

2.3.3.3 FDR Controlling Procedures

The other approach, based on the FDR, aims at controlling the proportion of false positives in the total number of voxels declared significant. The most extended procedure for controlling the FDR is that proposed by Benjamini and Hochberg [156]. The Benjamini and Hochberg method start with calculating the p-values of all voxels and ranking them so that:

$$p_1 \leq p_2 \leq \dots \leq p_i \leq \dots \leq p_N \quad \forall i = 1 \dots N \quad (2.7)$$

Let q be the a maximum FDR value that we can afford, for example 0.05. For each i , we compute:

$$p_i \leq \frac{i}{N} q \quad (2.8)$$

The maximum i value that holds Eq. 2.8 is used as α , the significance level, and its corresponding statistical value (T_i in the case of a t-test) is used as the critical value. This test, under the family-wise null hypothesis H_0 , is equivalent to controlling the FWE. However, FDR methods are less conservative than other approaches such as the Bonferroni or other FWE-based corrections, leading to a gain in statistical power.

2.3.3.4 *Permutation Tests*

An empirical way to obtain p-values without relying on any parametric assumption is permutation testing [29, 132]. Permutation tests evaluate a statistic such as the F-statistic or the t-test using randomly target variables, in our case, the classes. The procedure is applied many times (up to 10,000), and for each permutation, only the maximum value of the computed statistic is considered. These values are used to build the null distribution, from which the family-wise corrected p-values are computed. Results obtained in permutation tests are comparable to those obtained using Random Field Theory [29], and far less conservative than when applying the Bonferroni correction.

2.4 Machine Learning in Neuroimaging

Machine learning is a current trend in neuroimaging. It provides computers with the ability to learn from data, using a set of statistical and computational tools. Rather than being explicitly programmed for a certain task, machine learning systems are able to find relevant data, discover patterns and predict the outcome of the input data. Its application to medicine is often known as Computer Aided Diagnosis (CAD) [7].

There are two major branches of machine learning: supervised and unsupervised learning. The former explores the patterns that lead to a certain outcome, whereas on the other hand, unsupervised learning explores the underlying structure of the data. Most of the CAD systems rely on supervised learning, since their intention is to discover patterns that can effectively predict a disease.

For simplicity, in this thesis, when talking about CAD, we will always refer to automatic CAD systems. That is, those that, once trained with previously known data, can predict the outcome of new, unseen data. A typical CAD system, like the one in Figure 2.5, consists of input data (in our case, neuroimaging), feature extraction, feature selection and a classification step. The most basic is the Voxels As Features (VAF) approach, in which all voxels are considered as features, and then used as input to the classifier [120]. However, many more advances can be made in this field by adding different combinations of feature selection and feature extraction algorithms.

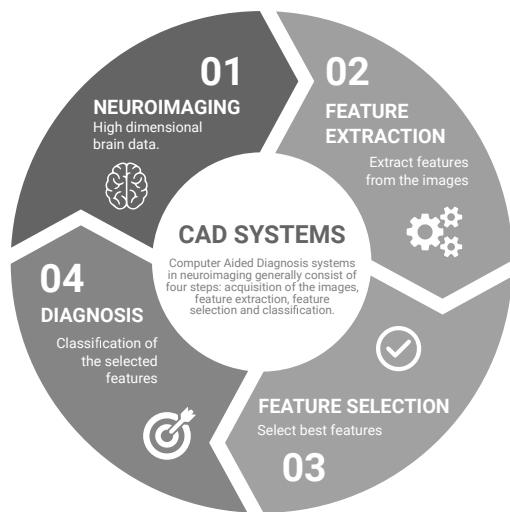


Figure 2.5: Illustration of a typical neuroimaging CAD system.

2.4.1 Voxels As Features (VAF)

Voxels As Features (VAF) [120] is an example of the simplest CAD system. It was originally proposed for evaluating and performing automatic diagnosis of AD using functional SPECT imaging. It uses a standard preprocessing (registration, intensity normalization) and a Support Vector Classifier (SVC) (See Appendix B) to predict the class of an image using all its intensities as features. Feature extraction, here, considers all voxels, and then there is no feature selection applied.

It has been used in many works as a baseline [7, 77, 78], since it is comparable to the performance achieved by expert physicians using visual analysis [120]. The weight vector of the SVC can be inverse transformed to the dimension of the original images, and therefore provide a visual map that reflects the most influential voxels, in a similar way to the Z-maps of SPM and VBM.

2.4.2 Multivariate Analyses

Many improvements can be made to VAF by adding and refining feature extraction and feature selection techniques. With this addition, we can avoid the Small Sample Size, in addition to the ability to discover higher level abstractions that can be more representative of the progression of the studied diseases.

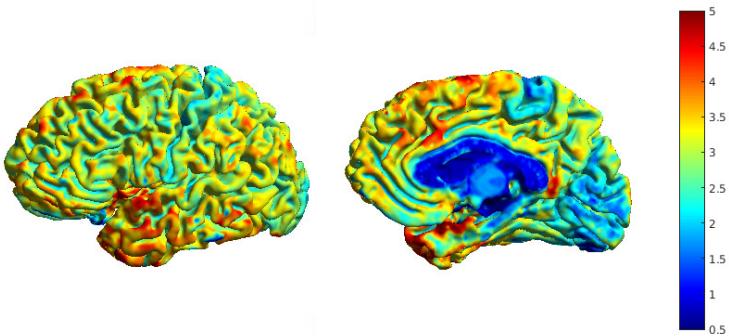


Figure 2.6: Example of the cortical thickness of a subject, obtained with the toolbox CAT12 in SPM12.

Feature extraction algorithms often change the strategy from a massive univariate approach, where a single feature is considered at each time, to multivariate analyses, where each feature can contain information from many voxels at the same time. Measures of total uptake of a given drug in nuclear imaging are a good example of this [100, 104], but also the widespread Cortical Thickness [142] provided by *FreeSurfer* or the CAT12 toolbox of SPM-12 (see Figure 2.6). Cortical thickness is an estimation of the amount of GM in a direction perpendicular to its surface. It first estimates the GM-WM and the WM-CSF separation surfaces, and then characterizes the thickness of the tissue, allowing a characterization of GM differences such as atrophy or hypertrophy.

Other more advanced algorithms are image decomposition techniques such as PCA or PLS, which have been extensively used in neuroimaging CAD systems [13, 40, 56, 60, 78]. In these approaches, a given image can be represented as the linear combination of different components, and while the component loadings are common to all subjects, the weights of these components are unique to each patient. This allows us to identify the patterns that better discriminate between classes, leading to a more accurate diagnosis.

For its part, feature selection refers to different strategies aimed at finding an optimal subset of the extracted features, according to a certain criterion. Irrelevant features are therefore discarded, making our models faster and more cost-effective [124]. Feature selection algorithms are often subdivided in three approaches [5]: filters, wrappers and embedded approaches.

Filters compute a feature relevance score from the data, which is then used to sort the different features. It is computed before the classification, and does not interact with it. Many scores can be derived from statistical features such as χ^2 , t-Test, Fisher's Discriminant Ratio (FDR) or others [5, 35]. The output

of these tests is already a very extended tool in voxelwise analyses, as we commented in section 2.3.

Wrappers are similar to filtering methods, since they assign a certain score to each feature. But in contrast to filters, the score is computed by estimating the performance in a predictive model, such as classifiers [157]. The most obvious measure here is accuracy, although other techniques such as Forward selection, backward elimination [124], genetic algorithms [157], or the expectation-maximization algorithm [70] have been used in the literature. And finally, embedded approaches use the very model that is being built to construct their optimal feature subset.

3 | GENERAL METHODOLOGY

Throughout this thesis, we will propose many CAD systems and analyses. We will apply them to many experiments, using similar data and techniques in them. In this chapter, we will focus on the methodology that is common to most of these experiments, particularly the preprocessing and evaluation of our systems.

To perform most automated analyses on neuroimaging, it is fundamental that images are comparable. Preprocessing comprises a series of algorithms that, applied after the acquisition and reconstruction of the images, produce directly comparable images in both structure and magnitude. Whether they have been used in one or all experiments, we can classify them in two major categories: spatial and intensity preprocessing. These are addressed in section 3.1 and 3.2 respectively.

Afterwards, in section 3.3, we will discuss how we evaluate our systems. Here we propose some performance measures and the procedure to obtain them by training and testing our systems.

3.1 Spatial Preprocessing

Spatial processing usually accounts for the differences in position, angles and structure that are generally found between images. A common pipeline in, for example, MRI preprocessing, is the one found at Figure 3.1, where the images are registered (or spatially normalized) to a template, smoothed and finally segmented. The smoothing is an optional step, often used in procedures like segmentation or VBM.

In this thesis, all the experiments in all image modalities involve spatial normalization. Smoothing, as well as segmentation, is only applied in some experiments that use MRI images, such as the segmented images in chapter 6 or the whole-brain analysis performed in chapter 7.

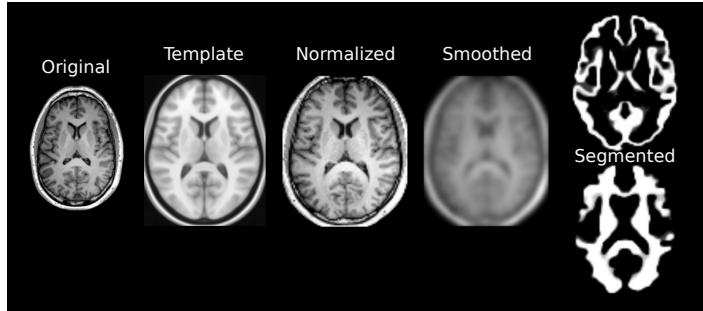


Figure 3.1: Typical pre-processing pipeline in MRI.

3.1.1 Spatial Normalization or Registration

Spatial Normalization, also known as registration, is the procedure by which every subject's brain is mapped from their individual space to a standard reference system. Registered images allows our system to overcome the individual differences in position and anatomy by establishing a common reference space in which a given coordinate represent the same anatomical position in all brains in the dataset.

There exist a number of pieces of software widely used for registering images, such as FreeSurfer [68] or FSL (in the FLIRT and FNIRT package) [119]. They them perform linear, non-rigid and elastic transformations or a combination of these. In this work we have used the software Statistical Parametric Mapping Software, version 8 (SPM8) [98] to perform registration of all the MRI, SPECT and PET datasets, and therefore we will focus on how it performs these procedures.

Linear registration usually refers to the affine transformation, a matrix multiplication that includes 12 parameters for translation, rotation, scale, squeeze, shear and others:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (3.1)$$

This matrix multiplication is performed globally, as it transforms the whole image, not accounting for local geometric differences. In equations 3.2, 3.3 and 3.4 we give an example of the parameters that are computed for scale, translation and shear in 3D:

$$\text{scale} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

$$\text{translation} = \begin{bmatrix} 1 & 0 & 0 & \Delta x \\ 0 & 1 & 0 & \Delta y \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.3)$$

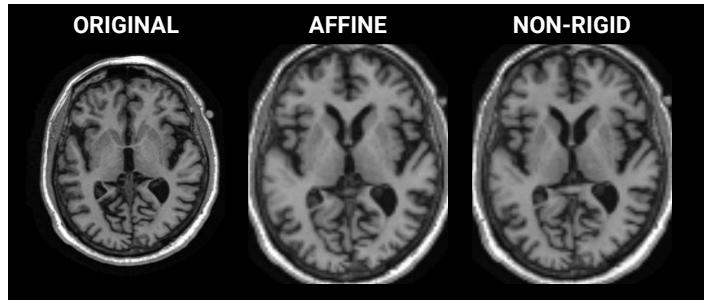
$$\text{shear} = \begin{bmatrix} 1 & h_{xy} & h_{xz} & 0 \\ h_{yx} & 1 & h_{yz} & 0 \\ h_{zx} & h_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.4)$$

The combination of all these operations result in the estimation of the twelve parameters in Eq. 3.1, which are the ones used in SPM8. The estimation of these parameters is performed via the optimization of a cost function, that in SPM8 can be the minimum squared difference between the source image and the template [98] in the case of within-modality registration, or the mutual information in between-modality registration. These functions are also used in FLIRT [134], whereas FreeSurfer uses the Tukey's biweight function (in `mri_robust_template`) [48].

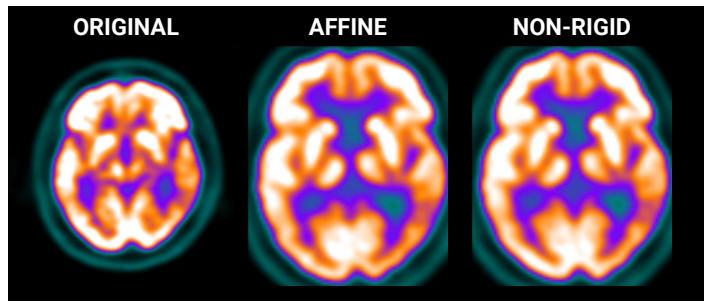
After the affine transform, the software usually performs a fine-tuning step via nonrigid transformations, to account for relevant anatomical differences between subjects. Nonrigid transformations range from the use of radial basis functions, physical continuum models and the large deformation models, or diffeomorphisms, that SPM8 uses. These procedures work by estimating a warp-field and then, apply it to the affine-registered images. An example of the differences of using only affine registration and applying diffeomorphisms can be found at Figure 3.2.

3.1.1.1 Co-registration

Sometimes we have several image modalities of the same subject, for example MRI and PET or functional MRI, often acquired at the same time. In this particular case, we can use the higher resolution MRI image to calculate the affine parameters and warping, and apply those to all modalities of the same subject. To do so, we perform a first co-registration, that is, a registration of the lower-resolution images (e.g. PET) to its correspondent MRI image. Being



(a) Comparison in MRI.



(b) Comparison in PET.

Figure 3.2: Comparison of the affine registration and the application of non-linear transformations to both MRI and PET images of the same ADNI subject.

anatomically similar, the co-registration usually comprises a single affine transformation. Afterwards, we can proceed with the registration of that MRI image to the template, and apply the same transformation to all its co-registered images.

3.1.1.2 *The MNI Space*

In this thesis, all images are coregistered to the Montreal Neurological Institute (MNI) space [136]. This is the most widely used coordinate system, recently adopted by the International Consortium for Brain Mapping (ICBM) as its standard template. The three-dimensional coordinate system defined in MNI was intended to replace the Tialarach space, an older system based on a dissected brain, that was used to compose an atlas by Tialarach and Tournoux [166]. The current template is known as ICBM152, and features the average of 152 normal MRI scans matched to a previous version of the MNI template using a nine parameter affine registration.

3.1.2 Segmentation

When using MRI images in this thesis, we often refer to grey matter (GM) and white matter (WM) maps, which is the result of the segmentation of the original data. Segmentation aims at producing maps of the distribution of different tissues, and it generally addresses GM, WM and CSF classes, although some software can also output data for bone, soft tissue or very detailed functional regions and subregions [128].

In this thesis we have used the VBM toolbox of the SPM8 software, which yields GM, WM and CSF maps. It features an Expectation-Maximization (EM) algorithm to model the distribution of the tissue classes as a mixture of gaussians and, by combining this distribution-based information with tissue probability maps using a bayesian rule, the software produces joint posterior probability maps for each tissue. To clean up the segmentation maps, a series of iterative dilations and erosions are used. Finally, since brain regions are expanded or contracted at the spatial normalization step, we can scale the segmented maps using modulation, producing final maps where the total amount of grey matter is preserved.

3.2 Intensity Normalization

Generally, structural modalities such as T1 and T2-weighted images are considered unitless, in contrast to functional imaging, in which each voxel's intensity represent the distribution of some biomarker, such as glucose metabolism, dopamine transporters, etc. These amounts are affected by many sources of variability that can affect the final values: individual contrast uptake, radio-tracer decay time, metabolism, etc. Therefore, along with the previous spatial normalization, there is a need to normalize the intensities of the images, so that the amount they represent is comparable between subjects.

In the case of intensity normalization, the method acts as a linear transformation of the image intensities, preserving fundamental information such as contrast between regions. This approximation estimates the new intensity values I' as:

$$I' = I/I_p \quad (3.5)$$

where I_p is a constant parameter, unique for each image. After this division, the new intensities should be directly comparable. The technique used to compute the normalization parameter varies, ranging from the simplest normalization to the maximum [47, 77] to complex methodologies that use assumptions about the image's Probability Density Function (PDF).

The *normalization to the maximum* strategy computes I_p as the average value of the 95th bin of the histogram of the image. In other words, this mean averaging the 5% higher intensity values and use this mean as I_p . Another useful approach is the so-called *integral normalization*, which computes I_p as the sum of all values in the image.

Other approaches involves some a-priori knowledge about the intensity distribution of normal subjects in a certain modality. This is the case of setting I_p to the Binding Potential (BP), a ratio between the intensities at specific and non-specific areas [113].

Finally, more advanced approaches use a general linear transformation of the image:

$$I' = aI + b \quad (3.6)$$

The parameters a and b are so that the PDF of a given matches a reference PDF. There exist methods that use the histogram [154], the gaussian distribution or the alpha-stable distribution [39]. In this latter case, the parameters a y b are computed as linear transformations of some distribution's parameters:

$$a = \frac{\gamma^*}{\gamma}, \quad b = \mu^* + \frac{\gamma^*}{\gamma} \mu \quad (3.7)$$

where γ^* and γ are the dispersion parameters of the alpha-stable intensity distribution of the non-normalized and the reference image respectively, and μ^* and μ are the location parameters of the same images.

Despite traditionally structural modalities such as MRI did not use intensity normalization, there exist a new tendency towards the use of quantitative T1 - weighted (qT_1) and quantitative T2 - weighted (qT_2) images [43] that provide biomarkers for absolute measures such as myelination, water and iron levels. This strategy is especially designed to overcome different sources of variability that affect multicentre studies, e.g. magnetic field inhomogeneity, noise, evolution of the scanners, etc. The role of those in multi-centre studies is addressed at chapter 7.

See Figure 3.3 for a comparison between different strategies of intensity normalization on the same DaTSCAN images.

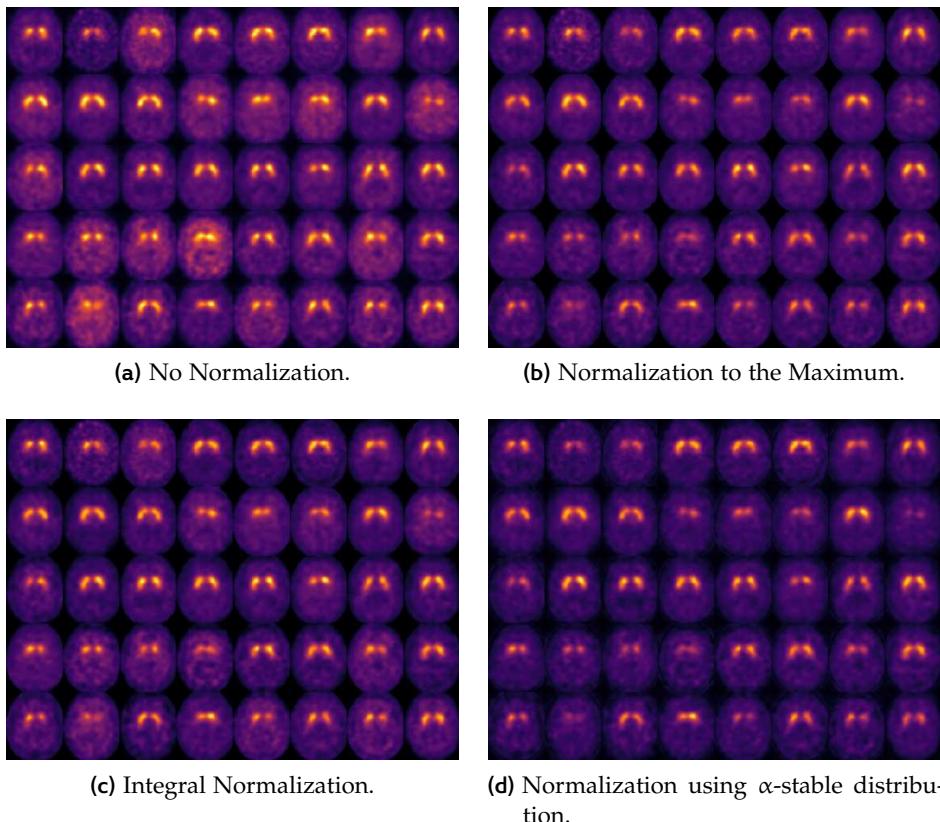


Figure 3.3: Comparison between different types of intensity normalization, applied to the VDLN-DAT dataset (see Appendix A).

3.3 Evaluation Parameters and Methodology

3.3.1 Cross-validation

Some machine learning applications such as digit or faces recognition use tens of thousands of images as input. In these cases, the common practice is to divide randomly the data in three subsets: training, validation and testing [151]. However, in neuroimaging, sample size is an issue. In contrast to those applications, we only have hundreds of patients in the best case, and the estimation of the performance using these subsets might not be reliable.

In these cases, Cross-validation (CV) is used to obtain more accurate performance measures. CV performs a division of the dataset into several subsets $X = S_1, S_2, \dots, S_k$ and iteratively use some of these subsets for training or test-

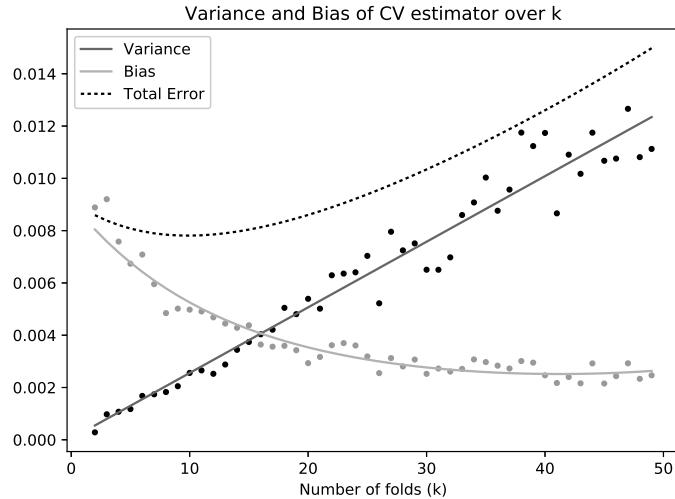


Figure 3.4: Evolution of bias and variance when increasing the number of folds in a k -fold CV.

ing. The simplest CV estimator is k -fold. This approach uses k equally-sized, non-overlapping subset. For each subset S_i (or “fold”, hence the name), the model is trained on all subsets $S_k \forall k \neq i$, and then evaluated on S_i . The performance measures, e.g. accuracy, are obtained as the average of the accuracies on each fold.

A particular case where $k = N$ (where N is the exact number of subjects in the dataset) is Leave-One-Out (LOO). This estimator is approximately unbiased for the true accuracy, but can have high variance because there is much overlapping between the N training set [71]. This imply that the learned models are correlated, and therefore, dependent. All CV strategies with $k > 2$ have overlap, and therefore, high variance. See how variance and bias evolve in a k -fold validation in Figure 3.4.

A $k = 10$ is assumed as a good compromise between variance and bias in many works [71, 157]. In this thesis, when referring to k -fold, we often use 10-fold stratified cross validation, which is a subclass of k -fold with 10 divisions where the distribution of classes within each fold is similar to the distribution of classes in the whole dataset, making the estimates more accurate [157].

3.3.2 Classification Performance

From each iteration in the CV loop, a confusion matrix is obtained, from which all performance measures will be obtained.

		Predicted	
		Positive	Negative
Actual	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Table 3.1: Confusion matrix and its parts

The confusion matrix (see Table 3.1) accounts for the number of correct and incorrect predictions: True Positives (TPs) and True Negatives (TNs) are correct predictions, and False Positives (FPs) and False Negatives (FNs) are incorrect predictions. It also allow us to identify which type of error is our model making, which in hypothesis testing are known as type I errors (FPs) and type II errors (FNs). The confusion matrix is the basis for computing other performance measures, such as accuracy (acc), sensitivity (sens) or specificity (spec).

$$\text{acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3.8)$$

$$\text{sens} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3.9)$$

$$\text{spec} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (3.10)$$

Sensitivity is also known as TP rate or recall in the literature, and specificity is known as TN rate. Sensitivity is widely used in the medical literature, since it gives an idea of how “sensitive” is our model to the patterns related to a disease (usually considered a positive detection or class).

Part II

REDUCING THE FEATURE SPACE

4

IMAGE DECOMPOSITION

In this chapter, we will focus on those CAD systems that use a combination of an image decomposition method and feature selection by means of hypothesis testing. These variety of methods have been published in [24, 35, 47, 59].

Image decomposition methods model a set of samples as a linear combination of c latent variables, also known as components. These variables can be considered as the basis of a c -dimensional space where each sample is represented by a feature or coordinate vector of length c . The i -th neuroimage in our dataset can be therefore decomposed as:

$$\mathbf{x}_i = \mathbf{s}_0 \mathbf{w}_0 + \mathbf{s}_1 \mathbf{w}_1 + \cdots + \mathbf{s}_c \mathbf{w}_c + \epsilon = \mathbf{sW} + \epsilon \quad (4.1)$$

Where s_i is the coordinate (or component score) of the current image in the i -th dimension of the new space defined by all the base vectors \mathbf{w}_i (component loadings), and ϵ is the error of the estimation. Figure 4.1 shows an illustration of the process.

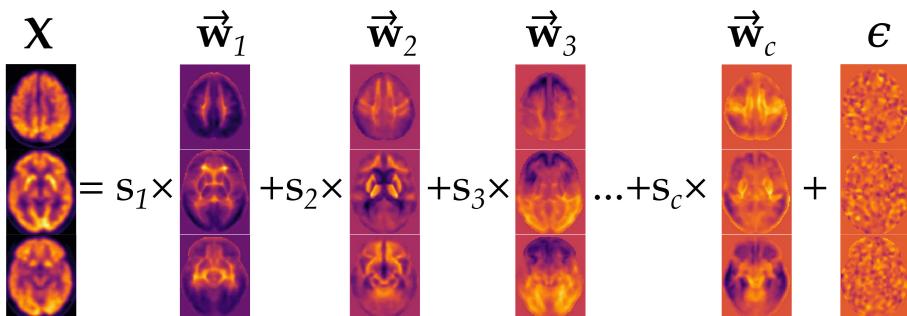


Figure 4.1: Illustration of how decomposition algorithms such as FA and ICA work on a PET-FDG brain image.

Many signal decomposition techniques are used in the literature, for example PCA or PLS [13, 40, 56, 60, 78]. We will focus on two less known decomposition algorithms Factor Analysis (FA) and Independent Component Analysis (ICA), which we will integrate in different CAD systems using a pipeline similar to the one displayed at Figure 4.2. This pipeline involves feature selection (for reducing the dimensionality), decomposition of the feature vectors and classification.

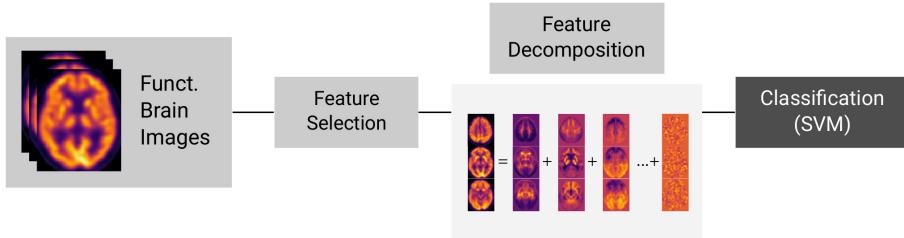


Figure 4.2: Illustration of the system used in chapter 4.

4.1 Feature Selection

Feature selection is our first strategy to reduce the feature space dimension [5], and it is often used along with feature extraction in order to build more complex pattern recognition systems. It refers to any strategy intended to find a subset of the original features containing the more suitable ones according to a certain criterion. Therefore, irrelevant features are discarded, and resultant models are faster and more cost-effective [124]. However, it usually requires an additional optimization to find the parameters for the optimal feature subset, and furthermore, it is impossible to guarantee that the optimal features for the subset are the same of the full feature set [122].

In this work, we will use filtering methods to perform feature selection. As we introduced in section 2.4.2, filtering methods are based on the computation of a feature relevance score directly on the data. The relevance score is used to sort the different features, discarding those with a lower score, and it is usually computed independently for each feature, in what is called a univariate approach [102].

Feature selection can be used before or after feature extraction. When using computationally-intensive algorithms such as FA or especially ICA, the selection of best features prior to the decomposition is key to obtain high performance while keeping the computation times small [47, 59]. This also removes noise in some cases where the decomposition algorithm cannot correctly compute the variance.

Three feature selection algorithms have been used in this thesis, not only in the CAD systems proposed in this chapter, but also in many other models that will be presented later: the t-Test, the Kullback-Leibler divergence or Relative Entropy, and the Mann-Whitney-Wilcoxon rank test.

4.1.1 t-test

The t-test is an old friend of statisticians. In this work we will use the independent two-sample t-test [66]. It quantifies the differences between two classes assuming independent variances. Let X_i^f a vector containing the f-th feature of all elements in class i. The t-score of the f-th feature can be computed as:

$$t_f = \frac{\bar{X}_1^f - \bar{X}_2^f}{\sqrt{\frac{\sigma_{X_2^f}^2 + \sigma_{X_1^f}^2}{n}}} \quad (4.2)$$

where $\sigma_{X_i^f}^2$ is the variance and \bar{X}_i^f is the average of the f-the feature within class i. The t-test is extensively used in the neuroimaging community, and it is the basis for the SPM and VBM analyses [98]. See figure 4.13a for an example of the t-test computed on the ADNI-PET database.

4.1.2 Relative Entropy

Another alternative is the relative entropy, also known as Kullback-Leibler (KL) divergence. It is a non-symmetric measure of the difference between two probabilities distributions. Let us assume that X_1^f and X_2^f , the vectors containing the f-th feature of all elements in class i, are two discrete random variables. Therefore, the relative entropy can be calculated with equation 4.3 [147].

$$RE_f = \left(\frac{\sigma_{X_2^f}^2}{\sigma_{X_1^f}^2} + \frac{\sigma_{X_1^f}^2}{\sigma_{X_2^f}^2} - 2 \right) + \frac{1}{2} (\bar{X}_2^f - \bar{X}_1^f)^2 \left(\frac{1}{\sigma_{X_1^f}^2} + \frac{1}{\sigma_{X_2^f}^2} \right) \quad (4.3)$$

using the same notation than in t-test. See figure 4.13b for an example of the computed relative entropy on the ADNI-PET database.

4.1.3 Mann-Whitney-Wilcoxon (MWW) U-test

The Mann-Whitney-Wilcoxon (MWW) rank test, also known as U-test, assigns a rank to all values in the vector corresponding to the f-th feature, X^f , without considering any class. The method used to assign a rank is the ‘average’, which means that each value is assigned with the average of the ranks that would have been assigned to all the tied values. This means that, for

example, in the case of the vector $X^f = (0, 2, 3, 2)$, the ranks assigned to each element would be $R^f = (1, 2.5, 4, 2.5)$.

Let n_1 and n_2 be the number of elements in class 1 and 2 respectively, and R^f the vector of ranked elements. We proceed by selecting the first n_1 elements in R^f by:

$$R_{n_1}^f = R_i^f \quad \forall i \in (0, n_1) \quad (4.4)$$

The U -score for the f -th feature and the first class will be:

$$U_1^f = n_1 n_2 + n_1 \frac{n_1 + 1}{2} - \sum R_{n_1}^f \quad (4.5)$$

And it can be computed for the second class as the remainder:

$$U_2^f = n_1 n_2 - U_1^f \quad (4.6)$$

The final U^f can be assigned to either U_1^f , U_2^f or $\min U_1^f, U_2^f$ [66], but the usual approach nowadays is to assign $U^f = U_2^f$. Unlike t-test, MWW test does not assume any distribution, and therefore is less likely to spuriously indicate significance because of the presence of outliers. Under the normal distribution, it performs relatively similar [66]. See figure 4.13c for an example of the MWW U -test computed on the ADNI-PET database.

4.2 Decomposition Algorithms

The feature selection algorithms presented above will perform a significant feature reduction, from hundreds of thousands of voxels to a few thousands. These few thousands voxels are considered the best in discriminating between CTL and affected subjects in each of the diseases. The feature selection strategy can be thought of as a mask, in which only the most relevant regions according to the tests are selected (see Figure 4.13).

However, this number of features is still large, and therefore, further feature reduction can be applied by performing a decomposition of the masked regions. We have used two algorithms in our CAD systems: Factor Analysis (FA) and Independent Component Analysis (ICA).

4.2.1 Factor Analysis (FA)

Factor Analysis (FA) was used in [47, 59] to perform feature extraction in CAD systems. This strategy assumes that each image in the database is a realization of a given experiment. FA then models each of the N observations

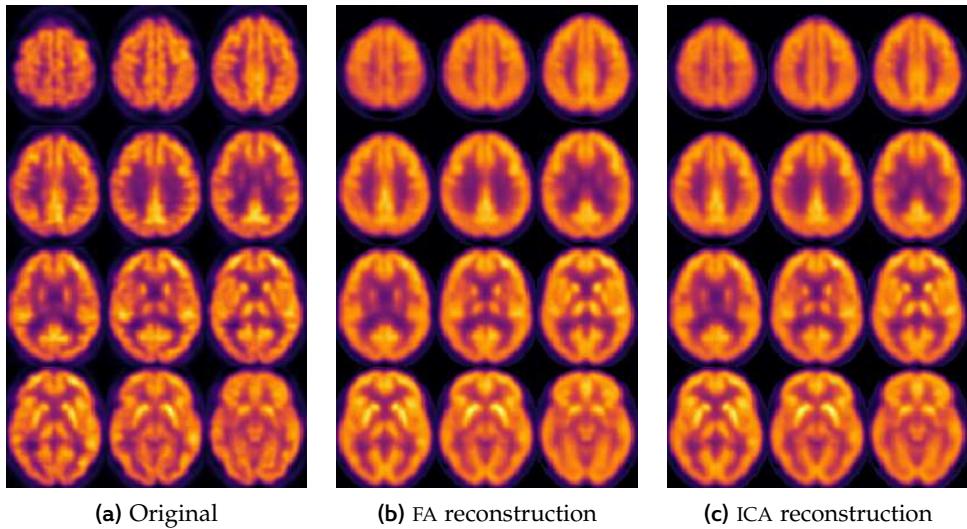


Figure 4.3: Original PET image from the ADNI-PET dataset, and examples of reconstruction using FA or ICA, with 10 components.

(or subjects) as the expression of c unobserved variables, known as factors. The model follows the general decomposition equation (Eq. 4.1), but assuming that the dataset matrix \mathbf{X} is zero-centred. That is, that we have subtracted the mean prior to the computation. In matrix form, Eq. 4.1 can be rewritten as:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{S}\mathbf{W} + \boldsymbol{\epsilon} \quad (4.7)$$

The columns of \mathbf{W} are known as factors, and the rows of \mathbf{S} are known as loadings (similar to the concept of component loading and component scores in PCA). Thanks to this, we can convert the original dataset \mathbf{X} of size $N \times f$ into \mathbf{S} , of size $N \times c$. The procedure of computing the decomposition imposes some assumptions on \mathbf{W} :

- \mathbf{W} and $\boldsymbol{\epsilon}$ must be independent.
- $E[\mathbf{W}] = 0$.
- $Cov(\mathbf{W}) = \mathbf{I}$, which ensures that the factors are uncorrelated.

Now we can rewrite Eq. 4.7 as:

$$Cov(\mathbf{X} - \boldsymbol{\mu}) = Cov(\mathbf{S}\mathbf{W} + \boldsymbol{\epsilon}) \quad (4.8)$$

Under the previous constraints, and setting $\boldsymbol{\Sigma} = Cov(\mathbf{X} - \boldsymbol{\mu})$, Eq. 4.8 becomes:

$$\boldsymbol{\Sigma} = \mathbf{S}Cov(\mathbf{W})\mathbf{S}^T - Cov(\boldsymbol{\epsilon}) \quad (4.9)$$

Since $\text{Cov}(\mathbf{W}) = \mathbf{I}$, and making $\text{Cov}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$, the diagonal matrix containing the specific variances of the reconstruction error, we obtain the alternative form of FA:

$$\boldsymbol{\Sigma} = \mathbf{S}\mathbf{S}^T - \boldsymbol{\Psi} \quad (4.10)$$

The mean μ , and the matrices \mathbf{S} and $\boldsymbol{\Psi}$ are obtained via Maximum Likelihood estimation. To guarantee an unique solution, we impose that $\mathbf{S}^T\boldsymbol{\Psi}^{-1}\mathbf{S}$ is a diagonal matrix. Then, we obtain the parameters by maximizing the log-likelihood given by the following expression:

$$\ell(\mu, \mathbf{S}, \boldsymbol{\Psi}) = -\frac{np}{2} \log 2\pi - \frac{n}{2} \log |\mathbf{S}\mathbf{S}^T + \boldsymbol{\Psi}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \mu)^T (\mathbf{S}\mathbf{S}^T + \boldsymbol{\Psi})(\mathbf{x}_i - \mu) \quad (4.11)$$

FA differs from PCA mainly because it performs an estimation of the noise, and needs the number of factors c as an input. Choosing c is not a naive task. A large c can yield a small reconstruction error, but the factors will not be representative enough, leading to overfitting of the subsequent model. Conversely, a small c can lead to a large reconstruction error, causing information loss. We have computed the reconstruction error variance over the ADNI-PET dataset, and plotted it in Figure 4.4 (similar graphs can be obtained for other databases). This proves that the error is asymptotical as we increase c , and therefore, once arrived at certain error, the improvements are not significant. To observe how the error affects the reconstruction, in Figure 4.3b we can compare a reconstructed image with its corresponding original.

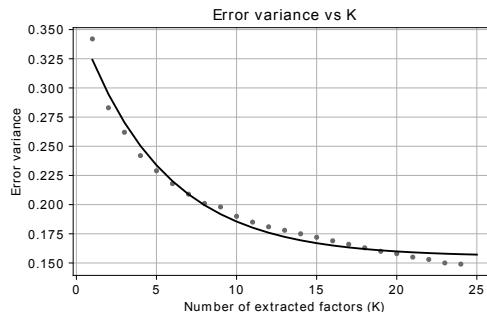


Figure 4.4: Specific variance of reconstruction error $\boldsymbol{\Psi}$ using FA, in function of number of factors extracted (K) for ADNI-PET database (the behaviour is similar in other datasets).

4.2.2 Independent Component Analysis (ICA)

Independent Component Analysis (ICA) [139] is an algorithm that performs decomposition imposing that the resulting components must be independent. It was used in [47, 59, 73] as part of a CAD system, and it had been used in other medical imaging applications such as segmentation [95].

ICA was born as a solution to the *blind source separation* problem, in which the aim is to estimate c independent sources from a series of mixed signals [139]. To do so, we assume the source signals to be non-gaussian, in addition to the independence assumption mentioned before. That is why their authors consider ICA to be a non-gaussian version of FA [125], although due to this assumption, the results are very different to those obtained in FA.

Unlike FA, ICA does not account for noise in the estimation procedure, and therefore the equation 4.1 remains:

$$\mathbf{X} = \mathbf{WS} \quad (4.12)$$

where \mathbf{S} are the component scores and \mathbf{W} are the component loadings, ‘sources’ or ‘mixing matrix’. Given that ICA lacks a noise term, there is a procedure called *whitening* that must be applied for the algorithm to converge [139]. The whitening implies a linear transformation of the i -th observed variable \mathbf{x}_i into a *white* vector $\tilde{\mathbf{x}}_i$ so that its covariance matrix equals the identity:

$$E\{\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T\} = \mathbf{I} \quad (4.13)$$

This procedure is often performed using the eigen-value decomposition (EVD) of the covariance matrix $E\{\mathbf{x}_i \mathbf{x}_i^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$. \mathbf{E} is the covariance matrix containing the eigenvectors of $E\{\mathbf{x}_i \mathbf{x}_i^T\}$, and \mathbf{D} is a diagonal matrix whose diagonal elements are the eigenvalues of $E\{\mathbf{x}_i \mathbf{x}_i^T\}$. Whitening is done using the following equation:

$$\tilde{\mathbf{x}}_i = \mathbf{ED}^{-1/2}\mathbf{E}^T \mathbf{x}_i \quad (4.14)$$

This procedure transform the mixing matrix to:

$$\tilde{\mathbf{x}}_i = \mathbf{ED}^{-1/2}\mathbf{E}^T \mathbf{Ws}_i \quad (4.15)$$

which is indeed orthogonal, as can be seen here:

$$E\{\tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_i^T\} = \tilde{\mathbf{W}} E\{\mathbf{s}_i \mathbf{s}_i^T\} \tilde{\mathbf{W}}^T = \tilde{\mathbf{W}} \tilde{\mathbf{W}}^T = \mathbf{I} \quad (4.16)$$

This property reduces the number of parameters to be estimated, since an orthogonal matrix contains $n(n - 1)/2$ degrees of freedom, in contrast to the n^2 degrees of freedom of the original mixing matrix \mathbf{W} .

As a consequence of the central limit theorem, we can assume that the sum of a large number of independent random variables is approximately normally distributed, regardless of the individual statistical distributions [110]. This property is used to maximize non-gaussianity and independence in the sources using any independence criteria such as the kurtosis or negentropy.

4.2.2.1 FastICA

In this work, we will use the FastICA algorithm. FastICA is a block fixed-point iteration algorithm [143, 152] based on negative entropy, or negentropy, as a non-gaussianity measure. Fixed-point algorithms converge faster than adaptive algorithms [143]. FastICA can be considered a neural algorithm [139], where we update the weight vector \mathbf{w} using a learning rule that finds the direction \mathbf{w} , a unit vector such that the projection $\mathbf{w}^\top \mathbf{x}_i$ maximizes non-gaussianity [143].

For a random vector \mathbf{y} , negentropy can be defined as:

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (4.17)$$

where \mathbf{y}_{gauss} and \mathbf{y} share the same covariance matrix, although \mathbf{y} is not a gaussian random variable, while \mathbf{y}_{gauss} is. There are many approximations to negentropy. The FastICA defines negentropy using the function:

$$J(\mathbf{y}) \propto [E\{G(\mathbf{y})\} - E\{G(\mathbf{v})\}]^2 \quad (4.18)$$

where we assume that \mathbf{y} is of zero mean and unit variance, \mathbf{v} is a Gaussian variable sharing the same mean and variance, and $G(x)$ can be any non-quadratic function. Many functions have been proposed, but in this algorithm we use either $G(x)_1 = (1/\alpha_1) \log \cosh(\alpha_1 x)$ with $1 < \alpha_1 < 2$ or $G(x)_2 = \exp(-x^2/2)$ [143]. With these measures, we can compute the derivatives of these functions by:

$$g_1(x) = \tanh(\alpha_1 x), \quad (4.19)$$

$$g_2(x) = x \exp(-x^2/2) \quad (4.20)$$

The algorithm for the one-unit version of FastICA can be defined [143] as:

1. Choose an initial (e.g. random) weight vector \mathbf{w} .
2. Let $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^\top \mathbf{x})\} - E\{g'(\mathbf{w}^\top \mathbf{x})\}\mathbf{w}$
3. Let $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$
4. If not converged, go back to 2.

The algorithm considers that the values of \mathbf{w} converge when their dot product is close to 1, that is, they are pointing in the same direction. Note that the expectations are computed as the sample mean. Additional modifications were presented in [139], in which step 2 is converted to a Newton iteration and further simplification is performed.

The four steps presented before compute one single component. Under the neural assumption, these steps are those performed at one single computational unit or neuron. We can then extend the procedure to c components by defining c neurons with weight vectors $\mathbf{w}_1, \dots, \mathbf{w}_c$ so that $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^\top$. The outputs $\mathbf{w}_1^\top \mathbf{x}, \dots, \mathbf{w}_n^\top \mathbf{x}$ must be decorrelated to prevent them from converging to the same maxima, using one of the three methods proposed in [139]. In this work, we have used a two-step iterative algorithm [139] to decorrelate the outputs after each iteration:

1. Let $\mathbf{W} = \mathbf{W} / \sqrt{\|\mathbf{W}\mathbf{W}^\top\|}$.
2. Let $\mathbf{W} = \frac{3}{2}\mathbf{W} - \frac{1}{2}\mathbf{W}\mathbf{W}^\top\mathbf{W}$

And repeat step 2 until convergence. For simplicity, the norm in step 1 can be computed as any norm but the Frobenius norm, for example, the L2-norm or the largest absolute row sum.

4.3 Results

In this work we will analyse the behaviour of the system proposed in the introduction and illustrated at Figure 4.2. The system comprises the selection of the most relevant voxels using filtering methods (we will focus on t-test, relative entropy and wilcoxon) and a feature decomposition of these feature vectors using either FA or ICA. Finally, we perform classification using a SVC with linear kernel, and performance values are obtained via cross-validation (see section 3.3 for more information).

We have varied the number of selected voxels and the number of factors or components depending on the algorithm and the dataset, and evaluated the system with those characteristics. That way, we obtain an estimation of the performance of the system in different situations, so that we can draw conclusions on the disease patterns and the ability of the system in the detection of different diseases.

4.3.1 Alzheimer's Disease

We begin by applying the proposed feature selection plus decomposition pipeline to the two functional neuroimaging datasets related to AD: ADNI-PET and VDNL-HMPAO. For this experiment we will use a maximum of 20,000 selected voxels and 25 components.

4.3.1.1 Factor Analysis

First, we test FA as a decomposition technique. In Figure 4.5 we average the accuracy over the number of voxels or the number of components respectively, to look at how these variables affect the performance of the system, and we do this for the three filtering methods used.

We can observe that the results are generally better when using the ADNI-PET dataset than with the VDNL-HMPAO, and this is especially notorious when using the relative entropy selection criterion. The performance tends to slightly increase with the number of voxels selected, but that is not the case with the number of components. By looking at figures 4.5b, 4.5d and 4.5f, it seems that a relatively small number of components (approximately 6) is enough to obtain good performance, and afterwards, the performance holds or even decreases.

4.3.1.2 Independent Component Analysis

In this section, we compute the results of applying ICA to the ADNI-PET and VDNL-HMPAO datasets. Figure 4.6 depicts the average accuracy over the number of voxels or the number of components respectively for the different selection criteria.

The case is similar to the one presented in section 4.3.1.1, where the performance slightly improves when increasing the number of selected voxels. The performance is again better when using the ADNI-PET dataset than with the VDNL-HMPAO, although the behaviour remains similar.

The results change when varying the number of components. In this case, although good performance is obtained within the first 5 components in most cases, the model achieves similar performance in both datasets with components between 5 and 10, and then, the estimates diverge. In the VDNL-HMPAO, the performance starts to decrease after this number of components, whereas when using the ADNI-PET dataset, the higher average performance is obtained with $c > 15$, especially in the case of the t-test or the wilcoxon selection criteria).

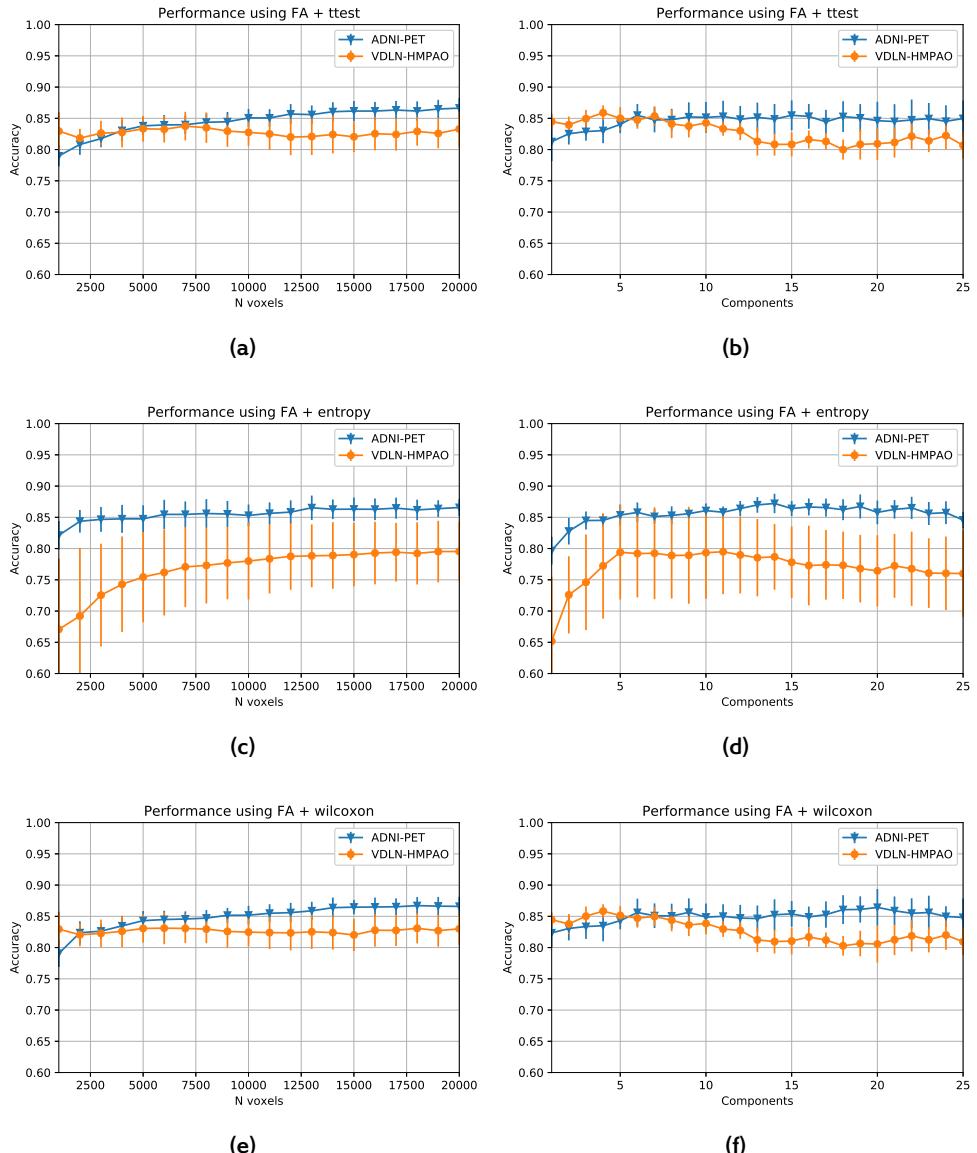


Figure 4.5: Average performance and standard deviation of the proposed system using the two AD datasets, FA and the three feature selection criteria: t-test - (a) and (b)-, relative entropy - (c) and (d)- and wilcoxon - (e) and (f)-.

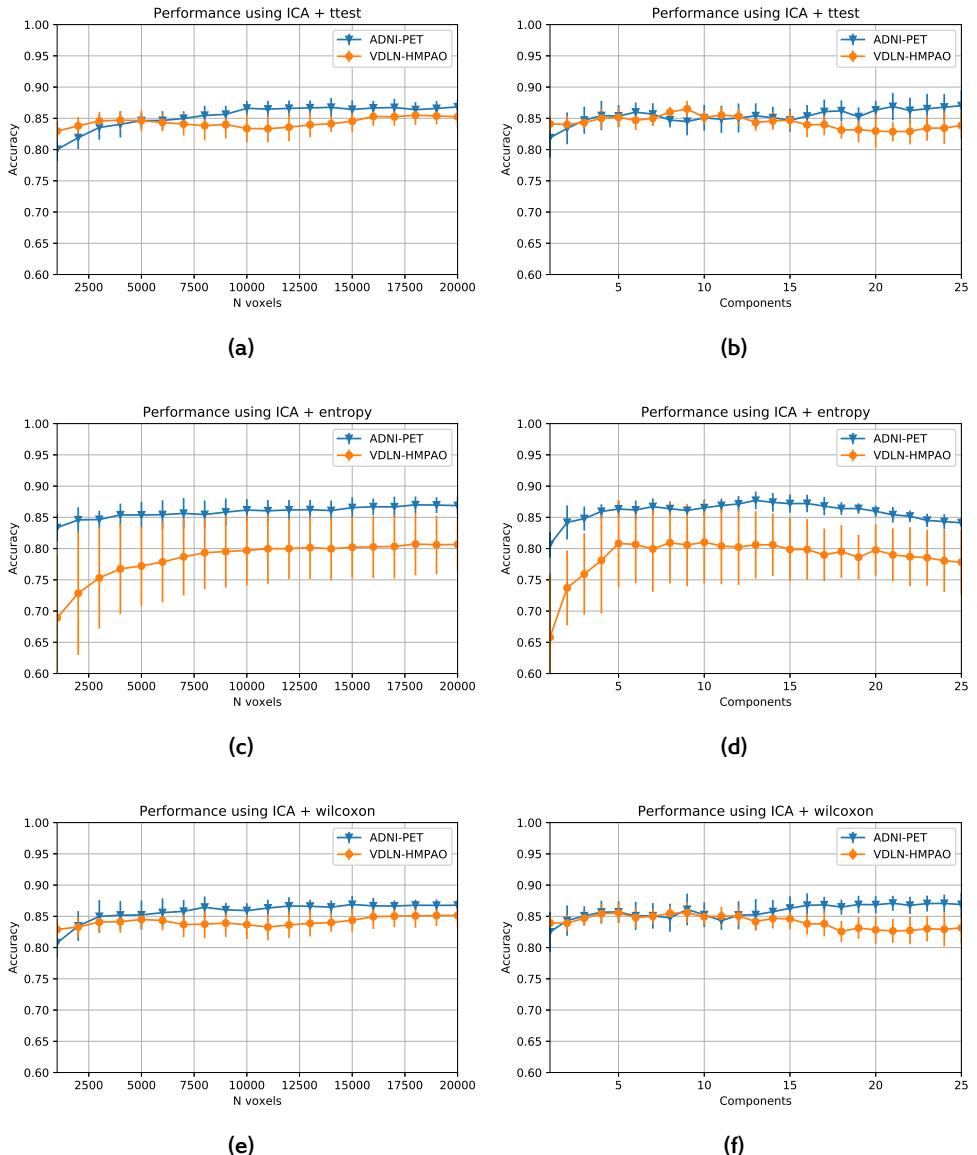


Figure 4.6: Average performance and standard deviation of the proposed system using the three AD datasets, ICA and the three feature selection criteria: t-test -(a) and (b)-, relative entropy -(c) and (d)- and wilcoxon -(e) and (f)-.

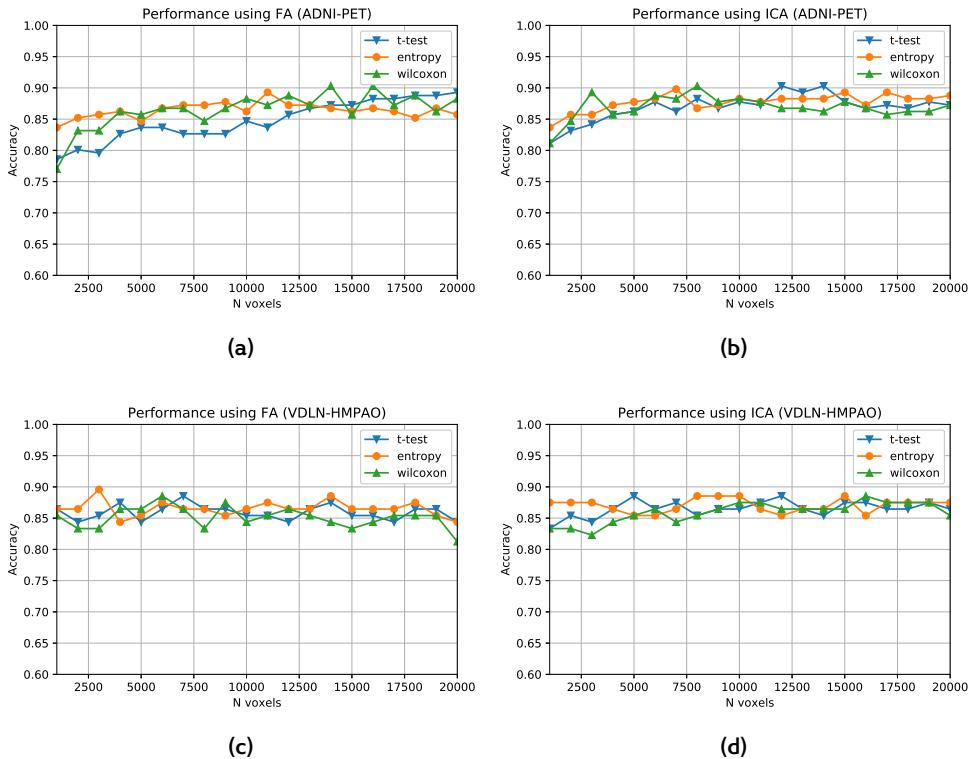


Figure 4.7: Performance of the proposed system using the two AD datasets: ADNI-PET and VDLN-HMPAO at the operation point, and how they vary over the number of selected voxels.

4.3.1.3 At the Operation Point

Now we focus on non-averaged values, the values for which our system is optimal: the operation point. In this scenario we see that the tendency is that all systems behave similarly.

When increasing the number of selected voxels, we can see that the system tends to slightly increase its performance in both datasets and decomposition strategies, (Figure 4.7). For the ADNI-PET dataset, the maximum accuracies are obtained with a large number of voxels $f > 15000$, however, in the case of VDLN-HMPAO, we obtain similar performance with fewer voxels, $f < 7000$ in all selection criteria.

Now we can focus on the performance variations over the number of components in Figure 4.8. The accuracy slightly varies almost in any case, and

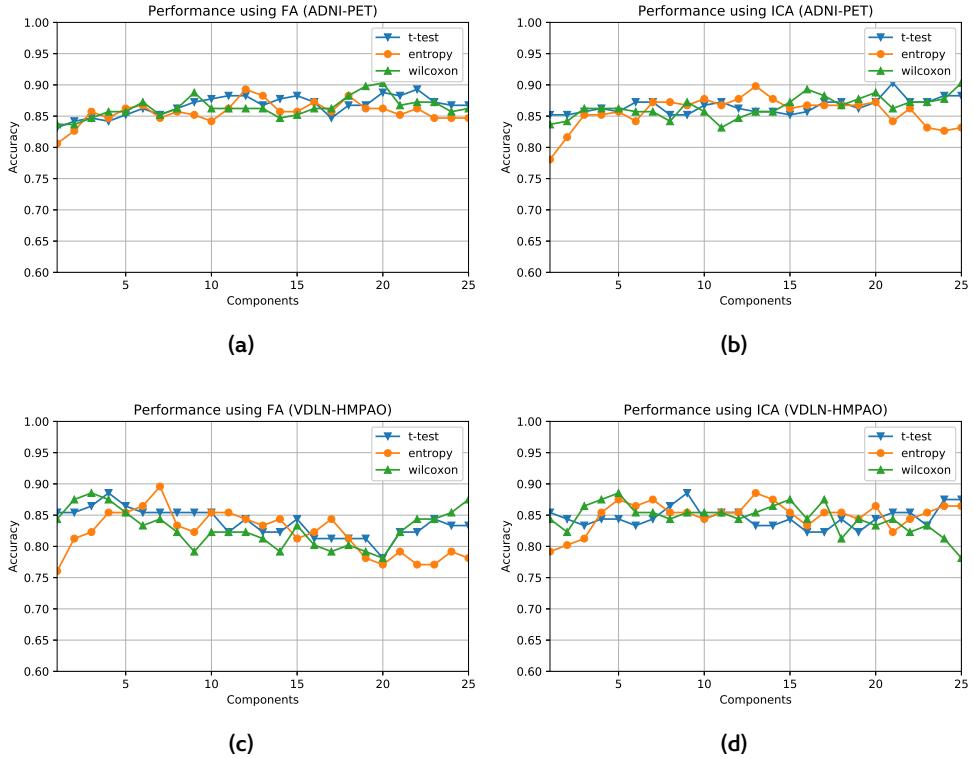


Figure 4.8: Performance of the proposed system using the two AD datasets: ADNI-PET and VDLN-HMPAO at the operation point, and how they vary over the number of components used in the decomposition.

there is a steep increase in the performance within the first five components in both FA and ICA.

A particular case is the combination of FA and the relative entropy selection criteria applied to the VDLN-HMPAO dataset. In this case there seems to be a trend to achieve a maximum performance at less than 5 components. However, with the ADNI-PET dataset the performance holds and achieves maximum accuracy with $c \approx 20$.

In Table 4.1 we show the performance values obtained at the operation point for our different tests combining decomposition algorithms and selection criteria. It is obvious that both datasets obtain similar performance, with values close to 0.9. These values are compared to the baseline classification performance, quantified using VAF [120]. From this, we can see that the decomposition systems always perform better than the baseline, but this difference is

DB	Dec.	Criterion	Accuracy	Sensitivity	Specificity
ADNI-PET	VAF	-	0.882 ± 0.064	0.876 ± 0.099	0.890 ± 0.097
		t-test	0.893 ± 0.074	0.886 ± 0.119	0.901 ± 0.101
		FA	0.893 ± 0.074	0.894 ± 0.092	0.891 ± 0.088
	ICA	wilcoxon	0.903 ± 0.066	0.917 ± 0.079	0.891 ± 0.082
		t-test	0.903 ± 0.071	0.893 ± 0.100	0.910 ± 0.107
		entropy	0.898 ± 0.059	0.917 ± 0.088	0.881 ± 0.084
VDLN-HMPAO	VAF	wilcoxon	0.903 ± 0.066	0.906 ± 0.097	0.901 ± 0.094
		-	0.802 ± 0.074	0.803 ± 0.088	0.805 ± 0.145
		t-test	0.885 ± 0.076	0.890 ± 0.127	0.875 ± 0.149
	ICA	FA	0.896 ± 0.092	0.907 ± 0.150	0.875 ± 0.139
		wilcoxon	0.885 ± 0.076	0.923 ± 0.130	0.825 ± 0.154
		t-test	0.885 ± 0.073	0.923 ± 0.130	0.825 ± 0.154
	FA	entropy	0.885 ± 0.076	0.903 ± 0.132	0.850 ± 0.130
		wilcoxon	0.885 ± 0.076	0.907 ± 0.130	0.850 ± 0.152

Table 4.1: Accuracy, sensitivity, specificity, and their standard deviation at the operation point for each method and its corresponding feature selection criterion, using two AD datasets.

especially large in the case of the VDLN-HMPAO dataset, which contains very noisy images.

Overall, all methods achieve similar values of accuracy, sensitivity and specificity. When analysing the ADNI-PET dataset, the wilcoxon selection criteria seems to outperform the rest especially with ICA, with a higher accuracy and sensitivity. On the other hand, with the VDLN-HMPAO dataset, either the t-test or the wilcoxon achieves higher sensitivity, but there is no difference in accuracy when decomposing with the ICA algorithm. When using FA decomposition, the relative entropy seems to perform better. In general, there seems to be little difference among methods, and a curious relationship between the relative entropy selection and the VDLN-HMPAO dataset that we will discuss later.

4.3.2 Parkinson’s Disease

Now we will look at how the proposed CAD system behaves when applied to the three DaTSCAN datasets: VDLN-DAT, VDLV-DAT and PPMI-DAT. For this experiment we will use a maximum of 1500 selected voxels and 25 com-

ponents, and images that have been previously intensity normalized using the integral normalization algorithm (see section 3.2).

4.3.2.1 *Factor Analysis*

Firstly, we will explore the average behaviour of the system that uses FA as a decomposition technique. For this purpose, as in previous sections, Figure 4.9 shows how the average performance varies when increasing the number of voxels selected and the number of components extracted.

In this case there is a clear difference between datasets, since some of them are more complex than others, usually due to a typical acquisition procedure in DaTSCAN and the subclasses of PKS included. In the VDNL-DAT, the images were often composed only of a few cuts around the striatum, whereas in PPMI and VDLV-DAT this rarely happens. That would explain the average outperformance of these two datasets over the VDNL-DAT in almost all cases.

Their behaviours are consistent. Usually, there is no variation with the number of voxels selected (except in the case of the VDNL-DAT dataset and the relative entropy selection criterion). However, there are repeated trends regarding the number of components or factors used in the computation. We can see that the performance increases in the first components, and once we have achieved a decent performance (between 4 and 6 components), it starts to decrease. This could mean that the use of more than 5 or 6 components only introduces noise and leads to a wrong decomposition.

Again, this behaviour is consistent with the VDNL-DAT dataset, except for one case. It does seem that the interaction between the VDNL-DAT dataset and the relative entropy selection criterion leads to a very different model. We will explore this question later, in the discussion.

4.3.2.2 *Independent Component Analysis*

Regarding the application of the ICA decomposition model to the DaTSCAN datasets. In Figure 4.10 we present the average accuracy achieved by this model using three different selection criteria and the three PKS datasets.

In average, PPMI-DAT and VDLV-DAT achieve similar performance, while VDNL-DAT generally achieves much poorer results. However, the behaviour of the system when varying the number of voxels or number of components is similar in all three datasets. The tendency is that accuracy does not significantly vary when increasing the number of voxels (except in the case of the relative entropy and VDNL-DAT). In contrast, when varying the number of components, we observe that the maximum performance is achieved in the first components, typically between 5 and 10.

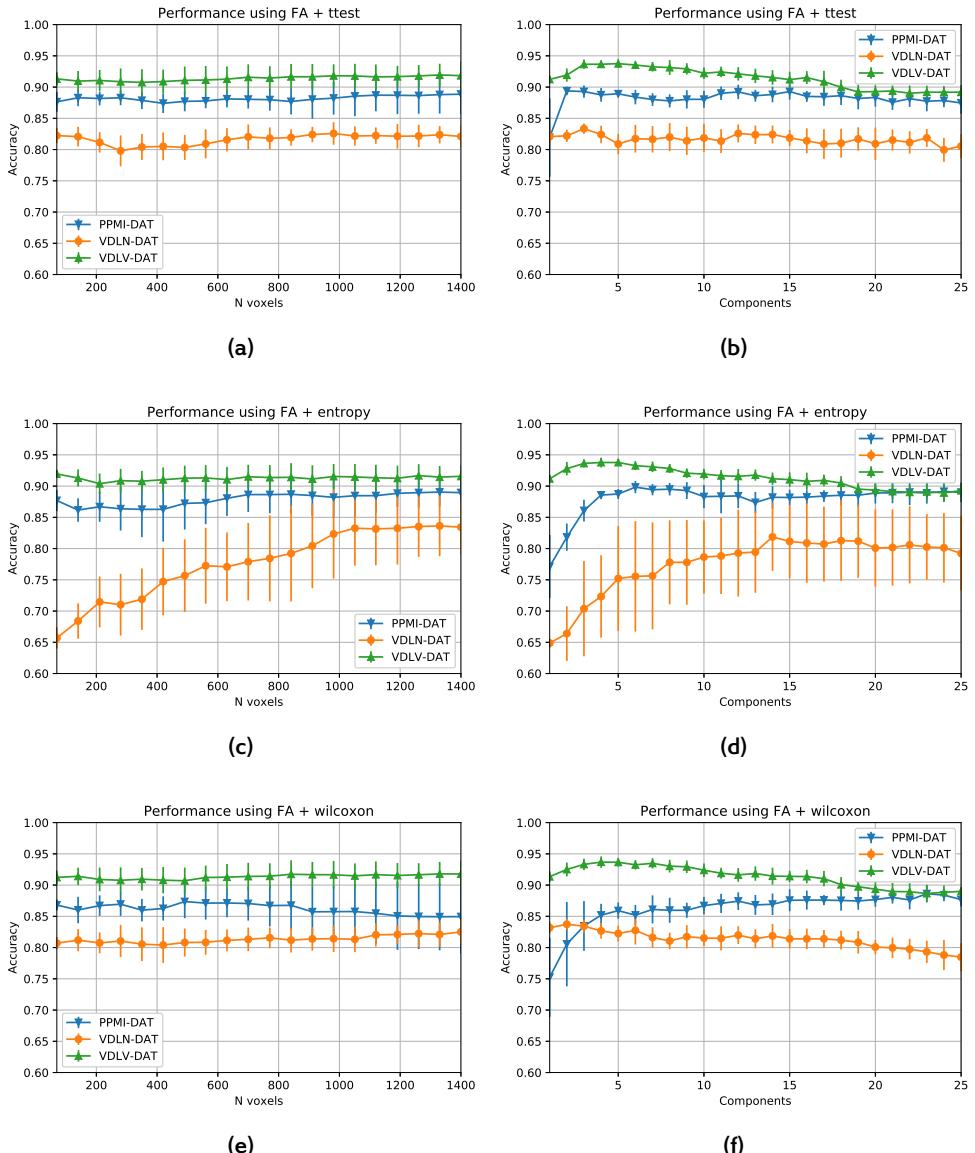


Figure 4.9: Average performance and standard deviation of the proposed system using the three PKS datasets, FA and the three feature selection criteria: t-test -(a) and (b)-, relative entropy -(c) and (d)- and wilcoxon -(e) and (f)-.

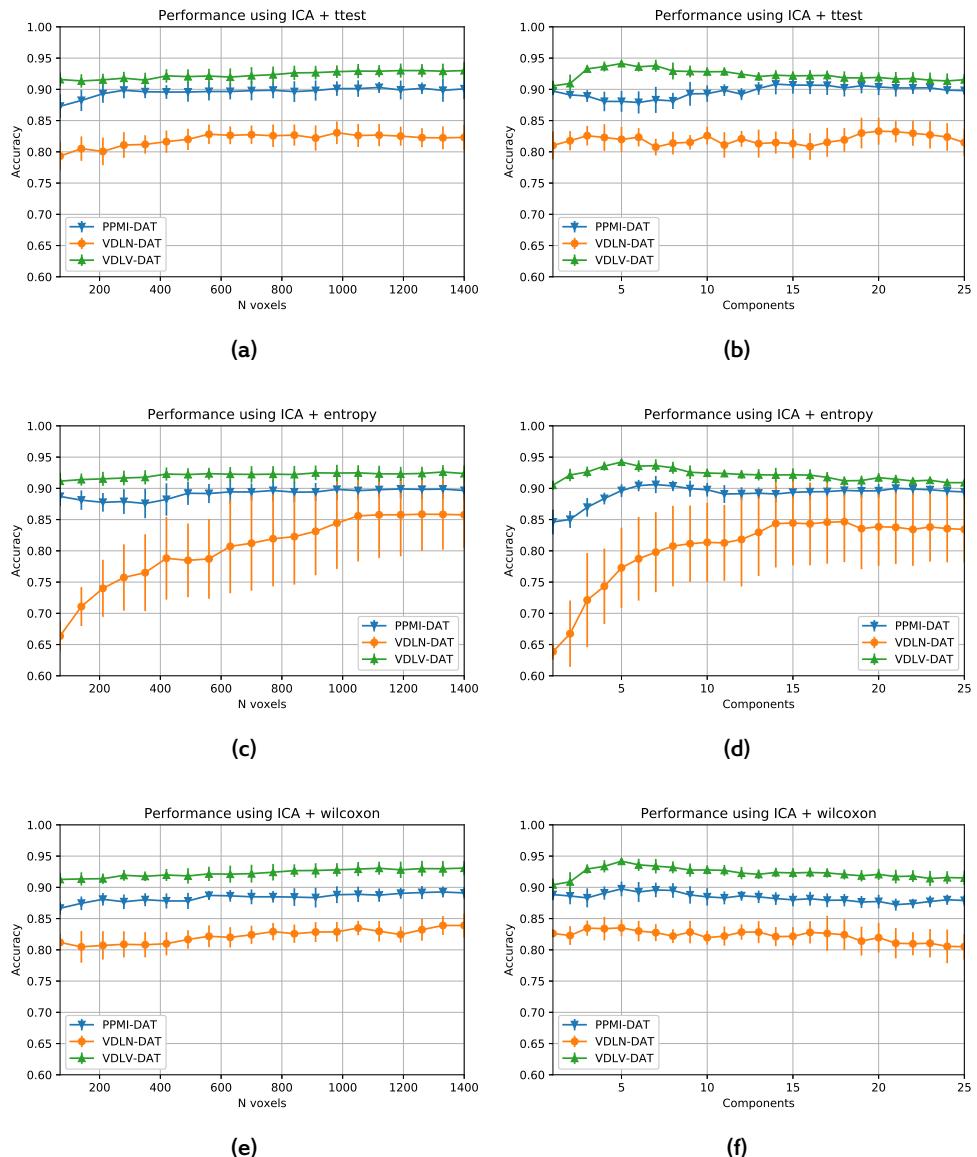


Figure 4.10: Average performance and standard deviation of the proposed system using the three PKS datasets, ICA and the three feature selection criteria: t-test -(a) and (b)-, relative entropy -(c) and (d)- and wilcoxon -(e) and (f)-.

4.3.2.3 At the Operation Point

Let us now have a look at the behaviour of this system at the operation point, using the parameters c and f for which our system is optimal. When varying the number of voxels selected, we obtain the graphs presented at Figure 4.11.

In this case, we obtain again that our system maintains approximately the same performance regardless of the number of voxels selected on both the PPMI-DAT and the VDLV-DAT datasets. In contrast, the performance rises when we increase the number of selected voxels on the VDLD-DAT dataset, especially when using the relative entropy criterion. This latter dataset also performs worse, whereas the VDLV-DAT achieves the best performance.

As for the changes in performance when varying the number of components, the results for both FA and ICA based systems are shown in Figure 4.12.

In this first case, the most evident result is the general performance of our system in the three datasets. It is clear that the system performs better when tested on VDLV-DAT than when tested on PPMI-DAT, and both datasets outperform VDLD-DAT. It is even clearer that when testing on VDLV-DAT, the results are similar using any type of decomposition and selection criteria, with similar performance. We hypothesize several reasons for this behaviour at the discussion of this chapter.

The tendency of the performance at the operation point is similar to the average behaviour commented before. In general, there is an accuracy increasing in the first components (usually, between 5 and 10 depending on the decomposition) and then, the performance remains stable. Again, the combination of relative entropy selection and VDLD-DAT leads to striking results. For this dataset, the higher performance is obtained with more than 10 components ($c = 14$), and it shows higher variability than other datasets.

Now we will focus on the specific performance values obtained at the operation point, that can be seen in Table 4.2. In this table we observe the differences in performance between datasets and also between CADs using each decomposition strategy.

In general, the systems using ICA tend to perform slightly better than those using FA, although the difference is small. There is little difference between selection criteria as well, although the combination of relative entropy and ICA seems to work better, at least in the PPMI-DAT and VDLD-DAT (in VDLV-DAT all combinations perform equally well).

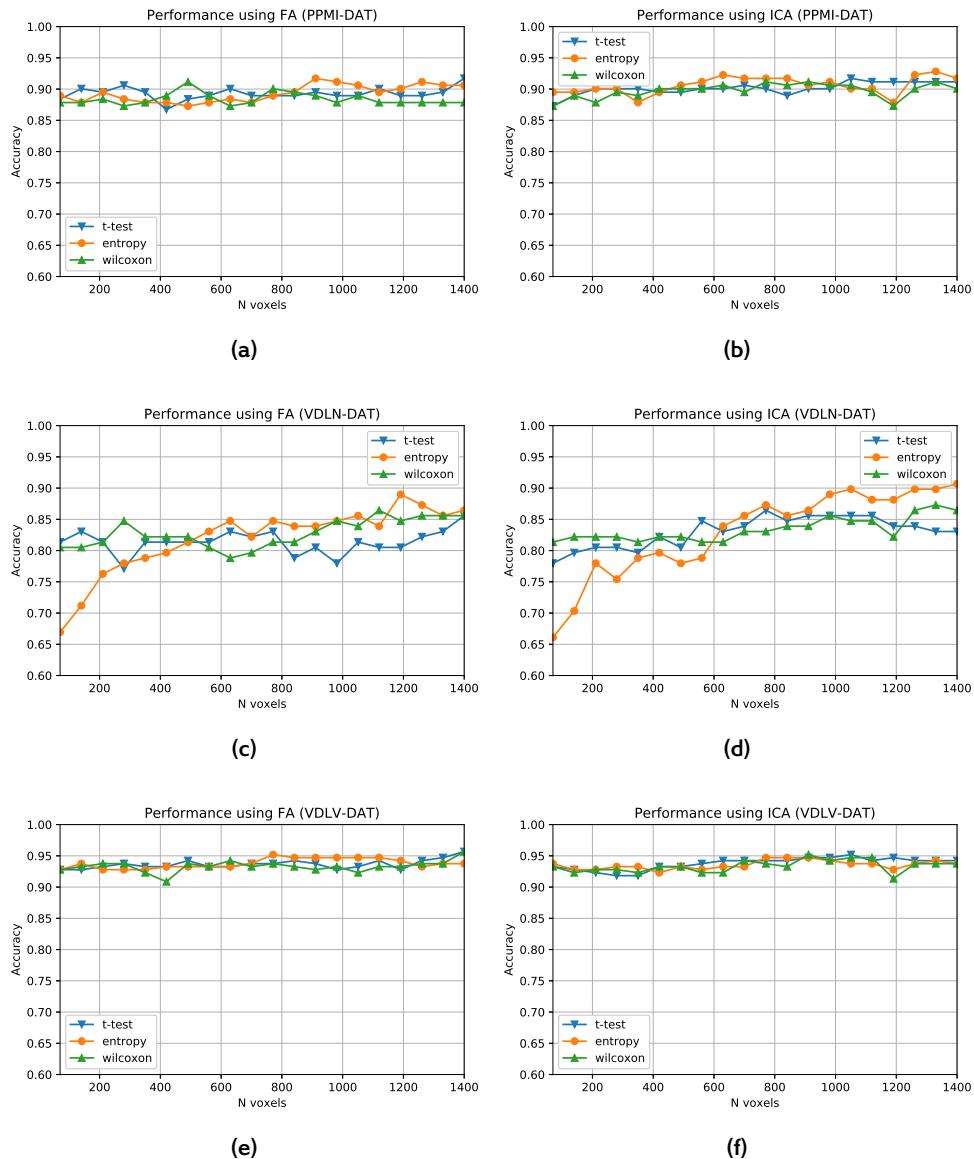


Figure 4.11: Performance of the proposed system using the two PKS datasets: PPMI-DAT, VDLN-DAT and VDLV-DAT at the operation point, and how they vary over the number of selected voxels.

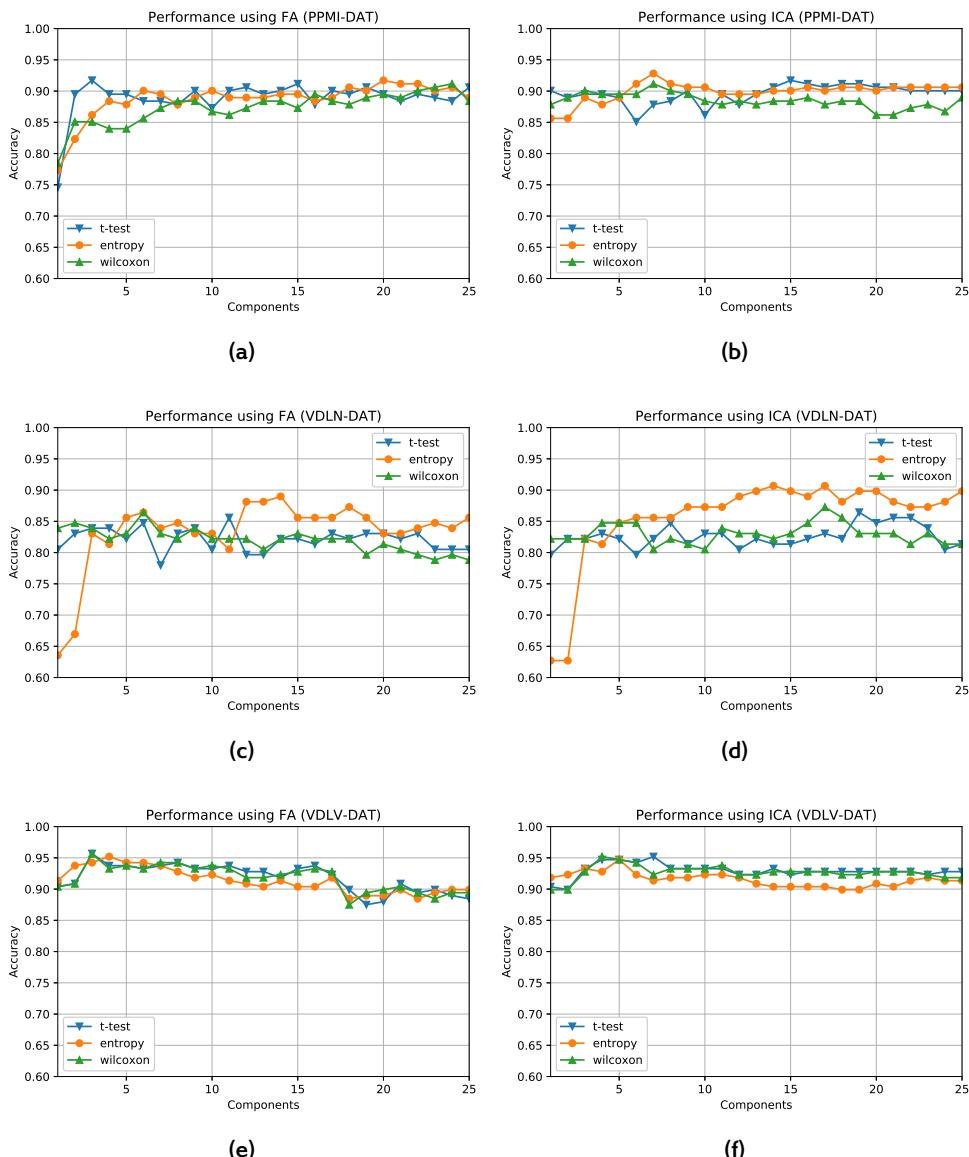


Figure 4.12: Performance of the proposed system using the two PKS datasets: PPMI-DAT, VDLN-DAT and VDLV-DAT at the operation point, and how they vary over the number of components used in the decomposition.

DB	Dec.	Criterion	Accuracy	Sensitivity	Specificity
PPMI-DAT	VAF	-	0.800 ± 0.071	0.831 ± 0.093	0.747 ± 0.112
		t-test	0.917 ± 0.037	0.918 ± 0.095	0.918 ± 0.091
	FA	entropy	0.917 ± 0.060	0.918 ± 0.076	0.921 ± 0.120
		wilcoxon	0.912 ± 0.056	0.927 ± 0.098	0.889 ± 0.102
	ICA	t-test	0.917 ± 0.056	0.900 ± 0.095	0.948 ± 0.109
		entropy	0.928 ± 0.055	0.909 ± 0.091	0.961 ± 0.090
		wilcoxon	0.912 ± 0.070	0.909 ± 0.100	0.920 ± 0.118
	VAF	-	0.796 ± 0.129	0.860 ± 0.143	0.675 ± 0.208
VDLN-DAT	VAF	t-test	0.856 ± 0.111	0.887 ± 0.178	0.795 ± 0.164
		FA	0.890 ± 0.098	0.875 ± 0.118	0.910 ± 0.116
	ICA	entropy	0.864 ± 0.070	0.916 ± 0.114	0.780 ± 0.183
		wilcoxon	0.864 ± 0.101	0.873 ± 0.174	0.840 ± 0.166
	VAF	t-test	0.907 ± 0.075	0.889 ± 0.124	0.935 ± 0.131
		ICA	0.873 ± 0.108	0.859 ± 0.181	0.890 ± 0.151
VDLV-DAT	VAF	-	0.918 ± 0.062	0.900 ± 0.094	0.926 ± 0.087
		t-test	0.957 ± 0.063	0.910 ± 0.094	0.973 ± 0.065
	FA	entropy	0.952 ± 0.037	0.940 ± 0.066	0.964 ± 0.064
		wilcoxon	0.957 ± 0.033	0.940 ± 0.066	0.973 ± 0.065
	ICA	t-test	0.952 ± 0.037	0.940 ± 0.066	0.964 ± 0.064
		entropy	0.947 ± 0.045	0.940 ± 0.066	0.955 ± 0.076
		wilcoxon	0.952 ± 0.037	0.940 ± 0.066	0.964 ± 0.064

Table 4.2: Accuracy, sensitivity, specificity, and their standard deviation at the operation point for each method and its corresponding feature selection criterion, using three PKS datasets

4.4 Discussion

Now we will discuss the general behaviour of the selection and decomposition algorithm in the CAD systems proposed and how they perform on the different diseases and databases.

Our CAD system performs reasonably well on the AD datasets, where it achieves around 90% accuracy, and more than 92% sensitivity. This is achieved in both systems composed by FA and ICA regardless of the selection criterion chosen. The system outperforms the visual analysis, estimated using VAF [120], in both cases.

In [59] and [47], the systems achieved better performance (up to 95.1% accuracy) when using multivariate quadratic classifiers and ICA, different from the SVC used here. We have chosen to evaluate the system only on linear SVCs for two main reasons. First, it favours a side-by-side comparison between all methods applied to different datasets in this thesis. And second, multivariate quadratic classifier could be more prone to overfitting, and conversely linear SVC has been proven to be better generalizable than other systems, even in environments where the small sample size is the norm [153].

The selected areas on these CAD systems correspond to the highlighted areas in Figure 4.13, in the case of ADNI-PET dataset.

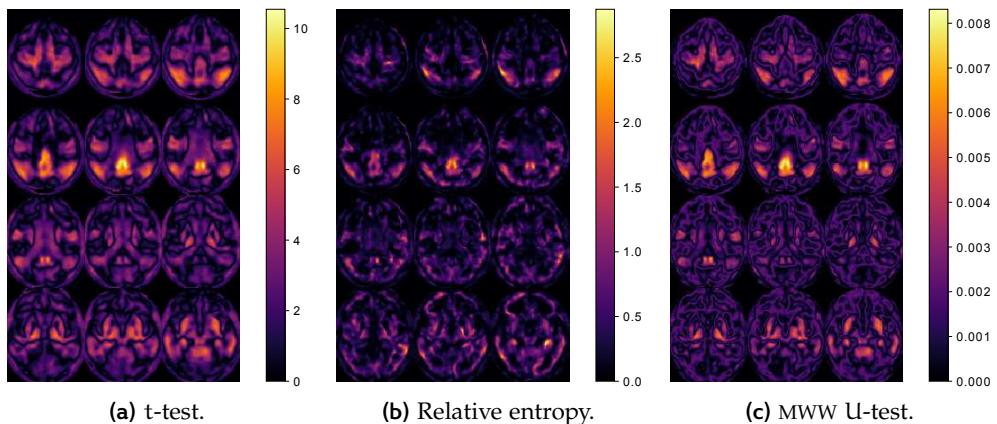


Figure 4.13: Comparison between the different filtering methods, and the regions selected by them, in the ADNI-PET dataset.

For its part, in the VDLN-HMPAO different areas are selected as we can see in Figure 4.14. This is mainly due to a change in the modality that deserves to be analysed.

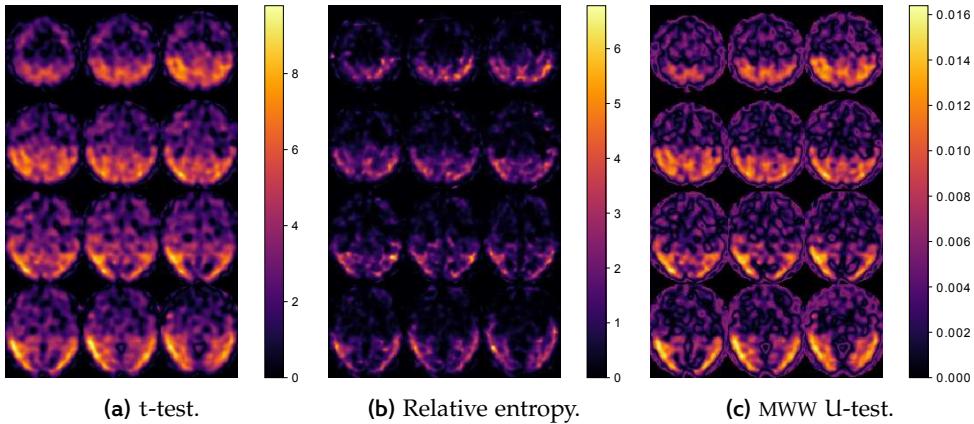


Figure 4.14: Comparison between the different filtering methods, and the regions selected by them, in the VDLN-HMPAO dataset.

To have a more profound look at the selected regions in both modalities, we provide Table 4.3, where the AAL regions with an overlapping higher than 0.5 in any of the modalities and selection criteria are displayed.

In the case of the VDLN-HMPAO dataset, the most interesting regions are located at the occipital lobe, the angular lobe and few of them in the temporal lobe. These are selected using almost any of the selection criteria. However, when using the ADNI-PET dataset, the only region with a significant overlapping is the angular gyrus, but other regions with a widely documented relation to AD such as the cingulum, hippocampus and parahippocampal gyrus are highlighted [96, 160].

It is clearly noticeable that the relative entropy selection criterion focuses on a variety of regions, but it is the only one able to detect the hippocampus or parahippocampal gyrus in the PET dataset, which other criteria ignore. It also focuses more on the different parts of the occipital lobe in the SPECT dataset. This difference in the selected areas could lead to the different overall performance observed in Figures 4.5c, 4.5d, 4.6c, and 4.6d.

For its part, wilcoxon and t-test often select similar regions. This can be due to their similarity under the normal distribution [66], and leads to a higher performance of the systems both in average and at the operation point (see Figures 4.5, 4.6, 4.7 and 4.8). From the selected regions, and since t-test and wilcoxon perform generally better, we can infer the more interesting regions for AD classification. For the ADNI-PET dataset, these would be the angular lobes and the cingulum, whereas for the VDLN-HMPAO, we can observe differ-

Region	ADNI-PET			VDDLN-HMPAO		
	entropy	t-test	wilcoxon	entropy	t-test	wilcoxon
Angular gyrus (left)	1.000	1.000	1.000	1.000	0.858	1.000
Angular gyrus (right)	0.729	0.782	0.809	0.980	0.566	0.697
Cingulum posterior part (left)	1.000	1.000	1.000	0.589	0.509	0.773
Cingulum posterior part (right)	0.843	0.852	0.791	0.524	0.228	0.442
Cuneus (right)	0.173	0.121	0.188	1.000	0.683	0.857
Fusiform gyrus (left)	0.495	0.156	0.167	0.652	0.561	0.713
Hippocampus (right)	0.807	0.375	0.413	0.235	0.155	0.425
Inferior occipital gyrus (left)	0.378	0.078	0.023	1.000	1.000	1.000
Inferior occipital gyrus (right)	0.380	0.080	0.011	1.000	0.794	0.930
Inferior parietal lobule (right)	0.497	0.414	0.439	1.000	0.339	0.477
Inferior temporal gyrus (left)	0.748	0.350	0.495	0.676	0.561	0.735
Middle occipital gyrus (left)	0.449	0.184	0.138	1.000	0.881	1.000
Middle occipital gyrus (right)	0.347	0.174	0.128	0.861	0.642	0.770
Middle temporal gyrus (left)	0.411	0.244	0.364	0.952	0.635	0.810
Middle temporal gyrus (right)	0.594	0.234	0.326	1.000	0.537	0.711
Parahippocampal gyrus (left)	0.778	0.444	0.465	0.098	0.132	0.262
Parahippocampal gyrus (right)	0.917	0.276	0.316	0.233	0.159	0.311
Precuneus (left)	0.322	0.429	0.463	0.708	0.485	0.660
Precuneus (right)	0.311	0.375	0.409	0.770	0.469	0.644
Superior occipital gyrus (right)	0.268	0.094	0.106	0.955	0.600	0.736
Supramarginal gyrus (left)	0.195	0.156	0.183	0.627	0.561	0.762

Table 4.3: Percentage of overlap between the selected areas by each method and the AAL atlas regions. For simplicity, overlapping values higher than 0.8 are displayed in bold.

ences in the angular lobe and also all over the occipital lobe and parts of the temporal lobe.

Regarding the decomposition method, there seems not to be any significant differences. Both ICA and FA perform similarly in any dataset, regardless of the noise contained in the images. However, the differences with the baseline system are much higher in the case of the VDNL-HMPAO dataset. This is perhaps due to the smoother nature of the ADNI-PET, in which several images from the same subject were averaged, and therefore, much of the noise was removed, whereas in the VDNL-HMPAO images, the noise could be removed afterwards by discarding many of the lower-significance components.

When applied to the PKS datasets, the results differ from the PPMI-DAT and VDNL-DAT to the VDLV-DAT. These databases differ in the number of subjects they contain. Whereas in the former there are different subjects with PKS, including subjects without evidence of dopaminergic deficit (SWEDD), in the later we only have PD and CTL subjects, which makes the classification easier.

This accuracy differences can be noticed throughout all figures and tables, although in general, the VDNL-DAT dataset has the lowest performance, while the PPMI-DAT and VDLV-DAT behave similarly in average.

When using FA with the PKS datasets, we observe a similar behaviour to that already seen with AD datasets: there is little difference in performance when varying the number of selected voxels, except when using the relative entropy criterion. In this particular case, there is a rise in the average performance when increasing the number of selected voxels, which is more noticeable when using the VDNL-DAT dataset.

For its part, there are more remarkable variations when increasing the number of components. Smaller c values lead to a fast increase in performance up to a maximum. Depending on the selection criterion used, this value varies from $c = 3$ when using the t-test to $c = 14$ for the entropy criterion applied to VDNL-DAT. The optimal c is usually located in the interval $c \in [3, 5]$ in most cases, as can be seen in Figure 4.9. A very similar behaviour is achieved when using the ICA decomposition, as depicted in Figure 4.10.

In AD we provided a table with the selected regions in both PET and SPECT modalities. Conversely, in DaTSCAN imaging, the selected regions are always located within the striatum, and the smaller resolution of these images hardly reveals the underlying structures. The selected regions with either t-test or wilcoxon fundamentally cover the whole caudate, putamen and globus pallidus, and some external structures as well. However, the relative entropy criterion focus almost exclusively in the striatum, with a strong preference for the posterior part, and discards all other regions, introducing less noise.

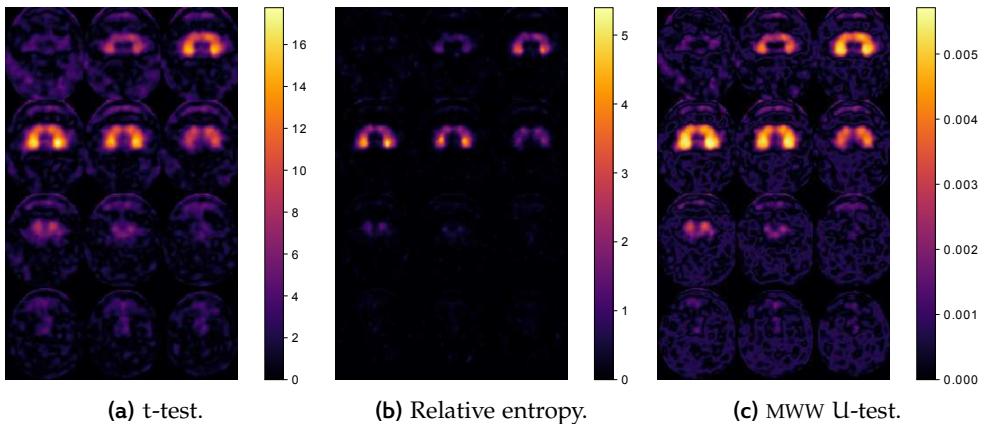


Figure 4.15: Comparison between the different filtering methods, and the regions selected by them, in the PPMI-DAT dataset.

This is clearly seen in Figure 4.11 where the performance around the operation point is displayed. Here, when looking at the VDLN-DAT dataset, the correlation between performance and number of selected voxels is more obvious. As can be seen in Table 4.2, in the two datasets where SWEDD subjects are included the system which uses relative entropy achieves better results. However, in the VDLV-DAT, where the system only involves PD and CTL subjects, the performance is very similar using all three selection criteria.

When looking at the evolution of the performance with the number of selected components, (Figure 4.12), the pattern observed in AD holds for the PPMI-DAT and the VDLV-DAT datasets. In these cases, maximum performance is obtained with a relatively small c (between 4 and 8, depending on the decomposition algorithm). However, with the relative entropy criterion, the VDLN-DAT still needs a higher number (more than 10) to reach the operation point.

All these differences in behaviour could be due to a higher variability in VDLN-DAT, compared to the other two datasets. The number of components needed, especially in FA, points to a more complex decomposition of those images. The sources of variability in this dataset probably correspond to a higher proportion of SWEDD subjects and the number of cuts used in the acquisition. The number of SWEDD in VDLN-DAT is 30 for a total dataset of 148 patients, whereas in the PPMI-DAT dataset we only have 31 SWEDD for 301 subjects. Furthermore, the number of cuts in the images of VDLN-DAT differs from one image to another, since they follow a common practice in which only the ROIs of the brain (the striatum) are acquired.

5

TEXTURE FEATURES

5.1 Introduction

Texture is a household word outside the image processing or related fields. However, in that context, it lacks a definition that allow us to measure and quantify it. Pattern recognition provides us with a mathematical definition that allow us to use texture as a feature in our CAD systems.

Texture analysis is any procedure by which we can quantify and classify the spatial variation of intensity of an image. In neuroimaging, texture has been widely used in segmentation (tissue classification) of MRI images [80, 121, 130], although there exist a number of works using it for feature extraction in CAD-like systems, like the works in [16, 135], or our work on PKS feature extraction [25, 34].

Texture features can be grouped in first, second and higher order analyses, depending on the number of variables used. First order statistics [5] are the most basic form of texture analysis, computing values such as average, variance or histogram of voxel intensity values [91].

The most popular form, with a very developed theoretical background, is second-order statistical texture analysis. This particular form is based on the probability of finding a pair of similar intensities at a certain distance and orientation of a certain image. From these probabilities, many measures can be derived, being the Haralick texture analysis [172] perhaps the most popular approach.

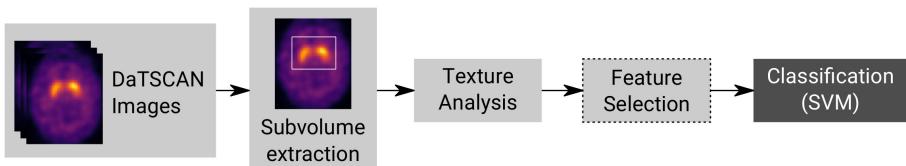


Figure 5.1: Schema of the proposed Texture-based CAD system, including an optional feature selection block.

In this work we have used Haralick texture analysis to extract features from DaTSCAN images and perform an automatic diagnosis of PD. It follows the pipeline depicted at Figure 5.1, as in [25, 34]. First we will provide an in-

introduction to the methodology followed at Section 5.2, including the volume selection tools, the Haralick texture analysis and the experiments used to validate the system. Later, in Section 5.3 we define the experiments and show their results. Finally, at Section 5.4 we discuss the implications of this systems and the evaluation results of our texture-based CAD system.

5.2 Methodology

5.2.1 Volume selection

Even the registered DaTSCAN images contain many voxels that are outside the brain. Therefore, to obtain a more robust estimation of the texture, it would be desirable to perform the computation of the features on subvolumes of those images (or subimages) that contain only voxels inside the brain. Many strategies can be performed for this, for example, force the texture analysis to ignore background voxels.

In this work, given the characteristics of the DaTSCAN ROIs, we opted for extracting a subvolume which contains only voxels higher than a certain intensity threshold I_{th} , which should be specified. To do so, we obtain the maximum and minimum coordinates for which \mathbf{I} is higher than the threshold:

$$p_{x,\min} = \arg \min_x (\mathbf{I} > I_{th}) \quad (5.1)$$

$$p_{x,\max} = \arg \max_x (\mathbf{I} > I_{th}) \quad (5.2)$$

And we do the same for the y and z axis of the array. Once this is computed, we can select the volume by:

$$\mathbf{I}_{\text{sub}} = \mathbf{I}[p_{x,\min} : p_{x,\max}, p_{y,\min} : p_{y,\max}, p_{z,\min} : p_{z,\max}] \quad (5.3)$$

The resulting subvolume \mathbf{I}_{sub} is the minimum box-shaped volume containing all the values for which $\mathbf{I} > I_{th}$, which allow us to select a I_{th} so that only the regions of interest are contained within.

Different subimages and sizes are obtained when applying different I_{th} . In Figure 5.2 we depict a comparison between the resulting images for $I_{th} = [0.25, 0.30, 0.35]$.

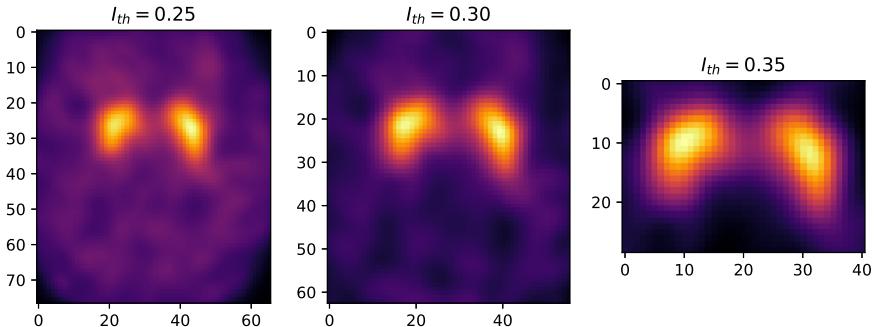


Figure 5.2: Comparison of the different I_{th} values for a random subject extracted from the PPMI-DAT database.

5.2.2 Haralick Texture Analysis

5.2.2.1 Gray Level Co-occurrence Matrix

The Haralick texture analysis is based on the computation of a Grey Level Co-occurrence Matrix (GLCM). This matrix is a summary of the probabilities of finding a pair of similar grey levels at a certain distance and in a certain direction.

The combination of the unitary vector dimension and the distance defines the offset $\Delta = (d_x, d_y, d_z)$, whose norm is the distance d and is defined in a given spatial direction. In this work, we use a three-dimensional approach to the computation of the GLCM, based on [90], that uses thirteen spatial directions to generalize the standard 2D GLCM to 3D. These offset define different angles and are used to get a degree of rotational invariance [90].

Medical images are described using different number precisions, which can vary from regular 8bit integers (256 values) to the type float64 (1.844×10^{19} possible values). Using all these values, even in the smallest case, would lead to 256×256 matrices, which would be both non representative of the real texture and computationally expensive. Therefore, prior to the GLCM computation, we posterize the image, that is, the image is quantified to use only 16 grey levels. This leads to more tractable GLCM without losing their representativeness.

Once images have been posterized, for two different grey levels i and j , the value of the co-occurrence matrix C over a $n \times m \times k$ three-dimensional image I is defined as:

$$C_{\Delta}(i, j) = \sum_{p=(1,1,1)}^{(n,m,k)} \begin{cases} 1, & \text{if } I(p) = i \text{ and } I(p + \Delta) = j \\ 0, & \text{otherwise} \end{cases} \quad (5.4)$$

where Δ is the three dimensional offset that we defined previously, and \mathbf{p} is the position of a given voxel inside the image.

We will compute one 16×16 GLCM for each of the combinations of direction and distances. This matrix \mathbf{C}_Δ is later modified to create the probability matrix \mathbf{P} as:

$$\mathbf{P}(i, j) = \frac{\mathbf{C}_\Delta(i, j)}{\sum_{i,j} \mathbf{C}_\Delta(i, j)} \quad (5.5)$$

from which the texture features will be derived.

5.2.2.2 Haralick Texture Features

In [165, 172], many texture features are derived from the probability matrix defined above. We have selected twelve of these features to use in this work. These features are:

$$\text{Energy} = \sum_i \sum_j \mathbf{P}(i, j)^2 \quad (5.6)$$

$$\text{Entropy} = \sum_i \sum_j \mathbf{P}(i, j) \log \mathbf{P}(i, j) \quad (5.7)$$

$$\text{Correlation} = \frac{\sum_i \sum_j ij \mathbf{P}(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (5.8)$$

$$\text{Contrast} = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{|i-j|=n} \mathbf{P}(i, j) \right\} \quad (5.9)$$

$$\text{Variance} \sum_i \sum_j (i - \mu_i)^2 \mathbf{P}(i, j) + (j - \mu_j)^2 \mathbf{P}(i, j) \quad (5.10)$$

$$\text{Sum Mean} = \frac{1}{2} \sum_i \sum_j (i \mathbf{P}(i, j) + j \mathbf{P}(i, j)) \quad (5.11)$$

$$\text{Inertia} \sum_i \sum_j (i - j)^2 \mathbf{P}(i, j) \quad (5.12)$$

$$\text{Cluster Shade} \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 \mathbf{P}(i, j) \quad (5.13)$$

$$\text{Cluster Tendency} \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 \mathbf{P}(i, j) \quad (5.14)$$

$$\text{Homogeneity} = \sum_i \sum_j \frac{\mathbf{P}(i, j)}{1 + |i - j|} \quad (5.15)$$

$$(5.16)$$

$$\text{Max Probability} = \max_{i,j} P(i,j) \quad (5.17)$$

$$\text{Inverse Variance} = \sum_i \sum_j \frac{P(i,j)}{(i-j)^2} \quad (5.18)$$

where μ_i , μ_j , σ_i and σ_j are the column and row-wise mean and variance respectively. These features measure things such as the randomness of the grey-level distribution (entropy), the number of repeated pairs (energy), the local contrast or homogeneity of the image, variance, the tendency to form clusters (cluster shade and tendency), among others.

For this work we have used a distance d ranging from 1 to 10, at each of the 13 spatial directions. Therefore, we have computed $13 \times 10 = 130$ GLCMs per image, from which 12 texture features are computed. Our final feature vector will therefore have 1560 features in total.

To further reduce the dimensionality of the feature vector, we have performed feature selection using the t-test, MWU U-test and the relative entropy (KL divergence) criteria (see Section 4.1).

5.2.3 Experiments

To evaluate the system proposed in this chapter, combining texture analysis and other feature selection algorithms, we propose two experiments:

- Experiment 1: Ability of the different texture features to differentiate between PD affected subjects and CTLs. Each of the texture features is analysed in two different ways: a "single approach", which only considers one type of feature using only the matrices at a distance d from the central voxel -using all the spatial directions- and a "cumulative approach" which considers one type of feature too, but this time using all matrices in distances ranging from 1 to d .
- Experiment 2: Impact of the introduction of a feature selection algorithm (of those presented at Section 4.1) after computing the texture features. This allow us to pool all texture features at all distances and directions, and then select the most discriminative ones according to any of these criteria.

All images are intensity normalized using either normalization to the maximum or integral normalization (see Section 3.2), and afterwards, a subvolume can be extracted using the intensity threshold methodology described at Section 5.2.1. In addition to the feature extraction technique using texture analysis, and the feature selection procedure defined for Experiment 2, we use a

linear SVC for classifying, and 10-fold cross validation strategy (see Section 3.3 for more details). Note that for this analysis we will remove the SWEDD subjects from the PPMI-DAT and VDLN-DAT datasets.

5.3 Results

5.3.1 Experiment 1: Individual Texture Features

In this experiment, the influence and effect of each texture feature is tested, as in [34]. We have tested the computation of the GLCMs over the image sub-volumes using different thresholds I_{th} (see Sec. 5.2.1) ranging from 0 to 50% of the maximum intensity value, and a range of distances $d = 1, 2, \dots, 10$ in the thirteen spatial directions.

To check which value of the intensity threshold is the best for computing the texture features, we can compute the general tendency of the system by averaging the accuracy values. Figure 5.3 depicts the general trend of the performance over the intensity threshold for either the no normalized or normalized images. This is done for the single and cumulative approach.

The most remarkable differences are found between normalization procedures. As a general trend, the integral normalization barely has any impact over the performance achieved with the registered images. Furthermore, the normalization to the maximum even drops the performance for high I_{th} , however it increases the general performance in the range 0.1-0.3. From these graphs it is patent that normalization has little impact on the performance achieved by our system, which can be consider an advantage, since it reduces the preprocessing needed.

In this regard, the VDLV-DAT dataset has an strange behaviour. It performance holds and even increases when using normalization to the maximum, but significantly drops when integral normalization is used. This is exactly the opposite to the other datasets, and will be discussed later.

In these images, we can observe a strong dependence of the system's performance with the intensity threshold I_{th} . In general, the performance increases with a more restrictive threshold (a smaller box around the striatum). This increase is probably due to removing the background from the texture analysis. According to Figure 5.2, a value between 0.30 a 0.35 could be indicative of a complete background removal from the computation. In most cases, $I_{th} \approx 0.35 \times I_{max}$ seems to be a critical value: either the global maximum or the inflection point from which accuracy stabilizes. This would prove

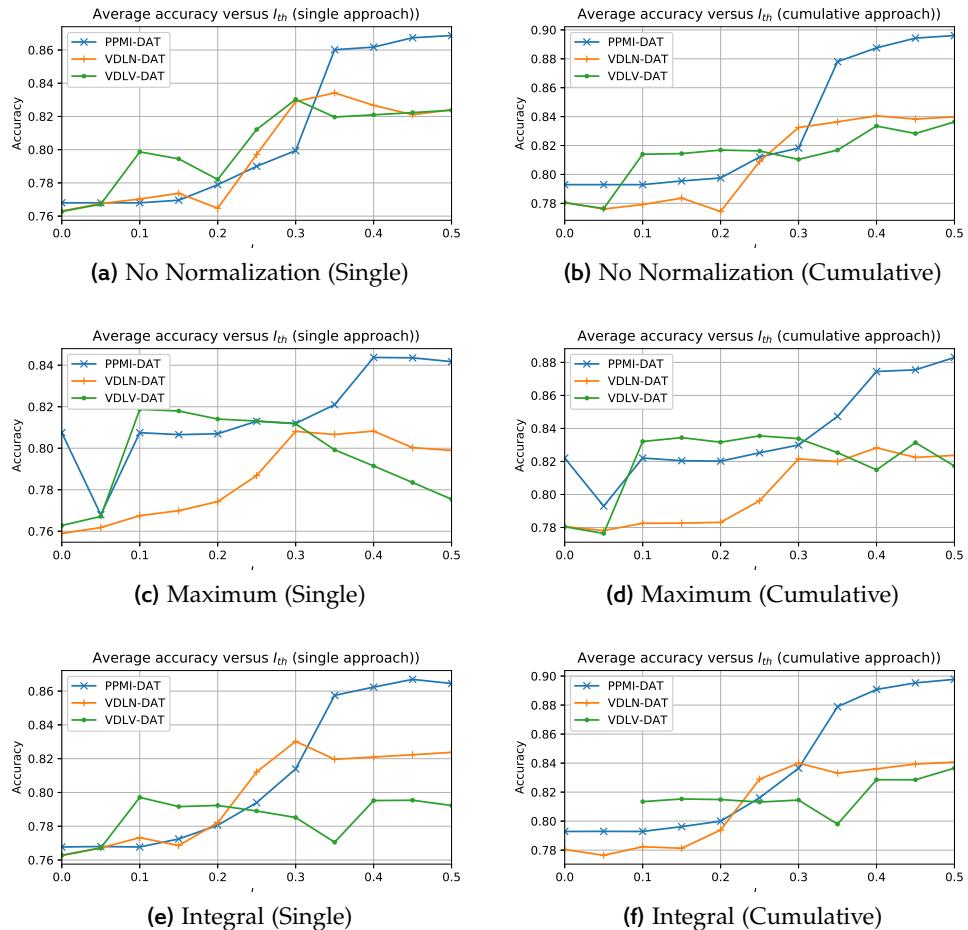


Figure 5.3: Evolution of the average accuracy values obtained for the single approach and the cumulative approach over the intensity threshold, using no normalization, normalization to the maximum and integral normalization.

that eliminating the background is beneficial for the texture analysis and a posterior classification of the images.

To obtain a deeper insight on the behaviour of each texture feature, we can use a violin plot. This plot is an evolution of the boxplot, frequently used in statistical distributions, in which the distribution of values is shown along with the mean and standard deviation of the set of given data. In our case, we display the violin plot of all accuracy values for each database, grouped by texture feature, at Figure 5.4. Since the performance of the integral normalization was similar to applying no normalization at all, we have only used data for the normalization to the maximum and the original images.

Figure 5.4 shows the differences between applying or not the normalization to the maximum (in colors), and also the differences in performance obtained depending on the database. We can observe that in average (the white dot), cluster tendency is the best performing feature. homogeneity, contrast and correlation also achieve good results. This behaviour is consistent along all three databases, from which we can consider cluster tendency the most discriminant feature for PD patterns.

In Tables 5.1 and 5.2, we take a look at the performance achieved by the maximum scoring feature in all databases with different normalization procedures. This is shown for both the single and the cumulative approach.

	Norm.	Feature	I_{th}	d	acc.	sens.	spec
PPMI-DAT	Original	Cluster Tendency	45	8	0.952	0.946	0.956
	Integral	Cluster Tendency	45	4	0.952	0.946	0.956
	Maximum	Cluster Tendency	45	8	0.948	0.955	0.943
VDLN-DAT	Original	Cluster Tendency	40	1	0.941	0.956	0.932
	Integral	Cluster Tendency	35	3	0.941	0.978	0.918
	Maximum	Inverse Variance	45	8	0.923	0.933	0.918
VDLV-DAT	Original	Cluster Tendency	35	7	0.941	0.978	0.918
	Integral	Cluster Tendency	30	6	0.904	0.889	0.920
	Maximum	Cluster Tendency	35	1	0.923	0.907	0.940

Table 5.1: Maximum scoring feature for each combination of database and normalization procedure, and the intensity threshold and offset distance for which this maximum performance is achieved, using the single approach.

The first noticeable feature is that, for both the single and cumulative approaches, the best scoring normalization method is using no normalization at all. Secondly, as anticipated, the best scoring feature is cluster tendency in

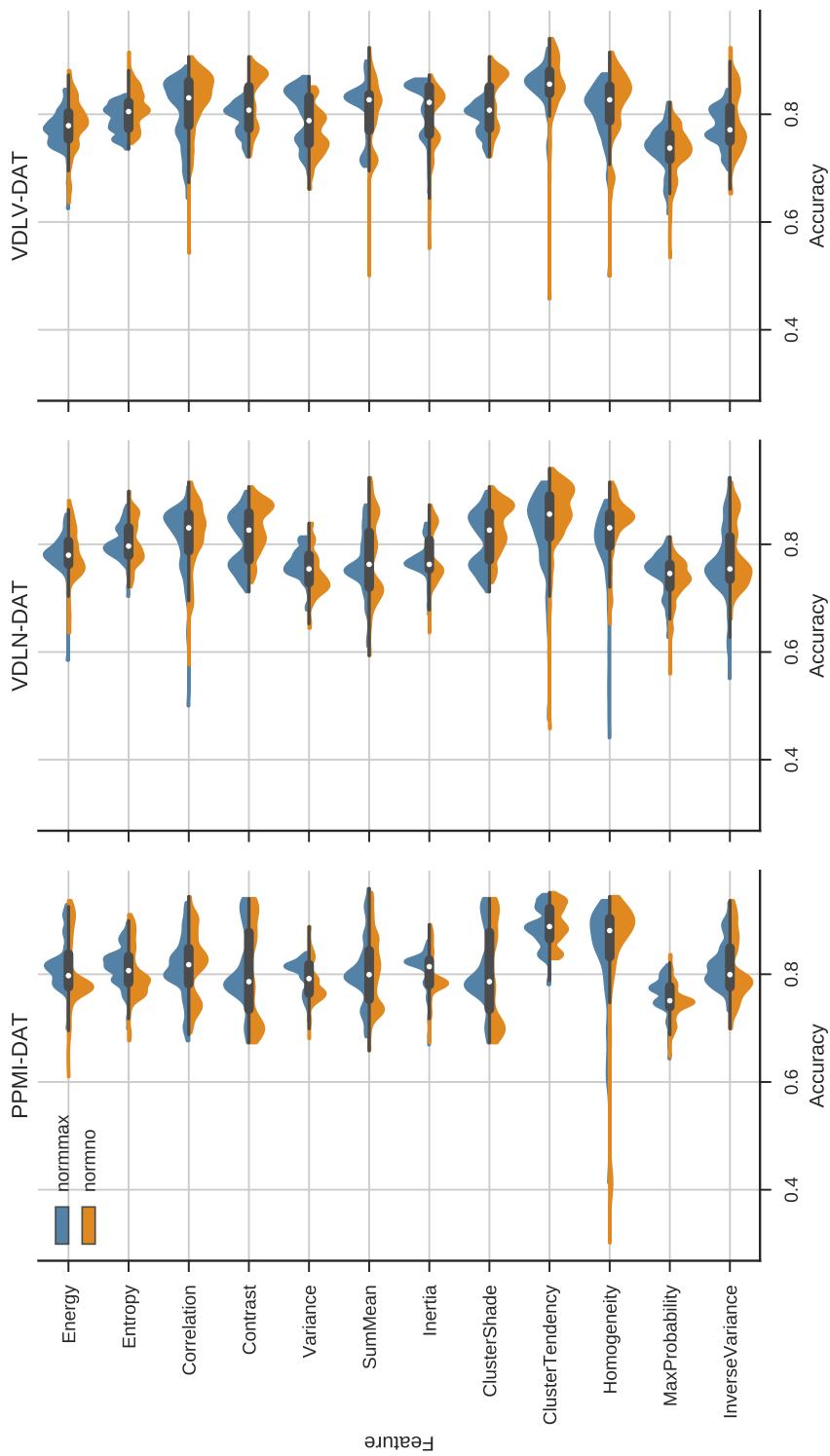


Figure 5.4: Violin plot of all accuracy values, grouped by database and showing the differences between normalization to the maximum and the original images.

	Norm.	Feature	I_{th}	d	acc.	sens.	spec
PPMI-DAT	Original	Cluster Tendency	45	6	0.970	0.982	0.962
	Integral	Cluster Tendency	40	5	0.966	0.982	0.956
	Maximum	Cluster Tendency	45	5	0.966	0.982	0.956
VDLN-DAT	Original	Cluster Tendency	45	7	0.966	1.000	0.945
	Integral	Cluster Tendency	45	7	0.966	1.000	0.945
	Maximum	Cluster Tendency	50	7	0.958	0.978	0.945
VDLV-DAT	Original	Inertia	50	7	0.918	0.963	0.870
	Integral	Inertia	50	6	0.918	0.963	0.870
	Maximum	Cluster Tendency	45	7	0.918	0.944	0.890

Table 5.2: Maximum scoring feature for each combination of database and normalization procedure, and the intensity threshold and offset distance for which this maximum performance is achieved, using the cumulative approach.

most cases, and the systems perform better when the intensity threshold is more restrictive.

When comparing the single approach with the cumulative one, it is obvious that the best results are obtained with the cumulative one, except for the particular case of the VDLV-DAT. Whereas with the single approach there was no evident choice for the offset distance d, in the cumulative one, the best results are obtained combining the first 5-7 distances at which the GLCM was computed. This is reasonable, since for the single approach we only use the features contained at each d, while in the cumulative approach we pool much more information for training and testing the system. Nevertheless, the single approach obtains decent results in many cases, which proves the value of the Haralick texture features for characterizing DaTSCAN images.

5.3.2 Experiment 2: Selected Texture Features

In experiment 2 we pool together all features derived from all GLCMs, and use a feature selection algorithm (see Section 4.1). This gives us 1560 features per subject, from a total 12 features computed at 13 directions and 10 distances. We will test how values such as I_{th} , the normalization algorithm or the percentage of selected voxels affects the performance of this system, and discuss the results.

In Figure 5.5 we show how the performance evolves when varying the I_{th} , as we did in Experiment 1, by displaying the average accuracy. As in the

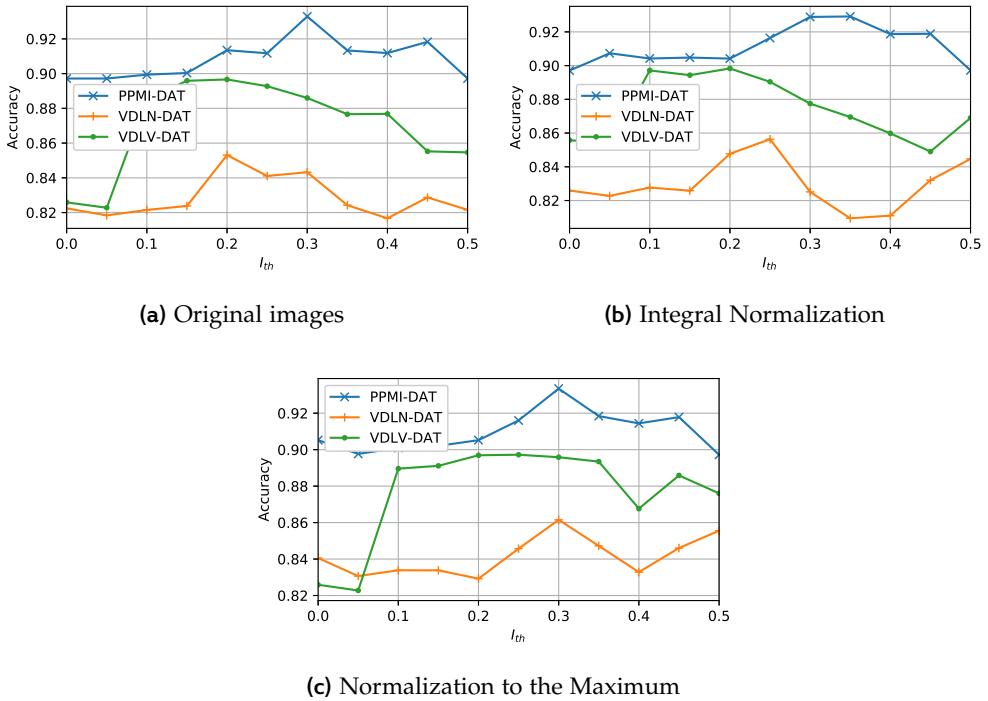


Figure 5.5: Accuracy obtained by averaging all accuracy values using a given volume selection threshold I_{th} .

previous case, the integral normalization performs similarly to the original, non-normalized images. In this case, the normalization to the maximum strategy seems to affect more the VDLV-DAT dataset, slightly improving its performance. Again, best I_{th} seems to be located at approximately 0.30 for all databases when using normalization to the maximum, although for other approaches, this happens with smaller I_{th} (0.20 for VDNL-DAT and VDLV-DAT original images). For these values, there are still background voxels included in the computation of textures, but its negative influence might be overcome by the feature selection procedure.

These results corroborate that our volume selection strategy is profitable in almost any case, using a intensity threshold between 0.25 and 0.45.

Regarding the different selection criteria, Fig. 5.6 analyses the behaviour of our system under the three proposed criteria. For this purpose, we average all accuracy values for a given proportion of selected features (from 1% to 100% of the 1560 total texture features, previously ranked according to the selection criteria), using I_{th} values ranging from 0.1 to 0.45.

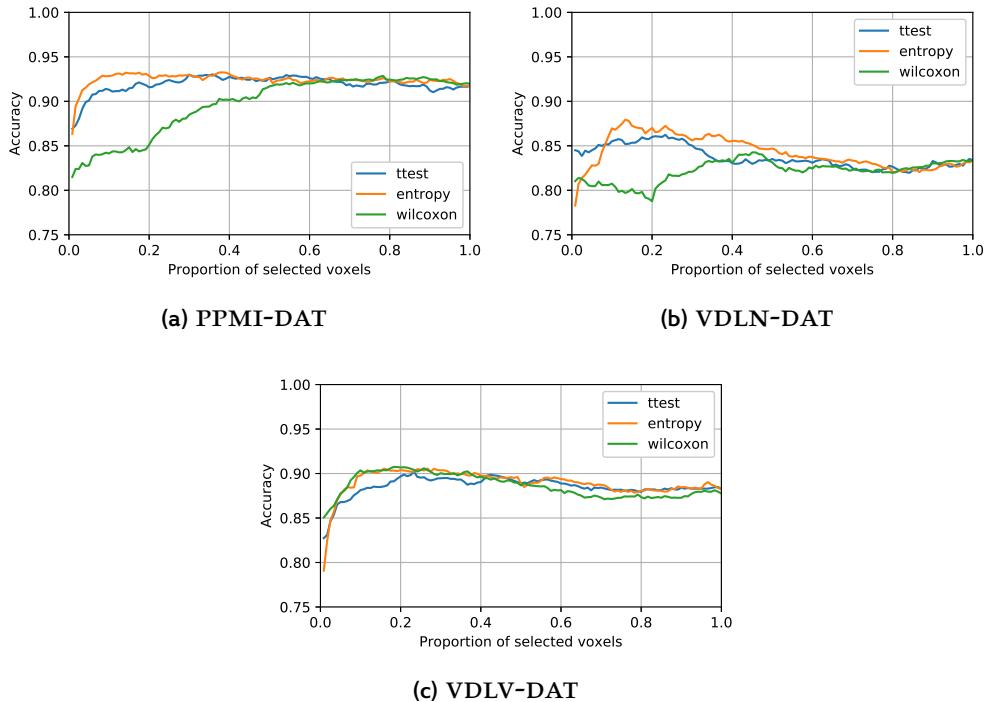


Figure 5.6: Average accuracy computed for each selection criteria, using all accuracy values for intensity thresholds of 0.10 to 0.45, compared to the proportion of selected features for the PPMI-DAT, VDNL-DAT and VDLV-DAT databases.

From these figures we can infer that either the t-test or the relative entropy criteria perform better than the MWW U-test. Of these two, the relative entropy seems to perform the best in the PPMI-DAT and VDNL-DAT datasets, whereas when applied to the VDLV-DAT, the wilcoxon achieves similar results to the relative entropy. In these cases, the relative entropy criterion obtains its maximum average accuracy using the first 10% of features, while wilcoxon needs more than 50% of the features, which makes it less efficient.

After describing the general behaviour, we can take another look at the different combinations of normalization, datasets and selection criteria. In Table 5.3, the peak results of these combinations are shown, including the percentage of selected features needed to achieve this value.

This table confirms that best values are obtained generally by the relative entropy criterion, with few exceptions. It also confirms that best performance is achieved when using the volume extraction method proposed in Sec. 5.2.1,

	Norm.	Selection	I_{th}	acc.	sens.	spec .	%
PPMI-DAT	Orig.	entropy	30	0.970	0.972	0.968	0.983
		t-test	30	0.966	0.972	0.962	0.966
		wilcoxon	30	0.959	0.954	0.962	0.858
	Int.	entropy	25	0.966	0.981	0.955	0.308
		t-test	25	0.973	0.990	0.962	0.358
		wilcoxon	35	0.947	0.954	0.943	0.983
VDLN-DAT	Orig.	entropy	25	0.966	0.981	0.955	0.308
		t-test	25	0.973	0.990	0.962	0.358
		wilcoxon	30	0.959	0.954	0.962	0.858
	Int.	entropy	30	0.932	0.933	0.931	0.175
		t-test	30	0.940	0.955	0.931	0.175
		wilcoxon	15	0.898	0.955	0.863	0.433
VDLV-DAT	Orig.	entropy	25	0.932	0.977	0.904	0.100
		t-test	25	0.915	0.955	0.890	0.100
		wilcoxon	25	0.915	0.911	0.917	0.966
	Max.	entropy	20	0.923	1.000	0.876	0.233
		t-test	30	0.932	0.933	0.931	0.225
		wilcoxon	45	0.932	0.933	0.931	0.033
VDLV-DAT	Orig.	entropy	35	0.937	0.935	0.940	0.133
		t-test	35	0.932	0.953	0.910	0.350
		wilcoxon	20	0.927	0.916	0.940	0.383
	Int.	entropy	20	0.937	0.935	0.940	0.983
		t-test	10	0.932	0.907	0.960	0.508
		wilcoxon	10	0.937	0.935	0.940	0.966
VDLV-DAT	Max.	entropy	20	0.937	0.953	0.920	0.608
		t-test	35	0.932	0.935	0.930	0.341
	Max.	wilcoxon	20	0.932	0.925	0.940	0.141

Table 5.3: Best results of the experiment 2 per database, normalization and selection criteria. The I_{th} and percentage of selected features (of the total 1560 features computed) at which each value is obtained is also shown for comparison.

with values of I_{th} between 0.25 and 0.30. As for the datasets, we see that PPMI-DAT yields the best performance in all cases, with all selection criteria, whereas, contrary to their average performance examined in Figure 5.5, VDLN-DAT and VDLV-DAT can actually perform similarly.

The application of intensity normalization algorithms does not imply an improvement in performance, and most systems achieve good accuracy without any normalization at all. However, there is only one benefit that can be inferred from the table, which is a decrease in number of selected features needed to achieve best performance with the PPMI-DAT. In this case, and partially in the case of VDLV-DAT as well, similar performance is obtained using either normalized or original images, but when using the normalized data, the percentage of selected features is smaller. Nevertheless, this does not hold for the VDLN-DAT, therefore we cannot consider it a general behaviour.

The choice of a best selection method is here a matter of trade-off between the computer performance (the number of features to estimate) and the accuracy needed. In the case of the PPMI-DAT, it is patent that using normalized images and the t-Test selection criterion achieves best performance. For VDLN-DAT it is more difficult to assure that any option will perform better than others, although the t-test still performs well with the original images. Finally, with the VDLV-DAT dataset, the preferred method will be again to use the original images, since the selection of features seems more optimal, especially when using relative entropy. Anyway, all the options reveal the ability of our system in the PD detection with an relevant performance (over 90% of accuracy in most cases).

5.4 Discussion

The system proposed in this chapter was published in [25, 34], and fundamentally defines a new pipeline for the CAD of PD based on texture analysis. It combines intensity normalization, a subvolume extraction algorithm, texture analysis and classification via SVC.

The application of a intensity normalization procedure was proved fundamental for feature extraction methods such as VAF [45], Singular Value Decomposition (SVD) [49] or PLS [38]. These methods strongly rely on the absolute intensity values found at each anatomical position. Conversely, texture analysis depends on the computation of the GLCM, which quantifies voxel relations. When analysing the behaviour of our system in both experiments, we can infer that normalizing the images does not pose any major improvement over using the original images themselves.

The slight differences between using or not normalization are variable. We find that the type of normalization depends on the database used, and no general trend could be stated. For example, VDLV-DAT has a preference for the normalization to the maximum, whereas the PPMI-DAT performs better when using integral normalization. There were even more cases where normalization decreased performance than those in which it improved the system. Therefore, we can assume that it yields no benefit at all, and therefore, our system can be intensity-independent, which indeed is considered a big advantage.

Now, regarding the volume selection algorithm, it has proved beneficial in almost any case. This can be due to a variety of reasons. Firstly, the optimum sub-volume (for an intensity threshold around $0.30I_{\max}$) has a size smaller than $40 \times 40 \times 50$. The maximum value at which we computed the GLCMs was $d = 10$, which correspond to at least a 20% of the subvolume selected with the previous value. Since the voxel size of the images in the three databases is approximately $2 \times 2 \times 2\text{mm}$, the maximum textural changes that we could analyse were computed at a distance of 20mm , approximately half the size of the striatum. This is more than enough to characterize texture in these noisy images, since lower frequency -smoother- textural changes can be obviated for diagnosis, enhancing the descriptive ability of the texture features.

Secondly, the descriptive ability of the texture features is also enhanced with the removal of the background introduced by this subvolume extraction algorithm. Within the subvolume, we can ensure that the texture changes only correspond to real changes in the dopamine distribution, and no to the contrast between the brain and the background. And thirdly, computing the texture features over a subvolume is always faster than over the whole brain, which makes our system faster.

The features that best describe the texture of DaTSCAN images are analysed in Experiment 1. In Fig. 5.4, we saw that features such as homogeneity, sum mean and, overall, cluster tendency, achieve the best performance of our system in either the single and the cumulative approach, in all three datasets. Since cluster tendency measures the grouping of voxels with a similar grey-level, it is obviously a good descriptor of the striatum shape, where most of the intensities of the images is concentrated. Higher values of cluster tendency can be associated with CTL, whereas lower values can be related to people affected by dopaminergic deficit.

The introduction of a feature selection algorithm using hypothesis testing improved the performance in the three datasets. Thanks to this, we can take best features according to a criterion, and use all of them regardless of what they measure. We would assume that most criteria would select fundamen-

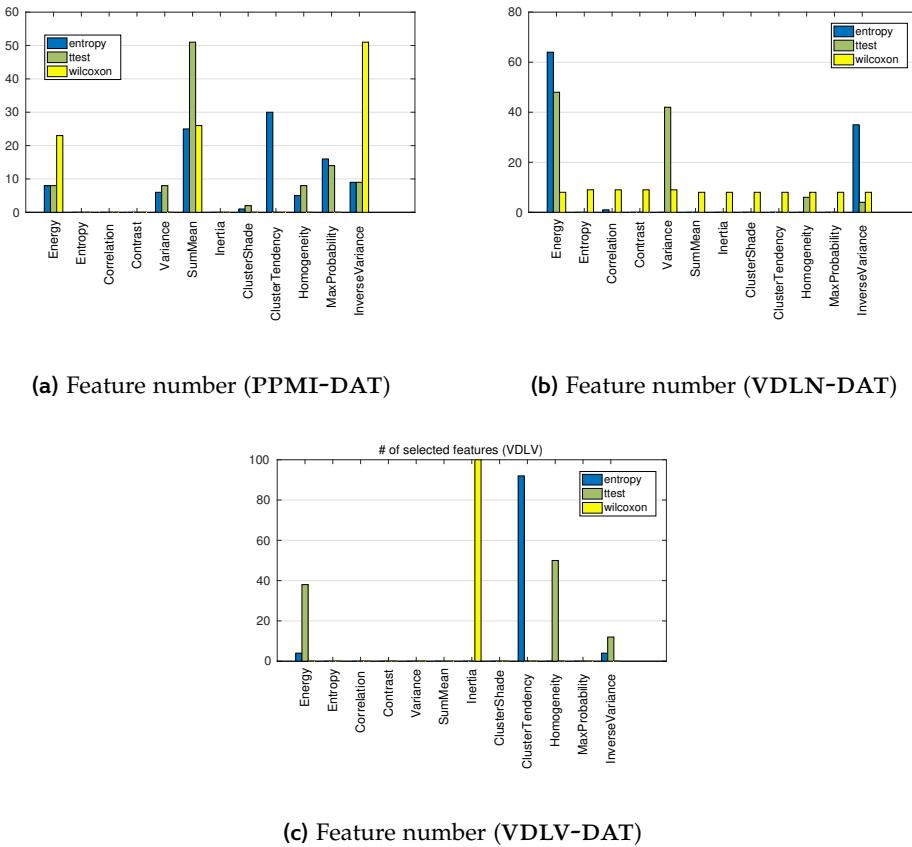


Figure 5.7: Distribution of the 100 first selected features by means of the different selection methods (using the $I_{th} = 0.35I_{max}$) for PPMI-DAT, VDLN-DAT, and VDLV-DAT database.

tally cluster tendency, Homogeneity and other high performance measures. In Figure 5.7 we count how many features of each type are in the 100 first ranked using each criterion and dataset. We see that the best-performing features such as cluster tendency, are not always selected, especially with the VDLN-DAT dataset, which could be responsible for its lower overall performance. It reveals that each database has internal characteristics, for example, the discarded cuts in VDLN-DAT, that are better modelled using other features, what could lead to the apparition of outliers in the computation, and a decrease in performance. However, the selection approach allow us to overcome the individual characteristics of each dataset, making the system extensible to other uses in clinical practice.

Finally, we will compare our proposed system with other methods used in PD diagnosis in the literature. We will compare with the baseline VAF from [45], and two additional methods. These are an asymmetrical SVD [49] that applied SVD on both sides of the brain (since PD often appears only in one hemisphere), and a Empirical Mode Decomposition (EMD) [38] using different Independent Mode Functions (IMF), particularly the IMF-3. These systems are compared with either the cumulative approach and the system of experiment 2 with different criteria in Table 5.4.

System	Acc	Sens	Spec
SumMean	0.951	0.972	0.936
Homogeneity	0.944	0.946	0.943
Cluster Shade	0.940	0.936	0.943
Cluster Tendency	0.952	0.946	0.956
Energy	0.936	0.954	0.924
Correlation	0.944	0.963	0.930
Entropy	0.970	0.972	0.968
t-test	0.966	0.972	0.962
Wilcoxon	0.959	0.954	0.962
VAF	0.840	0.807	0.862
VAF-IN	0.913	0.890	0.932
SVD	0.940	0.962	0.918
EMD-IMF3	0.950	0.951	0.948

Table 5.4: Comparison between our proposed system and other PD diagnosis systems in the literature: a VAF system using the intensity-normalized images, a combination of intensity normalization strategies and classifiers (VAF-IN) [45], a SVD-based approach [49] and EMD using the third independent mode function (IMF3) [38].

In this table, we compare the values at the operation point for Experiment 1 and 2 with other methods in the literature. In Experiment 1, values for six texture features, such as sum mean, homogeneity or cluster tendency are shown (using the single approach), and values for the three selection criteria are shown. The values of using only one texture feature match those obtained by state of the art methods like the ones proposed in [38, 49], whereas the methodology used in Experiment 2 outperform all previously used methods. This holds for all feature selection criterion, proving the ability of the Haralick texture analysis to detect PD patterns in DaTSCAN imaging.

6

SPHERICAL BRAIN MAPPING

6.1 Introduction

In this section we will present a novel feature extraction technique called Spherical Brain Mapping (SBM). SBM is based on the use of spherical coordinates to extract radial features from structural MRI images. Using the features at each coordinate, we can characterize the texture in each direction and project the information to a bidimensional map, which provides a significant feature reduction and a visual aid for diagnosis.

The most basic form is the standard SBM [6, 23], in which all voxels crossed by a rectilinear vector in a spherical coordinate pair (θ, φ) are selected, and then, a certain measure is extracted from that set. In this sense, statistical and morphological measures such as tissue thickness, average or entropy, among others, are computed.

Further improvements can be made to this simple approach, for example, with the layering extension [6], in which the mapping vector is divided in n subsets containing the same number of voxels. Therefore, instead of a single map, we can obtain n maps at different distances from the centre of the brain. Another useful approach is the characterization of texture features via Local Binary Patterns (LBP), computed around the mapping vector, which yielded very good results in [14].

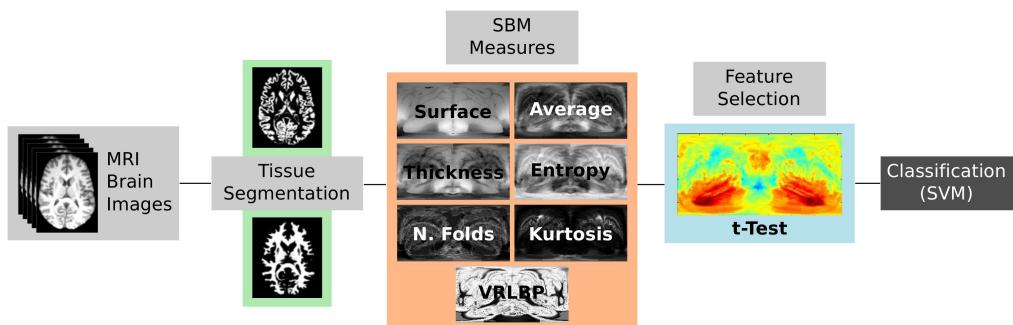


Figure 6.1: Flow diagram of the procedure used in the textural analysis of projected MR brain images.

The most relevant extension to the SBM was proposed in [7]. In this work, instead of using rectilinear vectors to select voxels, we developed a path tracing algorithm that follows a minimum-intensity change pathway towards an attractor placed in its corresponding spherical coordinate pair (θ, φ) . This way, the mapping paths follow the structural features of the MRI image, which could be used to create the bidimensional SBM maps as well as directly use the intensity distribution of the voxels along the path. This was extended to make use of the GLCM and Haralick texture analysis (see Chapter 5) to characterize the brain texture in each path and its neighbourhood.

6.2 Spherical Brain Mapping (SBM)

The original SBM proposed in [6, 23] was based on the use of spherical coordinates in the brain. The central voxel is used as the origin point from which a number of mapping vectors $\mathbf{v}_{\theta,\varphi}$ are defined for each inclination (θ) and azimuth (φ) angles in the range $0^\circ < \theta < 180^\circ$ and $0^\circ < \varphi < 360^\circ$ (see Figure 6.2). The voxels crossed by this mapping vector are selected, to form the sampled set $V_{\theta,\varphi}$, a set that contains the intensities of the P voxels crossed by the mapping vector $\mathbf{v}_{\theta,\varphi}$.

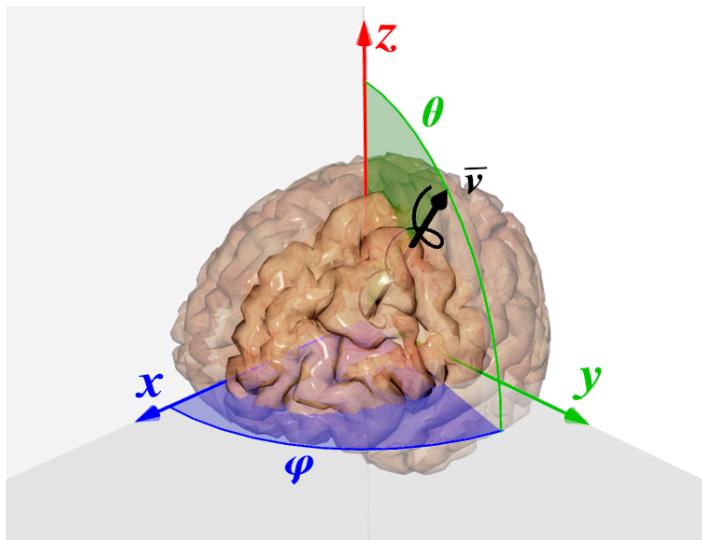


Figure 6.2: Illustration of the computation of the mapping vector $\mathbf{v}_{\theta,\varphi}$, the angles θ and φ and the r -neighbourhood of \mathbf{v} (see Section 6.2.2).

The basic form of SBM computes a mapping value v from each set $V_{\theta,\varphi}$ at each coordinate pair (θ, φ) . In [6, 23], six basic measures were proposed:

- A basic brain surface approach, which characterizes the surface of either GM or WM tissue. It accounts for the distance between the origin and the last tissue voxel in $V_{\theta,\varphi}$ greater than a threshold I_{th} . This might correlate with structural neurodegeneration and tissue loss in the surface of the tissue.

$$v_{surf} = \arg \max_i \{V_{\theta,\varphi}(i) > I_{th}\} \quad \forall i = 1, \dots, P \quad (6.1)$$

- The thickness of the tissue. It is defined as the distance between the last and first elements in $V_{\theta,\varphi}$ with an intensity greater than a threshold I_{th} (typically 0). This can be useful when measuring the thickness of segmented GM or WM maps, and, although less powerful than other implementations like Freesurfer's [142], it might be representative enough and easier to compute :

$$v_{thick} = \arg \max_i \{V_{\theta,\varphi}(i) > I_{th}\} - \arg \min_i \{V_{\theta,\varphi}(i) > I_{th}\} \quad \forall i = 1, \dots, P \quad (6.2)$$

- The number of folds represents the number of overlapping segments of tissue in the set $V_{\theta,\varphi}$. It is computed by counting the number of connected subsets in a thresholded $V_{\theta,\varphi}$ using the value I_{th} . Let $A_{\theta,\varphi}$ be the set that contains all the indices of the voxels in $V_{\theta,\varphi}$ with an intensity greater than I_{th} :

$$A_{\theta,\varphi} = \{i / V_{\theta,\varphi}(i) > I_{th}\} \quad (6.3)$$

where $A_{\theta,\varphi} \in \mathbb{N}$. Let us divide $A_{\theta,\varphi}$ in J disjoint connected subsets so that:

$$A_{\theta,\varphi} = A_{\theta,\varphi}^1 \cup A_{\theta,\varphi}^2 \cup \dots \cup A_{\theta,\varphi}^J \quad \text{so that} \quad A_{\theta,\varphi}^i \cap A_{\theta,\varphi}^j = \emptyset \quad \forall i, j \quad (6.4)$$

Therefore, our $v_{nf} = J$, the number of disjoint connected subsets in $A_{\theta,\varphi}$.

- The average of $V_{\theta,\varphi}$, which is related to tissue density:

$$v_{av} = \frac{1}{N} \sum_i V_{\theta,\varphi}(i) \quad \forall i = 1, \dots, P \quad (6.5)$$

- The entropy of $V_{\theta,\varphi}$, assuming it is a probability mass vector (probability of belonging to a certain tissue, normalized). It computes v as:

$$v_{ent} = \sum_i V_{\theta,\varphi}(i) * \log(V_{\theta,\varphi}(i)) \quad \forall i \in \arg_i \{V_{\theta,\varphi}(i) > 0\} \quad (6.6)$$

- The uncorrected kurtosis, also known as fourth standardized moment, of the set $V_{\theta,\varphi}$ in which v is calculated using:

$$v_{kurt} = \frac{\frac{1}{N} \sum_i (V_{\theta,\varphi}(i) - \bar{V}_{\theta,\varphi}(i))^4}{\left(\frac{1}{N} \sum_i (V_{\theta,\varphi}(i) - \bar{V}_{\theta,\varphi}(i))^2\right)^2} \quad \forall i = 1, \dots, P \quad (6.7)$$

where $\bar{V}_{\theta,\varphi}$ is the average of all voxels in $V_{\theta,\varphi}$ (same value as v_{av} , described in Eq. 6.5).

We can compute each of these six maps over the GM or WM tissue maps of a segmented MRI. Examples of these are depicted in Figure 6.3. In these maps, the value v computed at each direction (θ, φ) is represented, where the azimuth φ is represented in the x-axis, from 0° to 360° and the inclination angle θ in the y-axis, from 0° to 180° . The algorithm that produces these maps can be downloaded at <http://pakitochus.github.io/mapBrain/> (versions for MATLAB and Python are available).

This methodology defines the sampling set as the voxels crossed by the sampling vector $v_{\theta,\varphi}$. This implies a loss of information on the neighbourhood of $v_{\theta,\varphi}$ that increases with the distance to the origin. To overcome this problem, two different extensions have been suggested. In the first one, the sampled set $V_{\theta,\varphi}$ is divided in n equal parts, and one map is computed for each of the n parts, in the “Layered approach”. A second approach uses Local Binary Patterns (LBP) and helical sampling to map the neighbourhood of $v_{\theta,\varphi}$ and characterize texture. Finally, we will define new paths that adapt to the intensity changes of the brain images, using a HMM based approach.

6.2.1 Layered Extension

The layered extension is the simplest approach to keep relevant information of the different “layers” of tissue in our SBM maps. To do so, we divide each sampled set $V_{\theta,\varphi}$ in n equal subsets, from which n maps will be derived. For example, with a $n = 4$, 4 subsets will be used to compute 4 different maps at different distances from the origin, from the closest to the farthest. We assume that this approach features more detail, since overlapping structures placed at different depths will be contained within different maps.

6.2.2 Volumetric Radial LBP (VRLBP)

Another improvement can be made to the original SBM: the inclusion of the r -neighbourhood of the mapping vector $v_{\theta,\varphi}$ in the computation of v . We do so

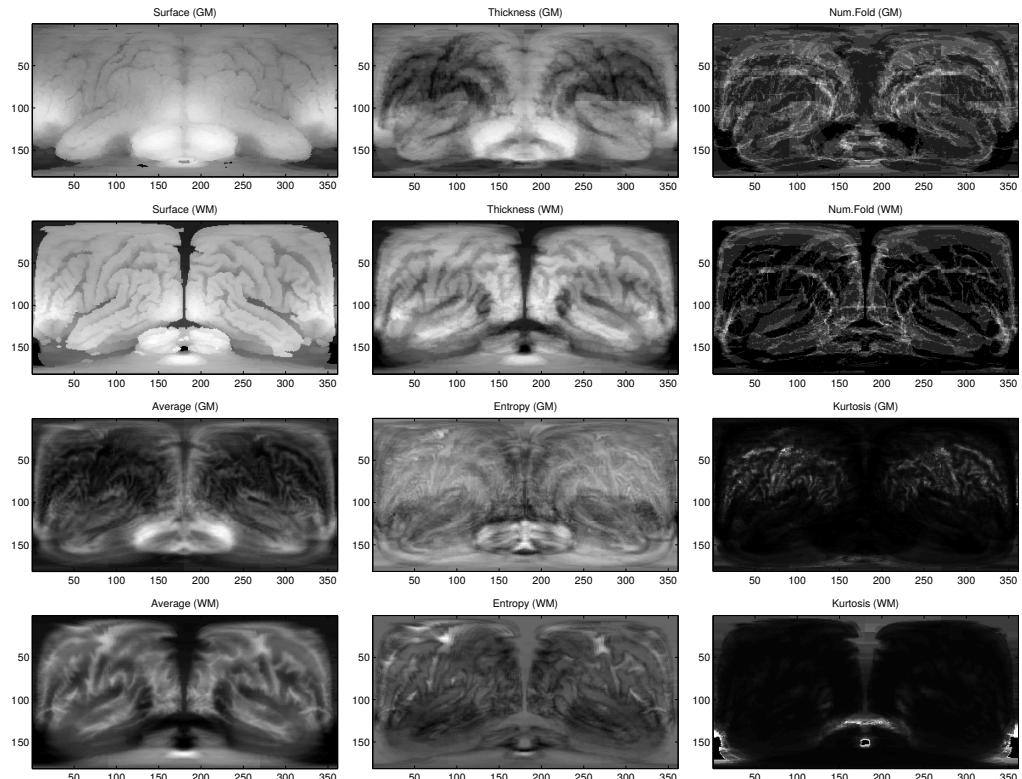


Figure 6.3: Resulting GM and WM maps of the same control subject using the six proposed measures: surface, thickness, number of folds, average, entropy and kurtosis.

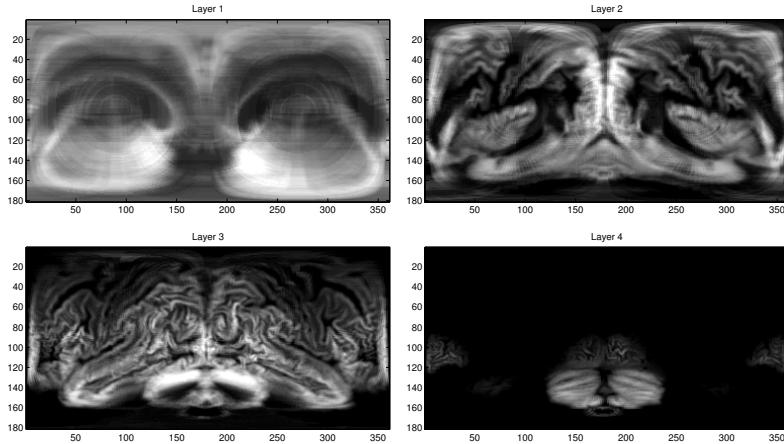


Figure 6.4: Example of the layered approach using the average measure on GM maps. Some internal GM structures such as the Putamen or Globus Pallidus can be identified at Layer 2 (see anatomical reference at Figures 6.7 and 6.8).

by computing the Volumetric Radial LBP (VRLBP), based on the LBP descriptors proposed in [155].

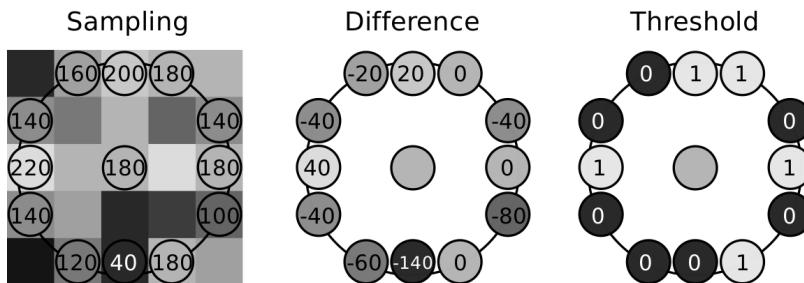


Figure 6.5: Example of how the basic LBP is computed.

LBP was devised to describe the texture of a image, with an initial application to face recognition. In its basic form, it consist of three steps: sampling, calculating the difference and thresholding (See Figure 6.5). The value of the LBP is defined as:

$$v_{LBP} = \sum_{p=0}^{P-1} s(I_p - I_c)2^p \quad (6.8)$$

where P is the number of neighbours at a distance r of the central voxel, and I_p and I_c are the intensities of the p^{th} voxel and the central voxel for which

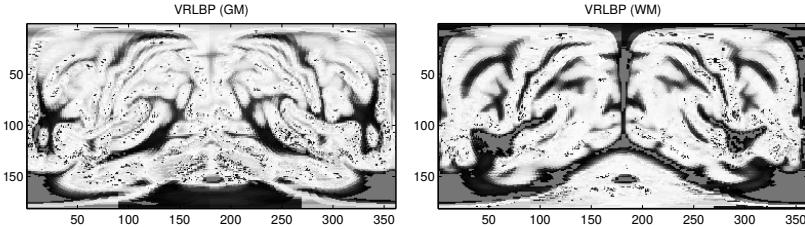


Figure 6.6: An example of the VRLBP projection for GM and WM Tissues.

the value v_{LBP} is being computed. The threshold step is performed using the sign function $s(x)$, defined as:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (6.9)$$

This basic approach was extended to Volumetric LBP [103], in which a 3D texture is defined in a local neighbourhood using a cylinder of radius r oriented in one direction. For this work, we updated the sampling procedure proposed in [103] using a helix around the mapping vector $\mathbf{v}_{\theta,\varphi}$ (see Figure 6.2). This new helical sampling of [14] defines the set of P sampled voxels on the image I using a r -neighbourhood $V_{\theta,\varphi}^{P,r}$ as:

$$V_{\theta,\varphi}^{P,r} = \{I(\mathbf{g}_{\theta,\varphi}^{0,r}), I(\mathbf{g}_{\theta,\varphi}^{1,r}), I(\mathbf{g}_{\theta,\varphi}^{2,r}), \dots, I(\mathbf{g}_{\theta,\varphi}^{P-1,r})\} \quad (6.10)$$

where the coordinate vector $\mathbf{g}_{\theta,\varphi}^{p,r}$ of each voxel are computed in the direction of $\mathbf{v}_{\theta,\varphi}$ by:

$$\mathbf{g}_{\theta,\varphi}^{p,r} = \begin{cases} x_{\theta,\varphi}^{p,r} = p \sin(\varphi) \cos(\theta) - r \sin(2\pi t p / P) \\ y_{\theta,\varphi}^{p,r} = p \sin(\varphi) \sin(\theta) + r \cos(2\pi t p / P) \\ z_{\theta,\varphi}^{p,r} = p \cos(\varphi) \end{cases} \quad p = \{0, \dots, P-1\}, P \in \mathbb{N} \quad (6.11)$$

being t the number of turns in the helical sampling. We use linear interpolation to estimate the intensities in positions that do not fall exactly at the coordinates computed in Eq. 6.11, as in [103].

If we fix P and r to constant values, the set of sampled voxels $V_{\theta,\varphi}^{P,r}$ becomes $V_{\theta,\varphi}$, which is equivalent to the definition of SBM found in Section 6.2. The value v of the VRLBP approach is therefore defined as:

$$v_{\text{VRLBP}} = \sum_p s(V_{\theta,\varphi}(p) - V_{\theta,\varphi}(0)) \cdot 2^p \quad \forall p = 1, \dots, P \quad (6.12)$$

The resulting texture maps for GM and WM tissues can be found at Figure 6.6.

6.2.3 Anatomical Reference

To better understand the SBM maps, and the location of different features, we have projected the widely known Automated Anatomical Labeling (AAL) atlas [131] using SBM. That way, we have an anatomical reference of the different structures and their position in the different coordinate pairs (θ, φ) . The regions are displayed in Figures 6.7 and 6.8.

6.3 Sampling Paths via Hidden Markov Models (HMMs)

The rectilinear mapping vector used in the original SBM [6, 14, 23] has some limitations, partially overcome by the VRLBP and the layered extension. However, a more flexible sampling could be beneficial for the computation of texture measures. In this section we present a technique used to define minimum intensity change sampling paths via HMMs, firstly proposed in [7]. These paths are defined so that the resulting sampled sets contain information about both intensity and structure of the brain.

To define the paths, we consider each three-dimensional image as a tuple containing spatial information in the image range (the coordinates $\mathbf{p} \in \mathbb{I}$) where $\mathbb{I} \subset \mathbb{R}^3$ as well as intensity information ($I(\mathbf{p}) \in \mathbb{R}$). The intensity information will be interpreted as a sampling of the underlying tissue density, and therefore, an estimation of the probability of finding tissue in each position.

Following the notation of the SBM defined in Section 6.2, we formulate a 3D path tracing algorithm that defines a curvilinear mapping set of positions $\mathbb{P}_{\theta, \varphi}$ directly linked to each direction (θ, φ) that is, at the same time, representative of the underlying intensity distribution. We then use both spatial and intensity information to construct the minimum intensity change paths oriented in the direction (θ, φ) . Thus, we could note our 3D path in a certain direction as a Markov Model [82]:

$$\mathbb{P}_{\theta, \varphi} = \{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\} \quad (6.13)$$

Therefore, our optimum path would be the one that maximizes the probability of the path:

$$\mathbb{P}_{\theta, \varphi}^{\text{opt}} = \arg \max_{\mathbb{P}_{\theta, \varphi}} \{P(\mathbb{P}_{\theta, \varphi})\} \quad (6.14)$$

or its equivalent, the probability of all the nodes:

$$P(\mathbb{P}_{\theta, \varphi}) = P(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) \quad (6.15)$$

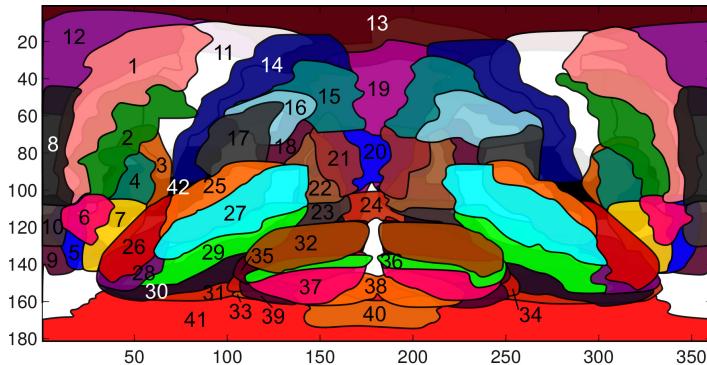


Figure 6.7: SBM mapping of different cortical regions. In the Frontal region, we can find: 1) Frontal Sup., 2) Frontal Mid., 3) Frontal Inf. Oper., 4) Frontal Inf. Tri., 5) Frontal Sup. Orb, 6) Frontal Mid. Orb, 7) Frontal Inf. Orb, 8) Frontal Sup. Medial, 9) Rectus, 10) Frontal Med. Orb., 11) Precentral, 12) Supp. Motor Area. In the Parietal region: 13) Paracentral Lobe, 14) Post-central, 15) Parietal Sup., 16) Parietal Inf., 17) Supramarginal, 18) Angular. In the Occipital region: 19) Precuneus, 20) Cuneus, 21) Occipital Sup., 22) Occipital Mid., 23) Occipital Inf., 24) Lingual. In the Temporal region: 25) Temporal Sup., 26) Temporal Pole Sup., 27) Temporal Mid., 28) Temporal Pole Mid., 29) Temporal Inf., 30) Fusiform, 31) Parahippocampal. The Cerebellum, divided in: 32) Cerebellum Crus 1, 33) Cerebellum 3, 34) Cerebellum 4-5, 35) Cerebellum 6, 36) Cerebellum 7b, 37) Cerebellum 8, 38) Cerebellum 9, 39) Cerebellum 10. And additionally, the 40) Medulla, 41) Brain Stem and 42) Insula.

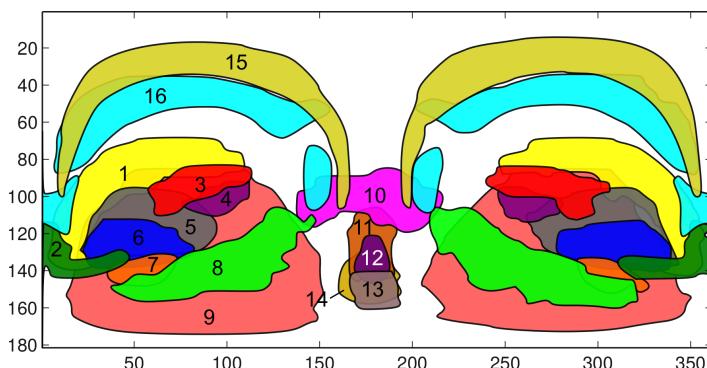


Figure 6.8: SBM mapping of some important subcortical regions and organs. We observe the following subcortical structures: 1) Caudate Nucleus, 2) Olfactory Bulb, 3) Rolandic Operculum, 4) Heschl's gyri, 5) Putamen, 6) Globus Pallidus, 7) Amygdala, 8) Hippocampus, 9) Thalamus, 10) Lingual, 11) Vermis 4-5, 12) Vermis 7, 13) Vermis 9, 14) Vermis 1-2, 15) Cingulate Gyrus, 16) Corpus Callosum

with \mathbf{p}_0 being the origin of the spherical coordinates, and \mathbf{p}_N is the last possible coordinate within \mathbb{I} in the current direction (θ, φ) . When using HMM paths we placed \mathbf{p}_0 at the Anterior Commissure (AC) of the image, although other options, such as setting the origin at the middle point of \mathbb{I} could be considered. This choice is a convention when using the MNI coordinates [163], sharing connectivity with both hemispheres, therefore allowing the optimal computation of our paths. If we assume a first-order Hidden Markov Model (HMM) for the tracing of the path, the probability of the i -th node in the path can be approximated as:

$$P(\mathbf{p}_i | \mathbf{p}_{i-1}, \mathbf{p}_{i-2}, \dots, \mathbf{p}_0) \approx P(\mathbf{p}_i | \mathbf{p}_{i-1}) \quad (6.16)$$

Using this assumption, Eq. 6.15 becomes:

$$P(\mathbb{P}_{\theta, \varphi}) = P(\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_N) = \prod_{i=1}^N P(\mathbf{p}_i | \mathbf{p}_{i-1}) \quad (6.17)$$

Using this HMM definition, the hidden state of each node will be its intensity $I(\mathbf{p}_i)$. Similarly to the original SBM, let $V_{\theta, \varphi} = \{I(\mathbf{p}_0), I(\mathbf{p}_1), \dots, I(\mathbf{p}_N)\}$ be the set containing all the intensities at each node of the path. Thank to this, our optimal path (Eq. 6.14) can be defined as:

$$\mathbb{P}_{\theta, \varphi}^{\text{opt}} = \arg \max_{\mathbb{P}_{\theta, \varphi}} \{P(\mathbb{P}_{\theta, \varphi} | \mathbf{I})\} \quad (6.18)$$

$$P(\mathbb{P}_{\theta, \varphi} | \mathbf{I}) = P(\mathbf{p}_0, \dots, \mathbf{p}_N | I(\mathbf{p}_0), \dots, I(\mathbf{p}_N)) \quad (6.19)$$

$$= \frac{P(I(\mathbf{p}_0), \dots, I(\mathbf{p}_N) | \mathbf{p}_0, \dots, \mathbf{p}_N) \cdot P(\mathbf{p}_0, \dots, \mathbf{p}_N)}{P(I(\mathbf{p}_0), \dots, I(\mathbf{p}_N))} \quad (6.20)$$

where:

$$P(I(\mathbf{p}_0), \dots, I(\mathbf{p}_N) | \mathbf{p}_0, \dots, \mathbf{p}_N) = \prod_{i=1}^N P(I(\mathbf{p}_i) | \mathbf{p}_i) \quad (6.21)$$

and $P(I(\mathbf{p}_0), \dots, I(\mathbf{p}_N))$ is the *a priori* probability of the intensities in the path. We can ignore this term in the optimization process under the assumption that it is constant along the path, which is generally true.

To avoid computational overload, we will define a restricted set of candidates from which we will derive all the needed probabilities. This set of candidates are defined inside the L2-norm support ball $B_{2,r}(\mathbf{p} - \mathbf{p}_{i-1})$ of radius r centred in \mathbf{p}_{i-1} , resulting in the candidate set $\mathbb{P}_{\theta, \varphi}^c = \{\mathbf{p}_{c,1}, \mathbf{p}_{c,2}, \dots, \mathbf{p}_{c,M}\}$.

Individual probabilities $P(I(\mathbf{p}_i) | \mathbf{p}_i)$ needed in the definition of Eq. 6.21 can be computed under the assumption of a normally distributed intensity candidate set $V_{\theta, \varphi}^c$ (containing the intensities of the candidate set $\mathbb{P}_{\theta, \varphi}^c$) with mean

$I(\mathbf{p}_{i-1})$ and variance σ_c^2 . We will estimate the probability of the i^{th} candidate node \mathbf{p}_i as:

$$P(I(\mathbf{p}_i)|\mathbf{p}_{i-1}) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(I(\mathbf{p}_i) - I(\mathbf{p}_{i-1}))^2}{2\sigma_c^2}\right) \quad (6.22)$$

This supports the assumption of minimal intensity change paths, since the $I(\mathbf{p}_i)$ maximizes its probability when similar to $I(\mathbf{p}_{i-1})$.

Finally, we must restrict the direction of the computed path $\mathbb{P}_{\theta,\varphi}$, to match the definition of the SBM framework. We do this by setting an attractor located in the position \mathbf{p}_N , the last possible coordinate within \mathbb{I} in the current direction (θ, φ) . This defines the last term $P(\mathbf{p}_0, \dots, \mathbf{p}_N)$ in Eq. 6.20, the transition probability between states. In this case, we model this transition probability by means of an isotropic Radial Basis Function (RBF), defined in Eq. 6.23:

$$P(\mathbf{p}_0, \dots, \mathbf{p}_N) = P(\mathbf{p}_i|\mathbf{p}_N) \quad (6.23)$$

$$= \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2} (\mathbf{p}_i - \mathbf{p}_N) \Sigma^{-1} (\mathbf{p}_i - \mathbf{p}_N)\right) \quad (6.24)$$

where Σ is the covariance matrix of the RBF. For simplicity we will employ an isotropic RBF kernel, so that Σ is a matrix whose diagonal elements constant and equal to the Euclidean distance between \mathbf{p}_i and \mathbf{p}_N . This way, the attractor conditions the direction of the path, very slightly in the first nodes, and more strongly as it approaches the cortex, leading to a better representation of the underlying structure.

6.3.0.1 Step Size

This algorithm considers all candidate points $\mathbf{p} \in B_{2,r}(\mathbf{p} - \mathbf{p}_i)$ for each member of the final path $\mathbb{P}_{\theta,\varphi}$. Therefore, instead of a fixed step size, we will define the radius r of the support ball. To avoid computational overload while maintaining good results, we will set $r = 3$, which yields approximately 200 candidate points per iteration.

6.3.0.2 Stop Condition

The image \mathbb{I} contains not only information about the structure of the brain, but also many empty space. If the attractor is located at the last point $\mathbf{p} \in \mathbb{I}$, we expect the resulting path to reach that very point. However, what we are really interested in is the brain itself, so we define a stop condition that considers that the path is finished once it reaches the last voxel inside the brain. To do so, we use an intensity threshold.

This threshold is calculated under the entropic thresholding, as in [159]. If we note $G_m \equiv \{I_0, I_1, \dots, I_m\}$ the set containing all intensity levels in the image I (a vectorized image of length m), we can compute a histogram that characterizes the observed frequencies. From these frequencies we can derive the observed probability of the different grey levels. The entropic thresholding defines two distributions after normalization:

$$A \equiv \left\{ \frac{p_0}{P(I_s)}, \frac{p_1}{P(I_s)}, \dots, \frac{p_{s-1}}{P(I_s)} \right\} \quad (6.25)$$

$$B \equiv \left\{ \frac{p_s}{1 - P(I_s)}, \frac{p_{s+1}}{1 - P(I_s)}, \dots, \frac{p_m}{1 - P(I_s)} \right\} \quad (6.26)$$

where $P(I_s) = \sum_i^s p_{I_i}$ is the Cumulative Density Function (CDF) for the s -th grey level. The algorithm is called entropic thresholding because we choose the threshold $I_{th} = I_s$ so that the total amount of information provided by A and B (which we can consider the foreground and background of the image) is maximized. Therefore, we can define the total information provided by choosing the s -th grey level as:

$$TE(s) = E_A(s) + E_B(s) \quad (6.27)$$

$$= - \sum_{i=0}^{s-1} \left(\frac{p_i}{P(I_s)} \right) \log \left(\frac{p_i}{P(I_s)} \right) \quad (6.28)$$

$$- \sum_{i=s}^{m-1} \left(\frac{p_i}{1 - P(I_s)} \right) \log \left(\frac{p_i}{1 - P(I_s)} \right) \quad (6.29)$$

The s that maximizes that latter equation is the grey level that we choose as threshold.

A summary of our HMM-based path tracing method is shown in Algorithm 1. In Figure 6.9 we show all paths computed in all directions (θ, φ) for $0^\circ < \varphi < 360^\circ$ and $0^\circ < \theta < 180^\circ$ at an interval of 1° .

6.3.1 Radial Texture Features

The paths $\mathbb{P}_{\theta, \varphi}$ proposed above could be used as a feature selection tool to extract the set of intensities $V_{\theta, \varphi}$ as in the standard SBM, and even create hybrid HMM-SBM maps from that set. However, since these new paths contain geometric information, they encode the internal structure, which is itself an additional feature. This encoding could be used to characterize the texture in the neighbourhood of $\mathbb{P}_{\theta, \varphi}$.

Algorithm 1: HMM-based Path Creation

input : MRI Brain Image I of size $U \times V \times W$, p_0
output: List of nodes in the optimum path $\mathbb{P}_{\theta, \varphi}^{\text{opt}}$

Compute the $I_{\text{th}} = I_s$ where s maximizes $\text{TE}(s)$;

Set p_0 to the AC;

Compute the attractor position p_N in the direction (φ, θ) ;

$p_i \leftarrow p_0$;

while ($i < \text{IterLimit}$) & ($I(p_i) > I_{\text{th}}$) & ($p_i \in \mathbb{I}$) **do**

 Get the node candidates $\mathbb{P}_{\theta, \varphi}^c = \{p_{c,1}, p_{c,2}, \dots, p_{c,M}\}$ where

$p_{c,m} \in B_{2,r}(p_{c,m} - p_i)$;

 Get the intensities of the candidates $I(p_c) \quad \forall p_c \in \mathbb{P}_{\theta, \varphi}^c$;

foreach $p_c \in \mathbb{P}_{\theta, \varphi}^c$ **compute** $P(p_c|p_N)$ and $P(I(p_c)|p_i)$;

$p_{i+1} = \arg \max_{p_c} [P(I(p_c)|p_i) \cdot P(p_c|p_N)]$;

$i = i + 1$;

$\mathbb{P}_{\theta, \varphi}^{\text{opt}} \leftarrow \{p_0, p_1, \dots, p_N\}$;

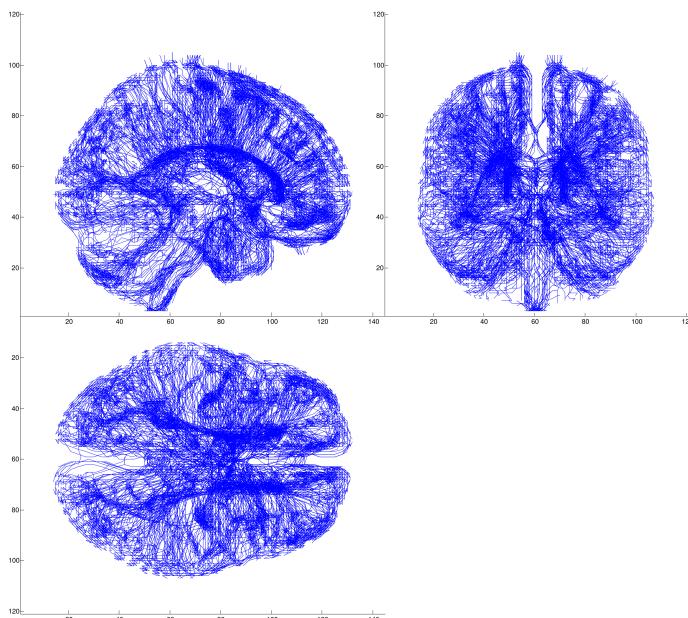


Figure 6.9: Set of HMM based paths over the MRI DARTEL template.

In [14] a helical sampling was proposed to define the neighbourhood of the mapping vector $\mathbf{v}_{\theta,\varphi}$. That is infeasible in this case, due to the topology of the HMM paths. Conversely, we propose a modification of the GLCM (see Section 5.2.2) that instead of computing the texture in a given direction, characterizes it along the defined HMM paths.

This is basically a node-wise GLCM, in which the number of grey-level transitions between adjacent nodes, which we could note as \mathbf{p}_i and \mathbf{p}_{i+1} , is stored along the whole path $\mathbb{P}_{\theta,\varphi} = \{\mathbf{p}_0, \mathbf{p}_1 \dots \mathbf{p}_N\}$. Mathematically, the computation of the GLCM in each point in the path will be:

$$\mathbf{C}_{\Delta_n}(i, j) = \sum_{n=0}^{N-1} \begin{cases} 1 & I(\mathbf{p}_n) = i, I(\mathbf{p}_{n+1}) = j \\ 0 & \text{otherwise} \end{cases} \quad (6.30)$$

where the offset is different for each pair of nodes $\Delta_i = \mathbf{p}_{i+1} - \mathbf{p}_i$.

The definition in Eq. 6.30 computes the values at each node, without considering the neighbourhood. We can generalize it to include the surrounding vicinity of the i -th node \mathbf{p}_i in the HMM path, which we have noted as the set \mathbb{X}_i . Under this generalization, equation 6.30 becomes:

$$\mathbf{C}(i, j) = \sum_{n=0}^{N-1} \sum_{\mathbf{p} \in \mathbb{X}_i} \begin{cases} 1 & I(\mathbf{p}) = i, I(\mathbf{p} + \Delta_n) = j \\ 0 & \text{otherwise} \end{cases} \quad (6.31)$$

Once the GLCM is computed, we can extract a variety of texture descriptors, as defined in Section 5.2.2.2. Specifically, in this work, we will use the aforementioned Energy (eq. 5.6), Entropy (eq. 5.7), Correlation (eq. 5.8), Contrast (eq. 5.9) and Homogeneity (eq. 5.15), along with other texture features proposed in the original Haralick's article [172] as well as in [145] and [127]. These are Dissimilarity[145] (eq. 6.32), Difference Variance[172] (D. Variance, eq. 6.33), Difference Entropy[172] (D. Entropy, eq. 6.34), Inverse Difference Normalized[127] (IDN, eq. 6.35) and Inverse Difference Moment Normalized[127] (IDMN, eq. 6.36).

$$\text{Dissimilarity} = \sum_i \sum_j \{|i - j|P(i, j)\} \quad (6.32)$$

$$\text{D. Variance} = \text{VAR} \left\{ \sum_{|i-j|=k} P(i, j) \right\} \quad (6.33)$$

$$\text{D. Entropy} = - \sum_{k=0}^{N_g-1} \sum_{|i-j|=k} P(i, j) \log \left\{ \sum_{|i-j|=k} P(i, j) \right\} \quad (6.34)$$

$$\text{IDN} = \sum_i \sum_j \frac{P(i, j)}{1 + |i - j|/N} \quad (6.35)$$

$$\text{IDMN} = \sum_i \sum_j \frac{P(i, j)}{i + (j - i)^2/N^2} \quad (6.36)$$

6.4 Evaluation

In this chapter, we have proposed a completely new framework for extracting features and visualizing structural MRI images from the ADNI-MRI database. To evaluate them, we will combine statistical significance assessment and classification analysis. For this purpose, we propose the following experiments:

- Experiment 1: assessment of the original SBM maps and the VRLBP over segmented GM and WM images. We will provide a statistical significance analysis and a classification analysis under the AD vs CTL scenario of the seven proposed maps.
- Experiment 2: assessment of the layered extension of the SBM over segmented GM and WM images. That way, we want to prove if dividing the sampling set, and thus, increasing the depth resolution of the system affects the overall performance of our system. We will provide statistical significance analysis and classification analysis under the AD vs CTL scenario .
- Experiment 3: evaluation of the HMM based paths on simulated datasets, to demonstrate the ability of this algorithm to adapt to different intensity distributions.
- Experiment 4: evaluation of the HMM based paths on a real MRI dataset, by taking the different paths as feature selectors, and evaluating the

performance obtained by the set of intensities selected by each individual path, a combination of them, and the further construction of hybrid HMM-SBM maps. We provide a classification analysis of the selected features under the AD vs CTL scenario.

- Experiment 5: evaluation of the texture maps derived from the HMM-based paths on the T1-weighted MRI dataset, under a classification analysis of AD vs CTL subjects.

In the first three experiments, we will used segmented GM and WM maps, where in experiments 5 and 6, we will use raw, T1-weighted images. The classification analysis is performed by using a linear SVC for classifying, and 10-fold cross validation strategy (see Section 3.3 for more details). For estimating statistical significance, we use the two-sample t-test defined in Section 4.1.1.

6.5 Results

6.5.1 Experiment 1: Original and VRLBP Spherical Brain Mapping

First, with experiment 1, we test the original and VRLBP SBM maps, by means of significance and classification analysis. To start with, we provide the significance maps computed under the AD vs CTL scenario in the six original SBM measures (surface, thickness, number of folds, average, entropy and kurtosis) in Figure 6.10.

For $p < 0.05$, in the database subset of the ADNI-MRI, the significance threshold can be established at $|t| > 1.96$. This means that, in Figure 6.10, the most relevant differences between classes can be found at the areas coloured in dark red (positive, CTL subjects have a higher measure) and dark blue (negative, AD subjects have a higher measure). For the anatomic structures, we refer to the anatomical reference presented in Section 6.2.3.

We first note that the surface measure, tested in both GM and WM maps, does not look relevant. Very few significant pixels are scattered throughout the image, although with a slightly higher concentration in the areas corresponding to the temporal lobe.

In the remaining GM maps, we observe similar behaviours, with higher absolute t-values located in the frontal, occipital and parietal lobes. But again, the most significant areas can be found at the temporal lobe. It is more obvious in the average and entropy measures, but can also be found at the thickness, or negatively in the number of folds and kurtosis maps. This points to the well known fact that most of the neurodegeneration in GM occurs within the

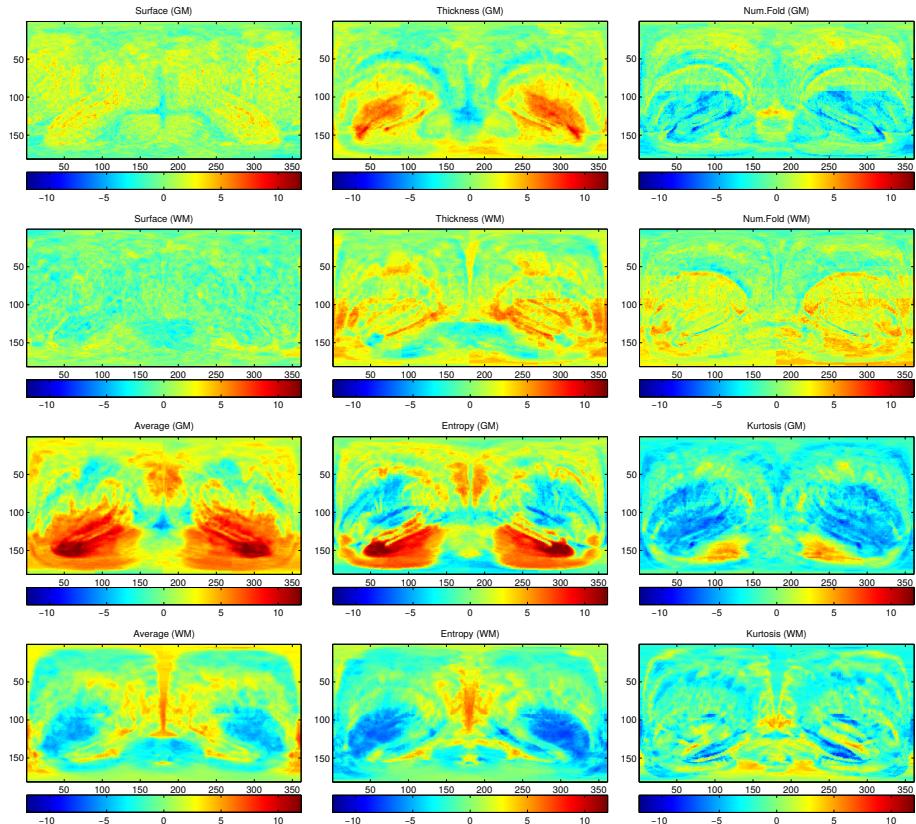


Figure 6.10: t-maps to assess the regions with a higher statistical relevance in the AD vs CTL paradigm, for each SBM measure and using GM and WM maps.

structures that are mapped to these directions, including mid temporal lobe, amygdala, hippocampus or parahippocampal gyrus, considered a strong indicator in the NINCDS-ADRDA criteria [96]. Other GM structures such as the caudate nucleus and putamen also appear with significant t values, especially in the entropy and kurtosis maps (with negative t) [37].

In the WM, the levels of significance achieved are smaller than in the GM, but still high. We see the higher levels of number of folds and thickness located in the vicinity of those obtained for GM, but in negative. For the average, entropy and kurtosis measures, we observe a different behaviours. These maps present large areas of negative t-values located in the Caudate Nucleus, Globus Pallidus and Putamen. Areas around the posterior cingulate gyrus and the adjacent precuneus also present reduced t values, which could be related to cell loss, as suggested in [133].

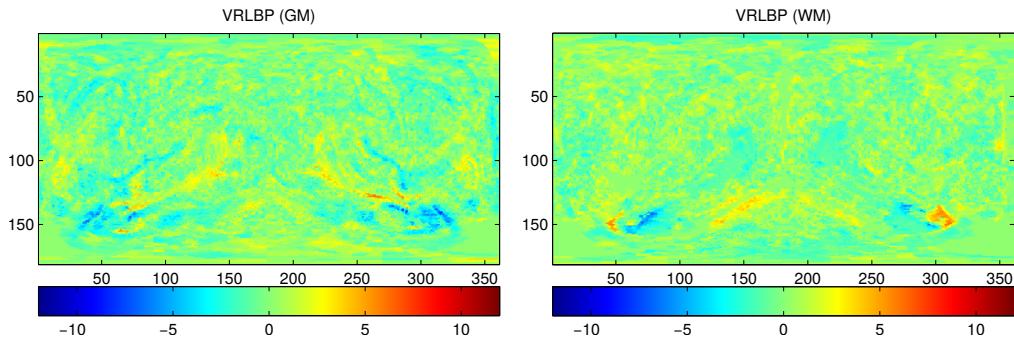


Figure 6.11: Maps that present the level of the t statistic in the AD vs. CTL paradigm, for the VRLBP projections mapping over a) GM and b) WM.

Finally, we take a look at the statistical significance of the VRLBP maps. In Figure 6.11, we can observe that most of the image features small t values, however there are small areas that contain higher significance. These areas correspond to the temporal lobe, amygdala and hippocampus (in the GM maps) and smaller regions located at the limits between the hippocampus and amygdala (in WM).

Now, we perform a classification analysis of the images. To this purpose, we will use the computed t-maps to select the most relevant pixels in the SBM maps, which will be used to train and test a SVC. The performance results for the six original measures and the VRLBP approach, including the percentage of selected pixels (perc.), can be found at Table 6.1.

We can see a general trend in which the original statistical measures (average, entropy and kurtosis) clearly outperform the morphological ones (surface, thickness and number of folds), although the tissue thickness is the best per-

Approach	Perc.	Accuracy	Sensitivity	Specificity
Surface (GM)	0.100	0.638 ± 0.006	0.660 ± 0.030	0.616 ± 0.024
Surface (WM)	0.100	0.672 ± 0.007	0.692 ± 0.018	0.652 ± 0.018
Thickness (GM)	0.725	0.781 ± 0.007	0.811 ± 0.011	0.751 ± 0.017
Thickness (WM)	0.925	0.758 ± 0.009	0.773 ± 0.017	0.744 ± 0.011
Num.Fold (GM)	0.600	0.749 ± 0.013	0.782 ± 0.019	0.716 ± 0.013
Num.Fold (WM)	0.500	0.757 ± 0.005	0.745 ± 0.006	0.768 ± 0.009
Average (GM)	0.575	0.879 ± 0.005	0.897 ± 0.006	0.861 ± 0.006
Average (WM)	0.150	0.800 ± 0.011	0.802 ± 0.013	0.798 ± 0.009
Entropy (GM)	0.825	0.846 ± 0.008	0.842 ± 0.009	0.849 ± 0.011
Entropy (WM)	0.525	0.796 ± 0.006	0.811 ± 0.009	0.781 ± 0.009
Kurtosis (GM)	1.000	0.753 ± 0.007	0.801 ± 0.011	0.704 ± 0.015
Kurtosis (WM)	0.175	0.697 ± 0.008	0.702 ± 0.018	0.693 ± 0.009
VRLBP (GM)	0.200	0.903 ± 0.010	0.890 ± 0.012	0.916 ± 0.018
VRLBP (WM)	0.150	0.909 ± 0.014	0.899 ± 0.028	0.919 ± 0.018

Table 6.1: Performance values (Average \pm Standard Deviation) for the different SBM approaches.

forming of these, as it could be expected. For GM maps, average (0.879 ± 0.005) and entropy (0.846 ± 0.008) achieve the best performance, followed by thickness (0.781 ± 0.007), kurtosis (0.753 ± 0.019), and finally, the number of folds (0.749 ± 0.013) and surface (0.638 ± 0.006). These results match what we presented previously with the statistical maps, in which the surface contained the less significant measures.

As for the performance of the SBM measures based on WM maps, the rank is very similar to that obtained for GM. Again, average (0.800 ± 0.011) and entropy (0.796 ± 0.006) are the best performing features among the original measures. They are followed by thickness and the number of folds, with respectively 0.758 ± 0.009 and 0.757 ± 0.005 . Finally, the kurtosis and the surface are the less discriminant measures again, with 0.697 ± 0.008 and 0.672 ± 0.007 . All the original measures are outperformed by the VRLBP approach, which achieves more than 90% accuracy in both GM and WM.

In Figure 6.12 we explore the evolution of the performance as the number of selected pixels varies. Very small differences in the accuracy exist for most measures, and so, we can consider that the performance of the SBM, once that a few thousand significant pixels (a 10% of $181 \times 361 = 65341$) have been selected, the system performs well independently of that number. That is, however, not the case of the surface approach, and more remarkably, of the VRLBP. In this latter case, for both tissues, the performance is high in the first 40% of selected voxels, but after that, it dramatically decreases down to less than 70% accuracy, which may be related to the small significant regions found in Figure 6.11.

6.5.2 Experiment 2: Layered Extension

In Experiment 2, we have assessed how the performance of our SBM varies when adding different layers, which might theoretically improve the accuracy of the SBM representation. First, we will take a look at the performance achieved by this extension on all six original SBM maps, using 4 layers and different t-thresholds (2, 4, 8 and 10). This is presented at Figure 6.13.

The first thing that we can observe is that for GM maps, the better performance is achieved within the second and third layers, and only for the second layer in the case of WM. In almost all cases, the average performance achieves higher accuracy than any other SBM measure, closely followed by entropy. It also tells us that, with less restrictive t-thresholds, the performance generally decreases, and the best values are between 2 and 4.

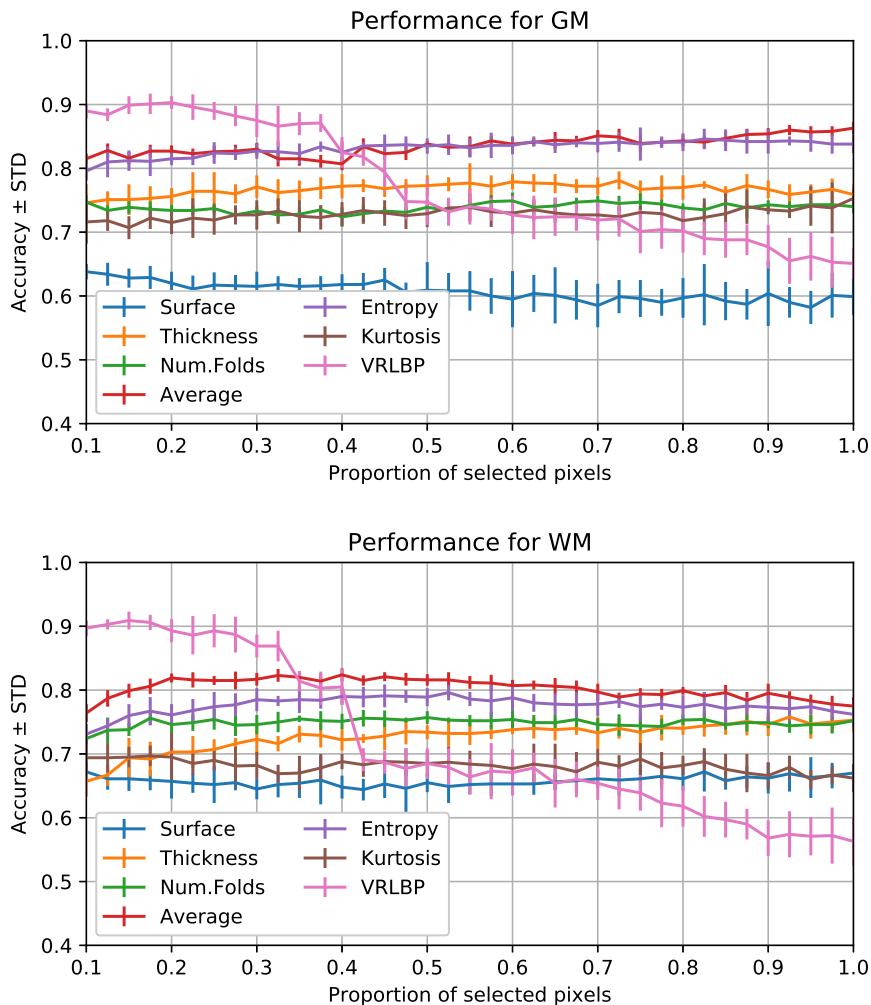


Figure 6.12: Performance for the different SBM approaches over the: a) GM and b) WM.

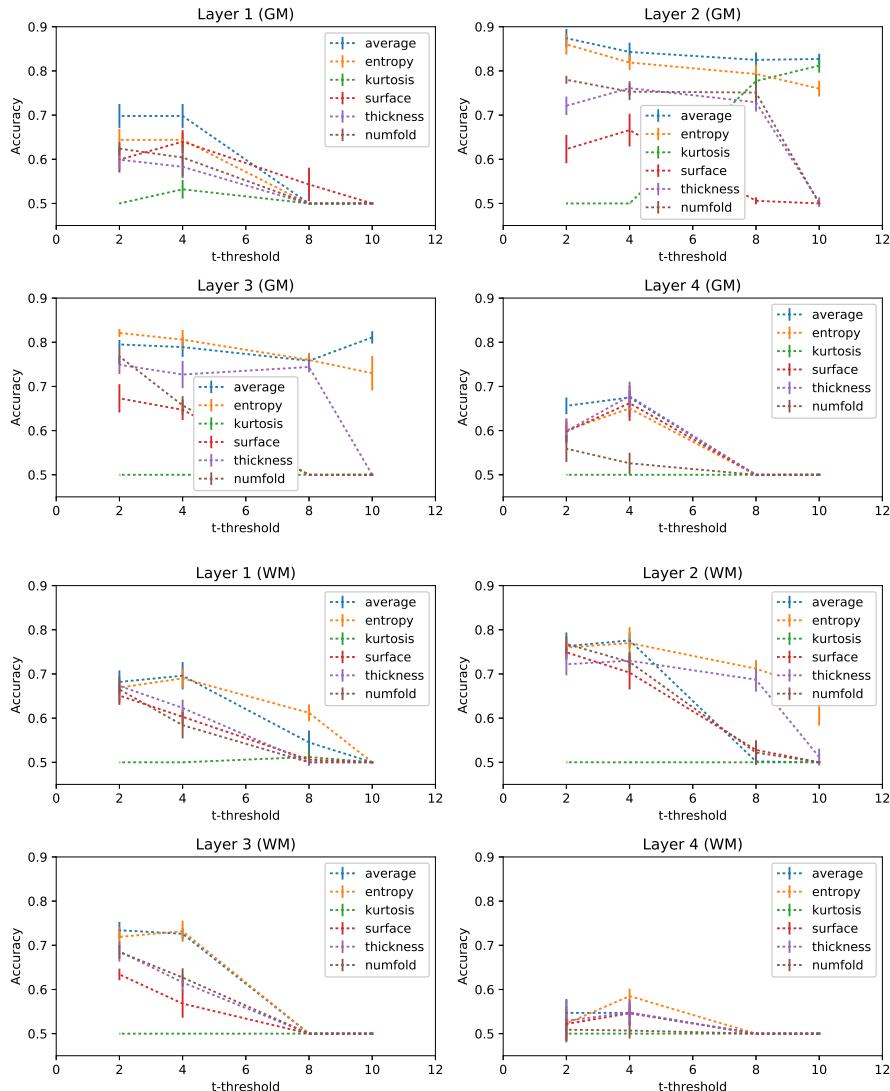


Figure 6.13: Performance for the different four-layered mappings over the: a) GM and b) WM at different levels of statistical significance.

Performing a significance analysis on the 4 layer extension of the 7 proposed SBM maps over the GM and WM images could be intractable, therefore we will only provide t-maps of the best performing measure: the average on the GM and WM tissues. The significance assessing of this case is found at Figure 6.14.

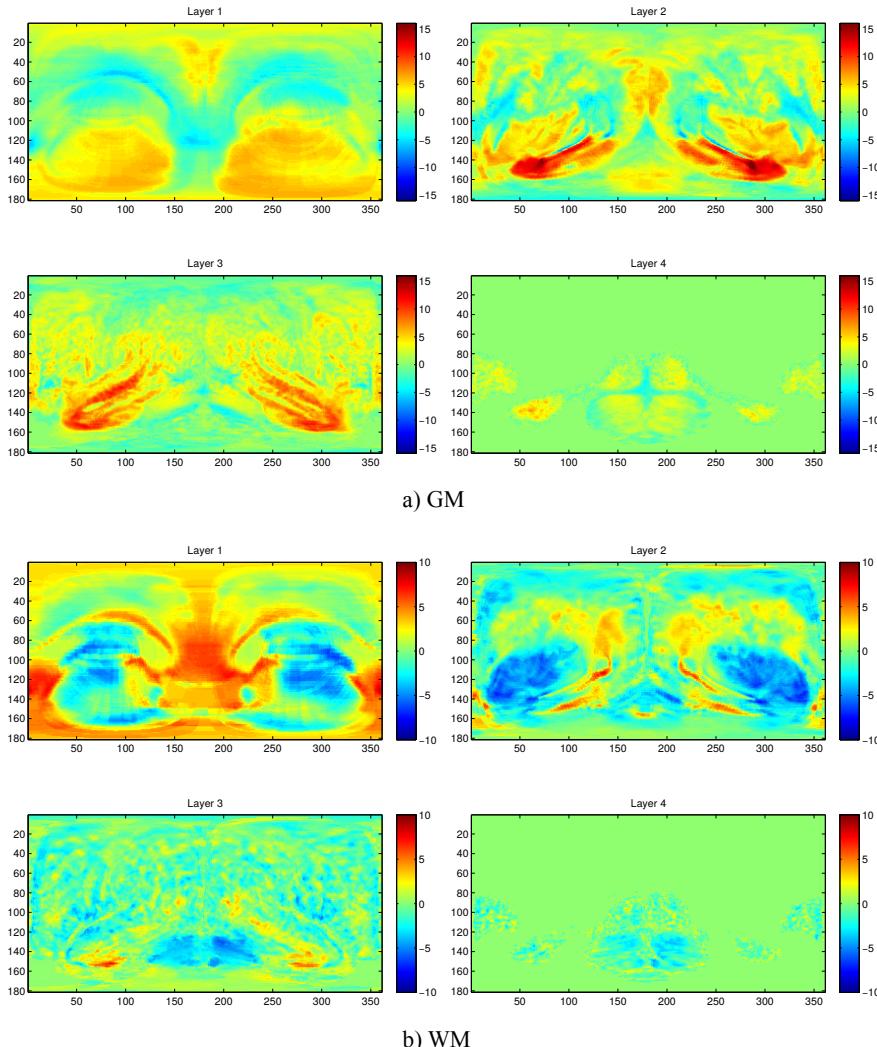


Figure 6.14: t-maps that present the level of statistical relevance in the AD vs CTL paradigm, for a four-layered average mapping over a) GM and b) WM.

In GM, the most obvious changes are located in layers 2 and 3, specifically at the hippocampus, parahippocampal gyrus and amygdala (layer 2), and the temporal lobe (layer 3), where the values achieved by CTL subjects are much higher than those found in AD. This could reveal atrophy in these organs,

as it has been reported in the literature [37, 96] and will be discussed later. For WM, we obtain large negative t-values in areas occupied by the rolandic operculum, heschl's gyri, putamen and globus pallidus, with positive values in parts of the hippocampus, and parts of the temporal lobe. Nevertheless, the most significant differences can be located in layer 1, at the borders between ventricles and thalamus, and the cuneus, precuneus and posterior cingulate gyrus, which have been reported in [133].

6.5.3 Experiment 3: HMM on Synthetic Datasets

Now, we will analyse the behaviour of the HMM paths on some two and three-dimensional examples. In Figure 6.15, we used a gaussian mixture probability density function, with four isotropic gaussian kernels, to generate a synthetic image. We wanted to test the direction of the HMM paths on this image, by locating the initial point $p_0 = (120, 20)$ and the attractor at the centre of one of the kernels, $p_N = (20, 60)$. This simple example illustrates how the generated paths maximizes both the orientation of the path (towards p_N) and the minimum change in the intensity values. This was specially noticeable in the last nodes of the path, where the approximation to p_N is very gradual, due to the restrictions imposed. In this example, we used a support ball with $r = 3$.

Afterwards, we wanted to demonstrate how these paths adapt to three-dimensional structures. In this case, we created a helix-shaped point distribution in Fig. 6.16. Since we need per-voxel intensity values (probabilities) to estimate the paths, we have approximate this by the number of points within each voxel over the total number of points. By setting p_0 to the point on the helix distribution with smallest z coordinate, and p_N to the one with maximum z, we trace the path shown in Fig. 6.16. This paths adapt to the distribution and follows it consistently until it reaches the attractor at the top of the helix.

Finally, and prior to the three-dimensional brain images, we test the algorithm on a bidimensional real-world example: a digital elevation model of the Iberian Peninsula. The image, whose intensity represents height, was acquired by the LANDSAT SRTM30+ mission. We tested multiple paths by establishing sequentially p_0 and p_N in ten different cities. The paths shown at Figure 6.17 optimize both the distance and height variation as well as resemble -in most cases- the main roads that connect these cities in the real world. Given the dimensions of the image, in this case, the L2-norm of the support ball has been set to $r = 30$.

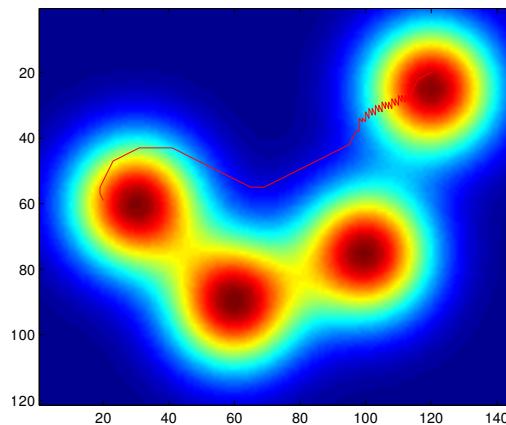


Figure 6.15: Path traced over a gaussian mixture distribution of 4 isotropic gaussian kernels.

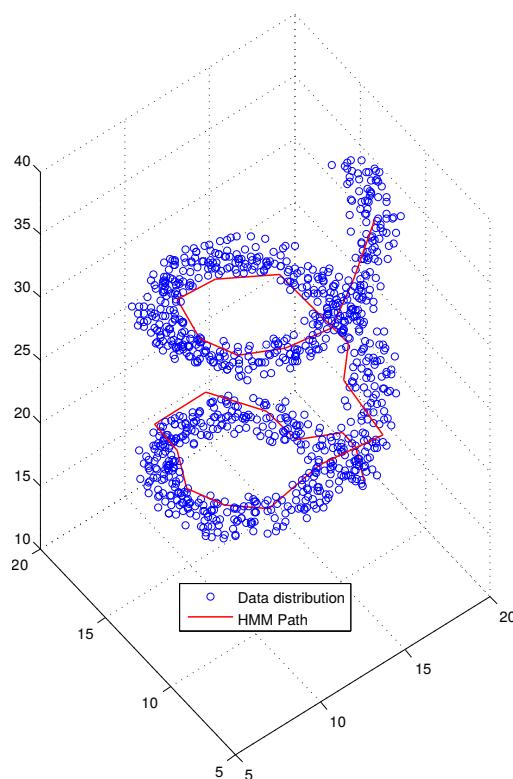


Figure 6.16: HMM path computed inside a density distribution defined by an helix.

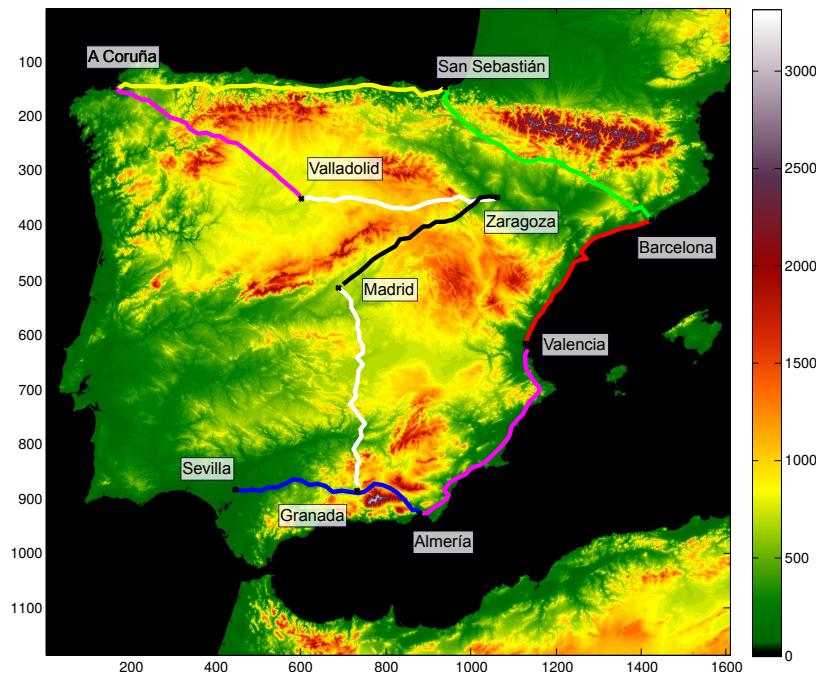


Figure 6.17: Simulation of the HMM-based path tracing over an Iberian Peninsula height map, interconecting different cities.

6.5.4 Experiment 4: Feature Selection using HMM Paths

In Experiment 4 we begin to apply the HMM paths to real MRI images from the ADNI-MRI dataset. We have first defined a set of paths over the DARTEL template, which we call the canonical paths. The canonical paths follow the anatomy of a normal subject to whom all other images have been registered. Therefore, the location of the different nodes on the paths are fixed throughout the general anatomy of all images, and we could characterize the differences between classes by analysing the intensity distribution on these paths –in other words, the tissue density–. In short, we apply the canonical paths as a feature selection algorithm to the different images.

To start with, we will take all the $180 \times 360 = 64800$ canonical paths computed in all spatial directions defined by $\varphi \in [0, 360)$ and $\theta \in [-90, 90)$. These paths have been used to select the voxel intensities where the nodes are placed, as in the original SBM. Due to the influence of the different variables (trace of the paths, stop condition), the amount of values varies from 2 to several tens. First we will test the accuracy achieved by each path itself, using the set of intensities as features in a SVC classifier. The accuracy reached by each path is represented by colour in Figure 6.18.

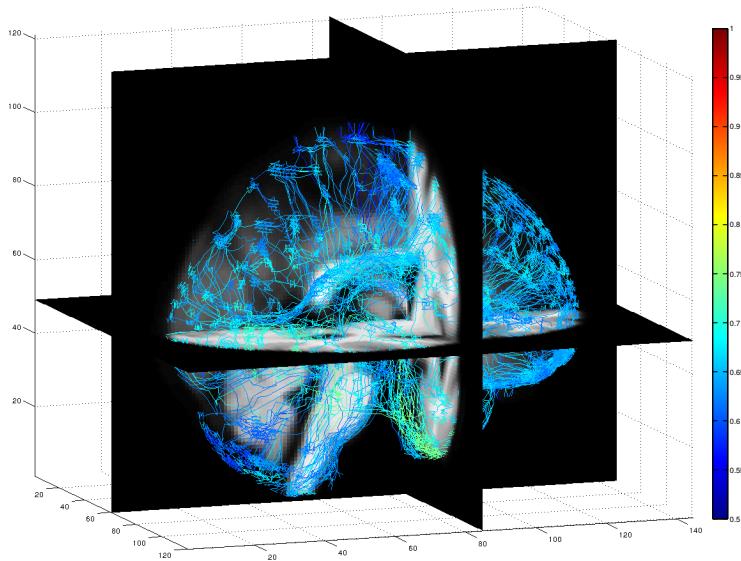


Figure 6.18: canonical paths computed in each direction (θ, φ). Each path's colour represent the accuracy in a differential diagnosis. Only one in every five paths are shown for clarity purposes.

The highest performance is achieved by some of the light green paths that cross the temporal lobe, where they reach a decent accuracy up to 0.8028 ± 0.0873 . Compared to the VAF approach obtained in the ADNI-MRI dataset using T1-weighted images, under the AD vs NOR scenario (0.789 ± 0.066 accuracy), this is a dramatic feature reduction while increasing the performance of the system.

Then, we want to test whether using all the intensities contained within the highest scoring paths can increase this raw performance. To do so, we select the paths that achieved accuracies ≥ 0.7 , and afterwards, we used a t-test over the set of intensities selected by these paths. We used a threshold $|t| > 1.96$ ($p < 0.05$) and all the significant intensities were used to construct a new training set. The performance for using either all voxels in the selected paths or only the significant voxels is presented at Table 6.2, for left, right and both sides.

Finally, we proceed as in the original SBM article [6], by computing different SBM measures using each set of intensities $V_{\theta, \varphi}$ selected by the HMM path at the direction of the attractor (θ, φ). We will compute the average, entropy and kurtosis measures, as well as an additional variance measure, the variance

Selection	Side	Accuracy	Sensitivity	Specificity
All	Left	0.769 ± 0.035	0.717 ± 0.061	0.822 ± 0.057
	Right	0.792 ± 0.080	0.706 ± 0.120	0.878 ± 0.101
	Both	0.806 ± 0.069	0.733 ± 0.073	0.878 ± 0.097
t-test	Left	0.733 ± 0.037	0.694 ± 0.099	0.772 ± 0.124
	Right	0.781 ± 0.085	0.711 ± 0.122	0.850 ± 0.083
	Both	0.828 ± 0.054	0.794 ± 0.095	0.861 ± 0.039

Table 6.2: Performance values ($\pm SD$) for the selected paths as features. First three rows correspond to using all intensities in the selected paths, and the following rows use only significant intensities (assessed by t-test).

Feature	Accuracy	Sensitivity	Specificity
Average	0.594 ± 0.062	0.661 ± 0.121	0.528 ± 0.106
Variance	0.750 ± 0.064	0.633 ± 0.131	0.867 ± 0.102
Entropy	0.603 ± 0.069	0.661 ± 0.071	0.544 ± 0.125
Kurtosis	0.756 ± 0.105	0.733 ± 0.165	0.778 ± 0.150

Table 6.3: Performance values ($\pm SD$) for the hybrid HMM-SBM maps, using some statistical measures.

of the intensity set $V_{\theta,\varphi}$. An example of the computed maps is provided in Figure 6.19.

The resulting maps are very different from the original SBM ones, and the different structures cannot be so easily located. This is mainly due to the adaptivity of the HMM paths, featuring a large number of twists and turns in their direction, and making them less powerful for visualization. However, the performance of using these hybrid HMM-SBM maps can still be high, as we can find at Table 6.3.

6.5.5 Experiment 5: Texture SBM Maps based on HMM Paths

After evaluating the HMM paths as a feature selection method, and the hybrid HMM-SBM maps, we want to compute new texture maps, similar to the approach used in the VRLBP extension, but using a path-based GLCM. To do so, we will derive a GLCM from each of the canonical paths, and extract one single value per texture feature and direction, which allow us to construct texture SBM maps. In Table 6.4 we show the performance values for the maps built using each of the texture features proposed in Section 6.3.1

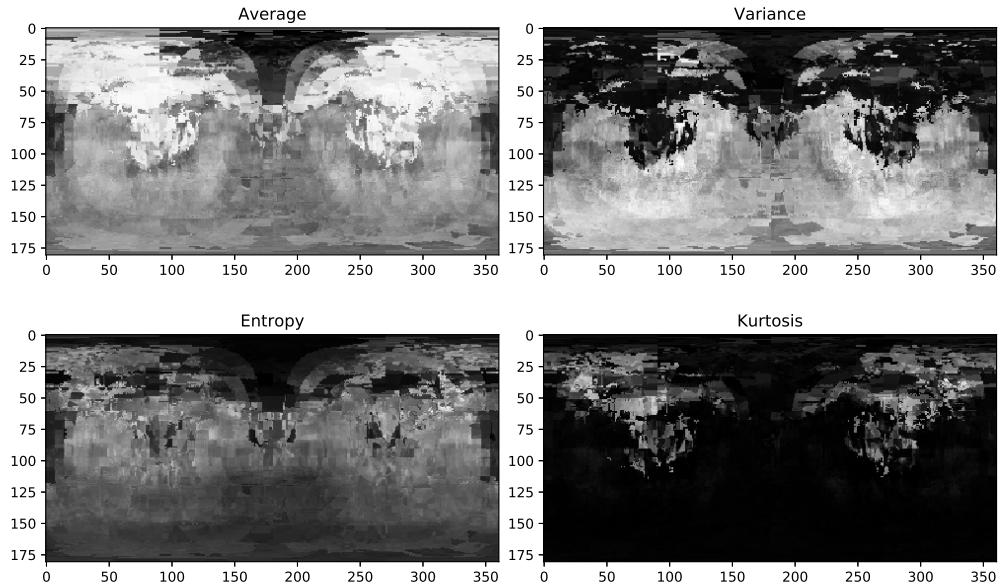


Figure 6.19: Example of the resulting hybrid HMM-SBM texture maps.

Feature	Accuracy	Sensitivity	Specificity
Energy	0.689 ± 0.061	0.700 ± 0.115	0.678 ± 0.073
Entropy	0.675 ± 0.101	0.672 ± 0.115	0.678 ± 0.159
Correlation	0.672 ± 0.068	0.672 ± 0.112	0.672 ± 0.100
Contrast	0.733 ± 0.060	0.689 ± 0.126	0.778 ± 0.105
Homogeneity	0.697 ± 0.058	0.700 ± 0.115	0.694 ± 0.106
Dissimilarity	0.711 ± 0.085	0.678 ± 0.110	0.744 ± 0.102
Difference Variance	0.736 ± 0.070	0.683 ± 0.098	0.789 ± 0.090
Difference Entropy	0.725 ± 0.122	0.683 ± 0.176	0.767 ± 0.114
IDN	0.719 ± 0.065	0.683 ± 0.108	0.756 ± 0.105
IDMN	0.717 ± 0.076	0.678 ± 0.125	0.756 ± 0.084

Table 6.4: Performance values ($\pm SD$) for each of the 10 texture features.

Difference variance achieves here the higher accuracy, 0.736 ± 0.070 , a value below the VAF approach in this database. The performance achieved by these texture features is always higher than 65% (most of them above 70%) which reveals a discrete discrimination ability, although much smaller than the one shown by the HMM paths, used as feature selection tool.

6.6 Discussion

6.6.1 Spherical Brain Mapping

The SBM defines a completely new approach to visualization and feature extraction of segmented MRI images. We have tested the original maps (including the VRLBP approach) in Experiment 1, and its layered extension in Experiment 2.

From the analyses performed, we can conclude that the most significant changes occur mainly in the GM tissue (see fig. 6.20), although some maps perform better with WM. According to the widely documented structural changes due to AD, these occur mainly in GM [37, 74, 108, 116, 120, 133], which is consistent to what we found, and the significance found in WM might be the “negative” influence of the GM loss. That is revealed in the statistical significance analysis of Figures 6.10 and 6.11, where the t-values in WM are negative in the areas where we found positive values in GM maps. This can be also appreciated in the layered approach, at Figure 6.14.

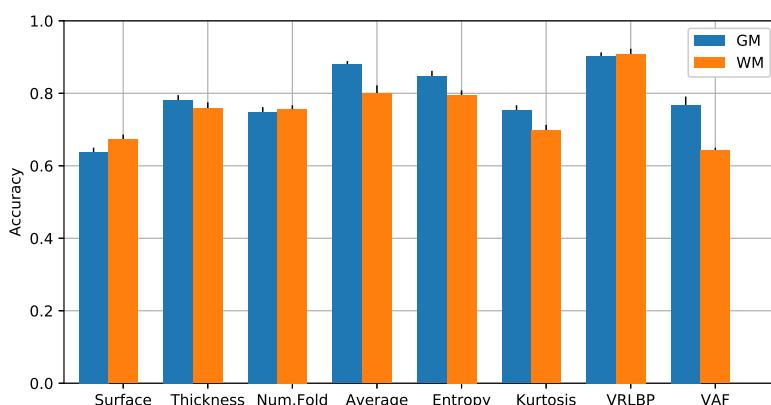


Figure 6.20: Performance at the operation point for the different mappings over the GM and WM, compared with the performance of VAF.

The mappings defined at Sections 6.2, 6.2.1 and 6.2.2 model a series of properties of the tissues crossed by the mapping vector $\mathbf{v}_{\theta,\varphi}$. We can group all the SBM measures in three categories: morphological (surface, number of folds and thickness), statistical (average, variance, entropy or kurtosis) and texture (VRLBP). The morphological approaches are easily interpreted, since they represent things such as the surface of the tissue, the complexity of the gyri and sulci folding, and the thickness. Surface and thickness are defined in a similar way to other widely-used software, such as Freesurfer [116, 142], although are less powerful in defining this complexity, due to the limitations in our rectilinear mapping.

Therefore, the surface measure is not a faithful representation of the surface of the cortex, as it can be seen in Fig. 6.3, and later in the t-maps at Fig. 6.10. That is why its performance is so small in GM, due to the superposing gyri and sulci, and better for WM, where it achieves a higher level of detail. As for the thickness measure, it cannot achieve the same level of detail of the Freesurfer's measures, although gathering much more information than other morphological approaches. Nevertheless, the thickness of these tissues can still be very descriptive, and it achieves a decent performance in the classification analysis. The number of folds is the last of these morphological measures. It is intended to measure the complexity of the cerebral cortex, although judging by its performance, it does not do a very good job.

The statistical measures, especially average, achieve higher performance in the classification analysis. That might be due to the precision of the numbers used. Whereas in the morphological features we obtained integers (as they count the number of voxels or zero-crossings), in the statistical measures values such as average or entropy compute high-precision floating values. This can help in further computation, although values such as kurtosis are outperformed by, for example, thickness.

The average is the best-performing measure in this group, and can be interpreted as the average amount of tissue -even tissue density- in the direction (θ, φ) . A comparison between the densities of the tissues in AD and CTL can reveal the neurodegeneration and tissue loss typically associated with the disease. This is perhaps why in the significance analysis we found high absolute t-values in areas typically affected by AD, such as the mid temporal lobe, amygdala, hippocampus, parahippocampal gyrus or some structures of the basal ganglia, such as caudate nucleus, globus pallidus or putamen [37, 96].

Entropy is a more complex statistical concept that comes from information theory, but is usually related to the amount of information, or in other words, the “randomness” of a source. In our particular case it could be interpreted as a measure of texture, that is, the grey-level variability in the direction of

$\mathbf{v}_{\theta,\varphi}$. These maps perform very similar to the average ones in both GM and WM, suggesting that the entropy accounts here for the tissue density as well.

The last measure of this group, the kurtosis, is also the less performing statistical feature. It is a fourth-order statistic that might be interpreted as the “peakedness” of a probability distribution. In the case of the SBM maps, it models how abrupt the intensity changes in the direction $\mathbf{v}_{\theta,\varphi}$ are. This relates kurtosis to the number of folds, and as in this case, the performance is also small. This is probably due to the fact that the main neurodegenerative processes are better modelled by atrophy-related measures, such as average or thickness, than with the complexity measures such as number of folds, kurtosis.

Finally, the VRLBP gathers not only the intensity values in the mapping vector, but also its neighbourhood, thanks to the helical sampling applied. This is probably why the VRLBP achieve the best performance in the classification analysis, with accuracies above 0.9 for both GM and WM tissues. Texture can also be associated with changes in the tissue density or distribution, which could lead to this higher performance. The most relevant regions in the VRLBP correspond to small areas in the temporal lobe, amygdala and hippocampus, where the neurodegeneration could be causing the texture changes that we detected.

The breakdown of the SBM maps into several layers might be considered a enhancement over the single maps, thanks to the level of detail achieved at different depths. However, it barely outperforms (0.874 ± 0.021 accuracy) the single-layer SBM average measure, much less the best-performing VRLBP. Therefore, its only contribution is a better visualization of overlapping structures. The most discriminant layer in a four-layer decomposition is layer 2, which is consistent with the presence of some AD-related organs such as the hippocampus, amygdala or putamen.

6.6.2 Paths via HMM

Now we will discuss the HMM paths proposed in Section 6.3, whose results were analysed in Experiments 3, 4 and 5. The paths were proposed as a curvilinear substituent of the SBM mapping vector $\mathbf{v}_{\theta,\varphi}$, capable of adapting to the structure changes found in the brain. Our paths $\mathbb{P}_{\theta,\varphi}$ are built upon a minimun intensity variation pathway that starts at the AC, and follows the orientation of an attractor placed in the limits of the image at the spherical coordinate pair (θ, φ) .

There is a fundamental reason why AC is the best choice as first node of the paths: it is a privileged position where the left and right hemispheres are connected. A different point would lead to suboptimal paths, whose construction might be quickly interrupted by the abrupt changes at the ventricles, and therefore, the resulting set of paths would not cover the entire brain.

It is important to note that these are not functional connectivity maps like Diffusion Tensor Imaging (DTI) fibre tracts, which have also been used in the diagnosis of AD [55, 89]. DTI tracts are based upon diffusion images that quantify the water molecule motion –in magnitude and direction– throughout a tensor processing. Our HMM path definition, on the other hand, only characterize grey level connectivity in MRI images, starting in white matter and progressively transitioning to GM, oriented in a fixed direction. This means that, although images like Figure 6.18 could resemble DTI fibres, they are not related at all, and their utility as a feature selection tool is completely different.

Our paths demonstrated in Experiment 3 that they can adapt easily to two and three-dimensional images, and even to the statistical distribution of points in space (as in the helix example). They were successfully tested against a real world example, aimed at tracing roads on a height map of the Iberian Peninsula, with the result of paths that closely resemble real roads in which the minimum variation in height is key.

Later, in Experiment 4, we used the HMM paths as a feature selection tool to construct the set of selected intensities $V_{\theta,\varphi}$, as in the original SBM. Individual paths as well as the pooled intensities of all paths in a hemisphere were tested using a SVC. The differences between intensity sets in AD and CTL were significant enough to achieve accuracies over 80% in some paths. The paths that achieved more than 75% accuracy are shown in Figure 6.21, along with the rendered structures crossed by them, according to the AAL brain atlas [131].

The structures that are crossed by the best-performing paths are the hippocampus, amygdala, thalamus, fusiform and inferior temporal gyrus. GM loss in the hippocampus has been extensively reported in the literature [85, 96, 133], as well as its surrounding structures (amygdala, parahippocampal gyrus and fusiform) [133]. Some studies have also reported neurodegeneration at the thalamus and putamen, also related to early AD [85], and in advanced AD, atrophy in most of the neocortex [85, 96, 133]. However, pooling all the intensities selected by these paths together did not increase the performance achieved by the best performing paths.

The last test of Experiment 4 was the hybrid SBM-HMM maps. Average, entropy, kurtosis and variance maps were tested, and we found that they were discriminant (accuracy above 70% in most cases), yet not very powerful. This is however coherent with the definition of the paths, where the inten-

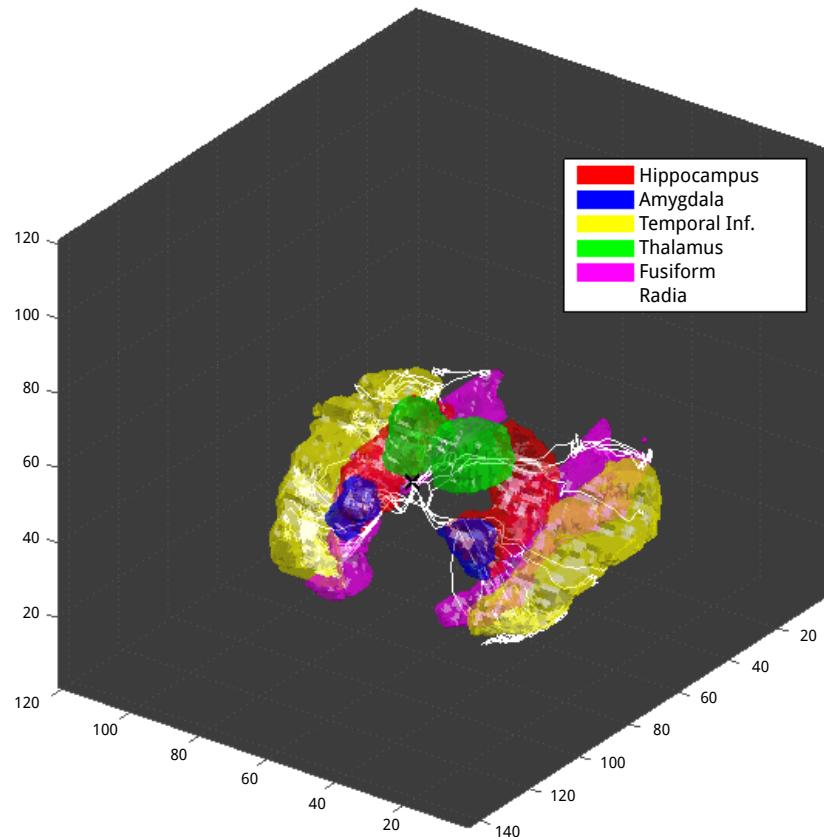


Figure 6.21: Paths that obtain more than 75% accuracy, and a three-dimensional representation of the structures crossed by them.

sity changes are small and progressive, and therefore, statistics that focus on higher order variability (such as variance or kurtosis) are more representative of this behaviour.

Finally, we tested the HMM paths' application to compute radial GLCM and texture features at Experiment 5. Here we saw that the performance achieved with any of the texture feature maps was smaller than using the best paths themselves, or even the most significant paths. However, the real utility of the texture HMM maps could reside in its application to longitudinal studies. Texture has been already related to the evolution of AD [16], and so, the evolution of the generated texture maps over the canonical paths throughout time could be a good indicator of the progression of AD.

Further improvements could be made by computing an independent set of paths per subject, and then extract morphological measures such as curvature or torsion, or extend the computation of textural features to longitudinal studies, as we mentioned previously. This seem a very promising approach that will be tested in future works.

6.6.3 SBM Framework Comparison

To obtain a general view of the SBM and its extension within the state of the art, in Table 6.5 we present some of the best results that have been already commented. We first present the values for the VAF approach [120] in T₁, GM and WM images. Then, we present the performance of the average and VRLBP maps of our SBM, and the HMM paths, both as feature selection tool (individual paths and selected features from the paths), the hybrid HMM-SBM tool (variance and kurtosis) and the best scoring texture feature (difference variance).

Afterwards, we provide reference values of some state of the art methodology that has been tested against the ADNI-MRI database. These are: LVQ) a Learning Vector Quantization (LVQ) [36] used to model the tissue distribution of CTL and AD images and project the images to a new space, just to use the projections in a SVC; and SCA) the Spatial Component Analysis proposed in [22], a bayesian network modelling of the dependencies between affected regions (or spatial components) in AD.

The original SBM achieve generally better results than its counterpart using HMM paths. That gives us an idea that the utility of these paths is not to generate the SBM maps, but to characterize the structure itself of the brain, which could have applications when applying a morphological characterization of the individual set of paths computed over each subject, or even in segmen-

App.	Feature	Tis.	Accuracy	Sensitivity	Specificity
VAF	Intensity	T ₁	0.789 ± 0.066	0.767 ± 0.094	0.811 ± 0.099
		GM	0.768 ± 0.011	0.752 ± 0.016	0.785 ± 0.016
		WM	0.642 ± 0.009	0.668 ± 0.012	0.617 ± 0.013
SBM	average	GM	0.879 ± 0.005	0.897 ± 0.006	0.861 ± 0.006
		WM	0.800 ± 0.011	0.802 ± 0.013	0.798 ± 0.009
	VRLBP	GM	0.903 ± 0.010	0.890 ± 0.012	0.916 ± 0.018
		WM	0.909 ± 0.014	0.899 ± 0.028	0.919 ± 0.018
HMM	Ind. Paths	T ₁	0.806 ± 0.069	0.733 ± 0.073	0.878 ± 0.097
	Sel. Paths	T ₁	0.828 ± 0.054	0.794 ± 0.095	0.861 ± 0.039
	Variance	T ₁	0.750 ± 0.064	0.633 ± 0.131	0.867 ± 0.102
	Kurtosis	T ₁	0.756 ± 0.105	0.733 ± 0.165	0.778 ± 0.150
	D. Variance	T ₁	0.736 ± 0.070	0.683 ± 0.098	0.789 ± 0.090
Other	LVQ	GM	0.869 ± 0.101	0.822 ± 0.120	0.890 ± 0.102
	SCA	GM	0.880 ± 0.0*	0.926 ± 0.0*	0.845 ± 0.0*
		WM	0.808 ± 0.0*	0.817 ± 0.0*	0.800 ± 0.0*

* SCA used leave-one-out cross-validation. SD is 0.

Table 6.5: Comparison between our algorithm performance values (best values for selected voxels in all paths and texture features) (\pm SD) and other methods in the bibliography

tation of the images. The seven SBM measures proposed here have been of great utility, clearly outperforming the VAF approach, and achieving results similar to state-of-the-art methodology in the field, while providing a useful two-dimensional representation of the brain. This is especially the case of the VRLBP, which according to the statistical significance analysis, was able to locate major differences in very small regions of the brain.

To end with, we want to remark that the SBM defines a whole framework that can easily be extended with different strategies. We already proposed the layered extension or the neighbourhood helical sampling that we tested with the VRLBP, but many others can be made. Higher order statistics on the set of intensities $V_{\phi,\theta}$, or the computation of GLCMs around the mapping vector can focus on different properties of the structure of the image. We could even, instead of using the HMM paths, use the fibre tracts defined in DTI modalities to compute features that are related to brain connectivity, and could be potential indicators of neurodegeneration, or even apply SBM to other imaging modalities such as PET or SPECT, where the structural information is not always present.

Part III

INCREASING THE SAMPLE SIZE

7

SIGNIFICANCE WEIGHTED PRINCIPAL COMPONENT ANALYSIS

Multicentre studies with structural MRI and functional MRI (fMRI) are increasingly common, allowing for recruitment of larger populations in shorter periods of time. However, the use of images acquired at different sites still poses a major challenge. In addition to logistical difficulties, such as regulatory approvals and data protection, a number of technical and methodological issues can potentially affect the resulting maps, introducing undesired intensity and geometric variance. This issue has been addressed in other neurological conditions, such as AD [92, 109], where group differences are well known, and demonstrating that the impact of a correction for site on the resulting neurobiological differences is relatively small. However, these effects have a stronger impact in psychiatric conditions where the atypical radiological signs on MRI are often subtle and require large samples of patients to observe on-average differences relative to control samples. Recent meta-analyses point to differences being inconsistently reported in schizophrenia [42, 106], psychosis [10, 17], and ASD (using the multi-centre ABIDE database) [21]

These inconsistencies can arise from a variety of variance sources, ranging from the multi-level (phenotypic, neurobiological, and etiological) heterogeneities of the conditions to technical issues that include differences in scanner make, model, manufacturer, static field strength, field inhomogeneities, slew rates and image reconstruction [79], as well as acquisition problems such as within-acquisition participant head motion. Field inhomogeneities are a source of misinterpretation of the data even when the same MRI system manufacturer and model are used [79]. Furthermore, results in [75] demonstrate that a single scanner can change with time, which makes some widely used strategies, for example collecting controls first and patients later, a flawed approach. Recent neuroimaging research on ASD [21] has shown that, while analyses performed on a particular database (acquired on a single platform) could yield coherent regions, the atypical structures are often inconsistent across the wider literature using different databases. Therefore, new methodologies focused on reducing multi-site variance may be potentially helpful in increasing the power to identify the characteristic neurobiological signature of autism, should there be one.

A number of approaches have been proposed to reduce between-site variance. Geometric distortions caused by magnetic fields inhomogeneities have been widely studied [92, 109]. Furthermore, there exist a number of diffeomorphic registration algorithms, such as DARTEL [94] or ANTS [61], intended to reduce inter-subject—and by extension, inter-site—variations. In the case of intensity variance, the issue has already been addressed in the MRC-AIMS database by leveraging sMRI sequences that yield quantitative estimates of relaxation times [83] which have been demonstrated to reduce single-site effects compared to weighted sequences. However, images acquired using this technique still yield between-site differences [27].

To address the problem of intensity variance and improve the homogeneity of the images across different sites, we have proposed a new post-acquisition methodology that enhances derived maps (e.g. grey or white matter volumes) by means of ameliorating site effects [1]. This method, called Significance Weighted Principal Component Analysis (SWPCA), can be applied as part of pre-processing before computing whole-brain or regional statistical analysis. The algorithm proceeds by performing a PCA over the whole database of images and later computing the statistical significance of each component in relation to a categorical variable, in this case the acquisition site. This information is used to reconstruct the datasets using a weighted strategy that effectively reduces intensity inhomogeneities due to site effects.

7.1 Significance Weighted Principal Component Analysis (SWPCA)

The Significance Weighted Principal Component Analysis (SWPCA) is an algorithm to reduce undesired intensity variance introduced by multi-site image acquisition. SWPCA takes any dataset of pre-processed images and decomposes them into their variance components to then provide a corrected dataset where the undesired variance has been reduced. To do so, PCA is applied to each modality in turn to obtain the component scores and component loadings. Since PCA is a data-driven approach, it was only used to decompose the source images, and after this procedure, a one-way ANOVA estimates the relation between each variance component and a given categorical variable, in our case, the acquisition site. The between-site variability in the variance component is then identified by its corresponding p -value. Finally, these p -values are transformed into a weighting matrix Λ that weights the influence of each variance component in a final PCA reconstruction of the corrected maps. The

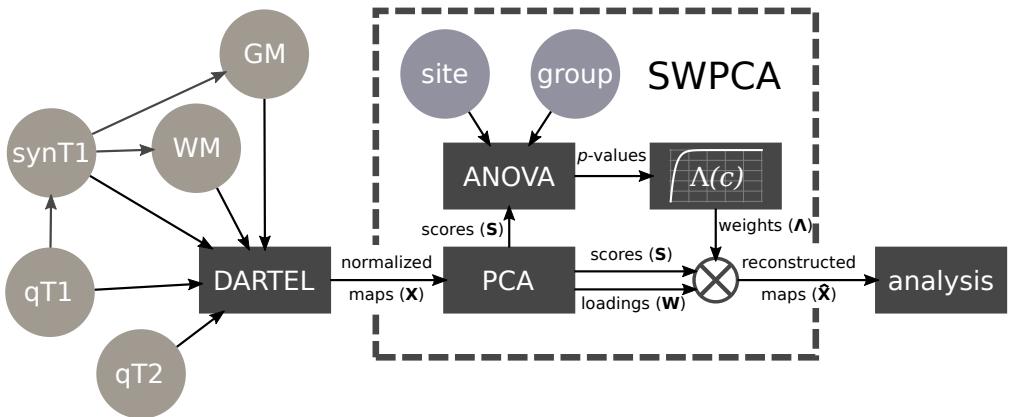


Figure 7.1: Summary of the SWPCA algorithm, along with its context in the pipeline used in this article. Circles represent the input data, both images (green shading) and class (group and acquisition site, purple shading). Rectangles represent the different procedures applied, comprising the DARTEL normalization and registration, the different steps contained in SWPCA, ANOVA and obtaining the weighting function $\Lambda(c)$ - and the subsequent analysis.

procedure is summarized in Figure 7.1. There is also a python class for performing SWPCA at <https://github.com/SiPBA/swpca>.

7.1.1 Principal Component Analysis (PCA)

The first step in the SWPCA algorithm is to perform a PCA decomposition of the dataset into a set of orthogonal components that model the variance present in the images.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations \mathbf{X} of possibly correlated variables, where \mathbf{X} is a $N \times K$ matrix, with N participants (in this case, with one image per participant) and K the number of voxels, into a set of K linearly uncorrelated variables called Principal Components (PC, also known as component loadings or the mixing matrix) \mathbf{W} of size $K \times K$ whose linear combination using a vector of component scores \mathbf{s}_N can perfectly recompose each image.

Intuitively, PCA defines a new space where the first spatial direction is defined so that it explains the maximum variance in the data. The subsequent directions will try to explain the remaining variance in decreasing order. The set of these component scores \mathbf{S} (size $N \times K$) is estimated as:

$$\mathbf{S} = \mathbf{X}\mathbf{W}^\top \quad (7.1)$$

where the columns of \mathbf{W} contain the eigenvalues of $\mathbf{X}^T \mathbf{X}$, the empirical covariance matrix of \mathbf{X} .

This transformation computes a sequence of PCs, maximally explaining the variability of the data while maintaining orthogonality between components. This is done, ideally, by obtaining the EVD of the empirical covariance matrix of the data. The mixing matrix \mathbf{W} is then defined as the set of eigenvectors obtained after applying the EVD to the covariance matrix of the data $\mathbf{X}^T \mathbf{X}$ (assuming that the column mean was subtracted from the data) so that:

$$\mathbf{X}^T \mathbf{X} = \mathbf{W} \Lambda \mathbf{W}^T \quad (7.2)$$

where Λ is a diagonal matrix whose diagonal elements contain the eigenvalues of each principal component, and the mixing matrix \mathbf{W} contains the eigenvectors.

The most popular approach to obtaining the decomposition takes advantage of some analogies between the SVD of \mathbf{X} and the EVD of $\mathbf{X}^T \mathbf{X}$. Since the EVD is often more computationally expensive, the SVD approach is often used. SVD performs the decomposition:

$$\mathbf{X} = \mathbf{U} \Sigma \mathbf{V}^* \quad (7.3)$$

where \mathbf{U} is an $N \times N$ orthogonal matrix, Σ is a $N \times K$ diagonal matrix with non-negative real numbers on the diagonal, and the $K \times K$ unitary matrix \mathbf{V}^* denotes the conjugate transpose of the $K \times K$ unitary matrix \mathbf{V} . Using that decomposition, the covariance matrix

$$\mathbf{X}^T \mathbf{X} = \mathbf{V} \Sigma \mathbf{U}^T \mathbf{U} \Sigma \mathbf{V}^T = \mathbf{V} \Sigma^2 \mathbf{V}^T \quad (7.4)$$

which, comparing with Equation 7.2, shows that we can obtain the mixing matrix as $\mathbf{W} = \mathbf{V}$, the eigenvalues $\Lambda = \Sigma^2$, and the transformed samples of Eq. 7.1 can be rewritten as $\mathbf{S} = \mathbf{XW} = \mathbf{U}\Sigma$.

We obtained both the component scores and estimates of the ‘eigenbrains’ \mathbf{W} . In PCA, the maximum variability in the data (information) is contained within the first components, and the remaining can be considered noise. Therefore, for this chapter, the truncated form of SVD was used such that only the first C components were considered, where most of the variability of the data was concentrated:

$$\mathbf{S}_C = \mathbf{U}_C \Sigma_C = \mathbf{XW}_C \quad (7.5)$$

where \mathbf{S}_C is the set of component scores using the first C components (size $N \times C$). To achieve reasonable performance with minimal information loss, it was assumed that the number of components was the same as the number of

images, $C = N$. Thus, a partial reconstruction of the original signal could be undertaken:

$$\hat{\mathbf{X}} = \mathbf{S}_C \mathbf{A}_C \quad (7.6)$$

where \mathbf{A}_C is the pseudoinverse of the truncated matrix of component loadings \mathbf{W}_C , and $\hat{\mathbf{X}}$ is the reconstructed set of images.

7.1.2 One-Way Analysis Of Variance (ANOVA)

The estimated PCs effectively model the variability of the image dataset. The next step was to assess each PC as a source of inter-site variance with one-way Analysis Of Variance (ANOVA). ANOVA estimates the F-statistic, defined as the ratio between the estimated variance within groups and the variance between groups:

$$F = \frac{MS_{within}}{MS_{between}} = \frac{SS_{within}/(G - 1)}{SS_{between}/(K - G)} = \frac{\sum_i n_i (\bar{Y}_i - \bar{Y})^2 / (G - 1)}{\sum_{ij} (Y_{ij} - \bar{Y}_i)^2 / (K - G)} \quad (7.7)$$

Where MS_{within} and $MS_{between}$ are the mean squares within- and between-groups respectively, G is the number of separate groups (in our case, two), \bar{Y} is the sample mean of a certain feature (in our case, the sample mean of all K values of a given component score), \bar{Y}_i is the sample mean of the features belonging to group $i = 1 \dots G$, Y_{ij} is the j th observation of a feature belonging to group i and n_i is the number of participants in the i th group. The F-distribution allows an easy computation of p-values, given the number of groups and degrees of freedom. The F-statistic and p-values were computed independently for each component score and acquisition site, and then used in the SWPCA algorithm.

7.1.3 Weighting Function

To obtain a set of corrected maps, a new signal matrix of all maps of the same modality, $\hat{\mathbf{X}}$, was estimated with the influence of the PCs with variance related to acquisition site, assessed via the p-values, reduced. To do so, equation 7.6 was modified to include a square matrix Λ (dimension $C \times C$) whose diagonal contains a weight λ_c for each component that depends on its p-value; that is,

$$\hat{\mathbf{X}} = \mathbf{S} \Lambda \mathbf{A} \quad (7.8)$$

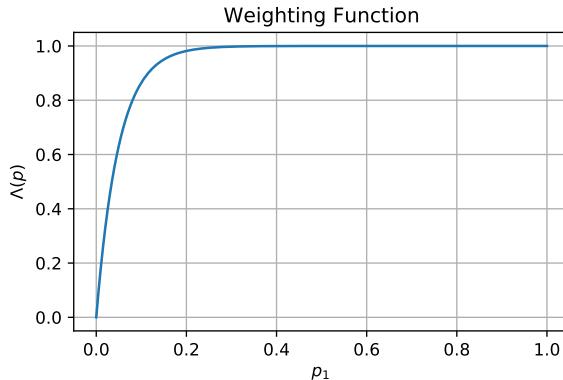


Figure 7.2: Weighting function $\lambda_c(p_c, p_{th})$ used in SWPCA.

The computation of each λ_c , for each component, was performed using the Laplace distribution, modified so that the weights were on the interval $[0, 1]$:

$$\lambda_c(p_c, p_{th}) = 1 - e^{\frac{-p_c}{p_{th}}} \quad \forall p_c \in [0, 1] \quad (7.9)$$

where p_c is the statistical significance of the c_{th} component with respect to the acquisition site and p_{th} is the statistical threshold for significance; that is, $p_{th}=0.05$. A plot of the univariate weighting function $\lambda_c(p_c, p_{th})$ can be found in Figure 7.2. This weighting ensured that most of the components of variance that are not related to the acquisition site are kept unchanged, while at the same time it strongly reduces the influence of components with p -values less than the threshold.

This procedure is illustrated in Figure 7.3, where a boxplot of the distribution of the first four principal component scores is shown. Since we have assumed that substantial differences imply a bigger influence of the acquisition site on the portion of variance modelled by that component, the resulting weight is reduced, and the contribution of that component to the reconstructed signal will be smaller. After computing all weights, most of the sources that are related to the acquisition site (for example, the second and third components with respectively $\lambda_2 = 0$ and $\lambda_3 = 6.14E-6$) have been parsed out while keeping all other sources of variance.

7.2 Evaluation of SWPCA

To validate the effects of the SWPCA algorithm on the inter-site variance, three experiments were undertaken to assess the reduction of the undesired

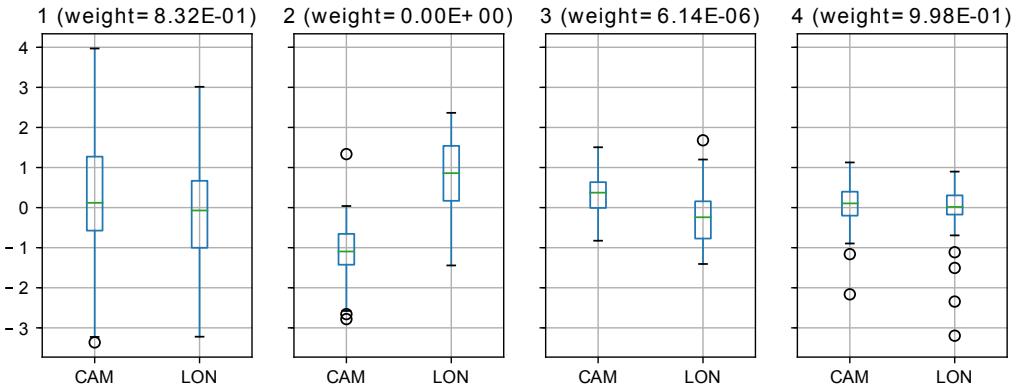


Figure 7.3: Box-plot of the distribution of the component scores at each site of the AIMS-MRI dataset (see Sections 7.3 and A.1.2) in the four first components. We assume that bigger differences between distributions imply a bigger influence of the acquisition site on the portion of variance modelled by that component and therefore, to parse out those differences, the resulting weight will be smaller.

site variance in the original datasets, and its impact on the between-group signal. Two kind of analysis have been performed: a characterization of voxel-wise differences, and a classification analysis.

Voxel-wise differences between groups were characterized using VBM [137], comprising preprocessing (registration, smoothing) and mass-univariate t-test on the smoothed maps from each modality. SWPCA is included (when needed) in this pipeline as a plug in, after the smoothing and before the hypothesis testing. Permutation testing assessed the significance of the relationship between the tested and target variables. A max-type procedure was used to obtain family-wise, whole-brain corrected p -values [169]. Additionally, a Component Based Morphometry (CBM), based on Source Based Morphometry (SBM) [81] was used. This procedure provided Z-maps for visual inspection comparable to those obtained in VBM, by selecting component loadings \mathbf{W} , scaling them to unit standard deviation and weighting their contribution to the final map with their statistical significance, computed using the same permutation inference as in VBM.

A classification analysis was undertaken using a common classification pipeline [13] consisting of preprocessing, feature extraction via PCA and SVC classification. SWPCA is used as a plug-in here as well, after the preprocessing and before the feature extraction step. The classification was validated using stratified 10-fold cross-validation [157].

For each modality independently, the following experiments were performed:

- **Experiment 1:** To demonstrate the ability of the SWPCA algorithm to reduce undesired effects due to acquisition site, the PCA + SVC pipeline was applied to the datasets labelled by acquisition site. Classification accuracy was compared to datasets with and without SWPCA. VBM was then applied to identify the spatial location of the between-site differences. This was undertaken on the whole database (ALL), and subgroups containing only ASD or ASD participants.
- **Experiment 2:** The discrimination ability of each modality, acquired at different sites was assessed by classification performance of individuals from London (LON) and Cambridge (CAM) was separately assessed, using group (ASD and CTL) as the labels.
- **Experiment 3:** To assess the impact of SWPCA on the datasets when characterizing the differences between ASD and CTL groups, the classification pipeline comprising PCA + SVC, as well as VBM and CBM, have been applied to all participants with group as the labels.

7.3 Results for AIMS-MRI Dataset

7.3.1 Experiment 1: Effect of Acquisition Site

The first experiment was to demonstrate the ability of SWPCA to reduce the intensity variance related to acquisition site. To do so, we first performed a VBM analysis in all five modalities (qT_1 , qT_2 , simulated T_1 - weighted Inversion Recovery ($synT_1$), GM and WM) separately, with the uncorrected (without applying SWPCA) and the corrected (after applying SWPCA) maps, using the acquisition site as labels.

To illustrate where the sources of variance of the acquisition sites are located, Figure 7.4 shows a brain t-map of significant ($p < 0.01$, $|t| > 2.57$) GM and WM between-site differences. The biggest reductions in variance were found in qT_1 and $synT_1$ maps, where high variability between acquisition sites, especially in the right hemisphere, was substantially reduced after the application of SWPCA. The reduction in the qT_2 , GM and WM maps was smaller, although noticeable.

To quantify the impact of this variance reduction on the between-groups effects, the classification analysis was undertaken. Higher accuracy values imply that the maps contain site-related patterns that were significant, whereas accuracy close to 0.5 indicates that the site-related variance was small. The test

was applied to ALL, and also to the ASD and CTL subgroups. The classification results are presented in Table 7.1.

Performance results indicate clear advantages of using SWPCA, in particular in the case of qT₁ and synT₁ which were associated with strong site-dependent variance. These results are also consistent with the reduction of significant between-group areas observed in Figure 7.4.

The between-site differences were smaller for GM and WM maps, possibly due their reduced sensitivity. Since fractional occupancy values are abstract, unitless values derived from each image they are less influenced by the acquisition site effects. For qT₂ maps, the site-related differences were greater for the CTL participants than ASD where, according to the classification accuracy, they were nearly indistinguishable. Acquisition site differences were therefore noticeably reduced in the CTL and ALL databases, but not in the ASD.

7.3.2 Experiment 2: Within-site Between-Group Differences

In this second experiment, accuracy, sensitivity and specificity in the between-group comparison were recorded for images acquired from each site and shown at Table 7.2. This is an estimation of the discrimination ability of the different modalities without the influence of the site effects. For all modalities, most of the values are close to a random classifier (~50%), indicative of having either no significant differences between groups, or having spatially heterogeneous patterns of SMRI measures across individuals where mass-univariate approaches are sub-optimal in detecting group differences. It is interesting to note that the London sample contained more between-group differences than those acquired at Cambridge.

7.3.3 Experiment 3: Effect of SWPCA on Group Differences

Finally, group differences were characterised with and without applying site-effects reduction via SWPCA to the five modalities.

Whole-brain VBM analysis was performed on the corrected and uncorrected maps from each modality. Figure 7.5 depicts the brain t-maps of significant ($p < 0.01, |t| > 2.57$) qT₁, qT₂, synT₁, GM and WM between-group differences, using ALL, with the GM+WM mask, before and after applying SWPCA, so that the reduction of site-related variability can be observed. Some of the highlighted areas after applying SWPCA are inconsistent across modalities, with spurious peaks and noise, including a large area around the ventricles in the qT₁ and synT₁ modalities related to some abnormal participants that will be discussed

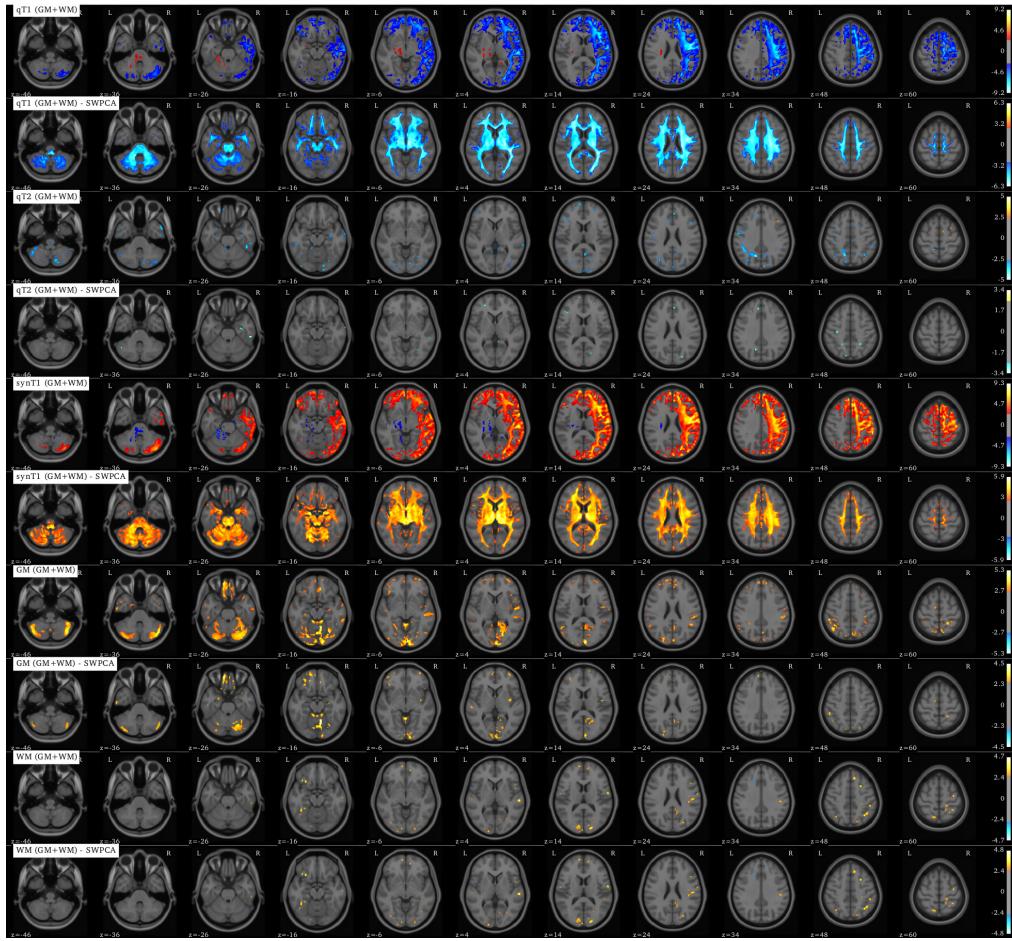


Figure 7.4: Brain t-map (VBM) of significant ($p < 0.01, |t| > 2.57$) GM and WM between-group differences using qT₁, qT₂, synT₁, GM and WM modalities before and after applying SWPCA to remove site effects.

Modality	Mask	all			CTL			ASD		
		no-SWPCA	SWPCA	no-SWPCA	SWPCA	no-SWPCA	SWPCA	no-SWPCA	SWPCA	no-SWPCA
qT ₁	GM+WM	0.875 ± 0.083	0.530 ± 0.130	0.847 ± 0.141	0.543 ± 0.115	0.769 ± 0.145	0.553 ± 0.093			
	GM	0.849 ± 0.085	0.535 ± 0.107	0.835 ± 0.154	0.501 ± 0.090	0.712 ± 0.161	0.575 ± 0.084			
	WM	0.865 ± 0.082	0.447 ± 0.071	0.876 ± 0.128	0.441 ± 0.058	0.813 ± 0.127	0.575 ± 0.153			
qT ₂	GM+WM	0.596 ± 0.128	0.503 ± 0.093	0.615 ± 0.196	0.454 ± 0.075	0.506 ± 0.192	0.476 ± 0.103			
	GM	0.596 ± 0.126	0.493 ± 0.097	0.549 ± 0.187	0.478 ± 0.108	0.497 ± 0.197	0.425 ± 0.091			
	WM	0.612 ± 0.131	0.560 ± 0.128	0.576 ± 0.195	0.550 ± 0.146	0.541 ± 0.185	0.575 ± 0.172			
synT ₁	GM+WM	0.904 ± 0.073	0.563 ± 0.060	0.919 ± 0.100	0.440 ± 0.057	0.807 ± 0.151	0.631 ± 0.098			
	GM	0.879 ± 0.090	0.576 ± 0.035	0.899 ± 0.108	0.526 ± 0.079	0.800 ± 0.145	0.587 ± 0.042			
	WM	0.904 ± 0.076	0.582 ± 0.047	0.894 ± 0.111	0.574 ± 0.038	0.859 ± 0.112	0.468 ± 0.101			
GM	GM+WM	0.595 ± 0.133	0.586 ± 0.141	0.582 ± 0.192	0.566 ± 0.093	0.481 ± 0.169	0.468 ± 0.152			
	GM	0.620 ± 0.141	0.585 ± 0.078	0.604 ± 0.227	0.574 ± 0.038	0.499 ± 0.188	0.525 ± 0.114			
WM	GM+WM	0.659 ± 0.139	0.448 ± 0.066	0.635 ± 0.180	0.507 ± 0.144	0.522 ± 0.206	0.525 ± 0.198			
	WM	0.639 ± 0.124	0.549 ± 0.072	0.578 ± 0.194	0.516 ± 0.126	0.549 ± 0.160	0.526 ± 0.136			

Table 7.1: Between-site classification accuracy (\pm standard deviation) for different modalities and masks without and with SWPCA correction.

Modality	Mask	LONDON			CAMBRIDGE		
		acc.	sens.	spec.	acc.	sens.	spec.
qT ₁	GM+WM	0.603 ± 0.175	0.512 ± 0.260	0.692 ± 0.237	0.504 ± 0.193	0.492 ± 0.276	0.515 ± 0.307
	GM	0.501 ± 0.157	0.440 ± 0.244	0.565 ± 0.245	0.484 ± 0.201	0.488 ± 0.300	0.480 ± 0.327
	WM	0.505 ± 0.174	0.485 ± 0.248	0.526 ± 0.242	0.451 ± 0.197	0.465 ± 0.297	0.435 ± 0.296
qT ₂	GM+WM	0.628 ± 0.168	0.535 ± 0.246	0.719 ± 0.237	0.467 ± 0.181	0.527 ± 0.307	0.417 ± 0.314
	GM	0.539 ± 0.149	0.425 ± 0.220	0.654 ± 0.222	0.491 ± 0.196	0.548 ± 0.316	0.430 ± 0.298
	WM	0.619 ± 0.194	0.585 ± 0.262	0.655 ± 0.250	0.472 ± 0.195	0.448 ± 0.283	0.492 ± 0.290
synt ₁	GM+WM	0.665 ± 0.158	0.578 ± 0.224	0.755 ± 0.238	0.479 ± 0.201	0.478 ± 0.318	0.475 ± 0.316
	GM	0.547 ± 0.159	0.475 ± 0.237	0.622 ± 0.252	0.514 ± 0.218	0.477 ± 0.322	0.555 ± 0.342
	WM	0.515 ± 0.185	0.520 ± 0.288	0.506 ± 0.254	0.509 ± 0.209	0.472 ± 0.317	0.542 ± 0.316
GM	GM+WM	0.513 ± 0.171	0.507 ± 0.252	0.518 ± 0.245	0.488 ± 0.202	0.445 ± 0.318	0.528 ± 0.285
	GM	0.586 ± 0.174	0.610 ± 0.247	0.564 ± 0.270	0.521 ± 0.187	0.522 ± 0.303	0.535 ± 0.289
WM	GM+WM	0.471 ± 0.181	0.455 ± 0.245	0.488 ± 0.278	0.489 ± 0.206	0.502 ± 0.319	0.483 ± 0.314
	WM	0.465 ± 0.174	0.445 ± 0.243	0.484 ± 0.268	0.468 ± 0.210	0.488 ± 0.292	0.448 ± 0.305

Table 7.2: Classification accuracy (acc), sensitivity (sens) and specificity (spec) ± standard deviation for each modality and mask using the participants acquired at the LON and CAM sites.

later. However, there were some areas that were consistent across modalities. Significant areas found across at least 4 of the 5 modalities correspond to the AAL [131] areas of: A) right superior frontal gyrus, Brodmann area 6 ($z=60$); B) the pars opercularis of the left inferior frontal gyrus, Brodmann area 44; C) the pars triangularis of the left inferior frontal gyrus, Brodmann area 45; D) the posterior part of the left middle temporal gyrus ($z=24$); CSF filled spaces on the margins of the ventricles ($z=-6,4,14,24$); and the left crus I of cerebellar hemisphere ($z=-26$).

The complementary CBM (see Section 7.2) analysis was performed on the most significant components. The resulting regions, statistically thresholded with $Z > 2.57$ (corresponding to $p < 0.01$), were superimposed on the MNI template, and are depicted in Figure 7.6. A reduction of significant between-group areas after applying SWPCA is evident in most modalities, but particularly noticeable in the qT₁ and qT₂. In WM no significant regions were observed, neither before nor after SWPCA. The significant regions identified in any modality corresponded to the AAL areas of the CSF filled areas around the ventricles (planes $z=-6, 4, 14, 24$), the right middle temporal gyrus (plane $z=14$) and the left crus I of cerebellar hemisphere (plane $z=-26$). However, none of these regions were repeated over more than two of the modalities, except for the large areas around ventricles that were caused by abnormalities in three participants, which will be discussed later.

Performance results for the classification analysis applied to ALL are shown in Table 7.3. Between-group results were quite similar before or after applying SWPCA, although reducing between-site variance generally reduced the performance towards a random classifier. The results in this table match the overall effects that were found in Figure 7.5, where most spurious significance peaks disappeared after applying SWPCA, but some regions were highlighted. These regions, where SWPCA did not seem to eliminate the significant areas but enhanced them, could be responsible for the accuracy increment in the analysis of the qT₂ modality, and the GM with GM mask.

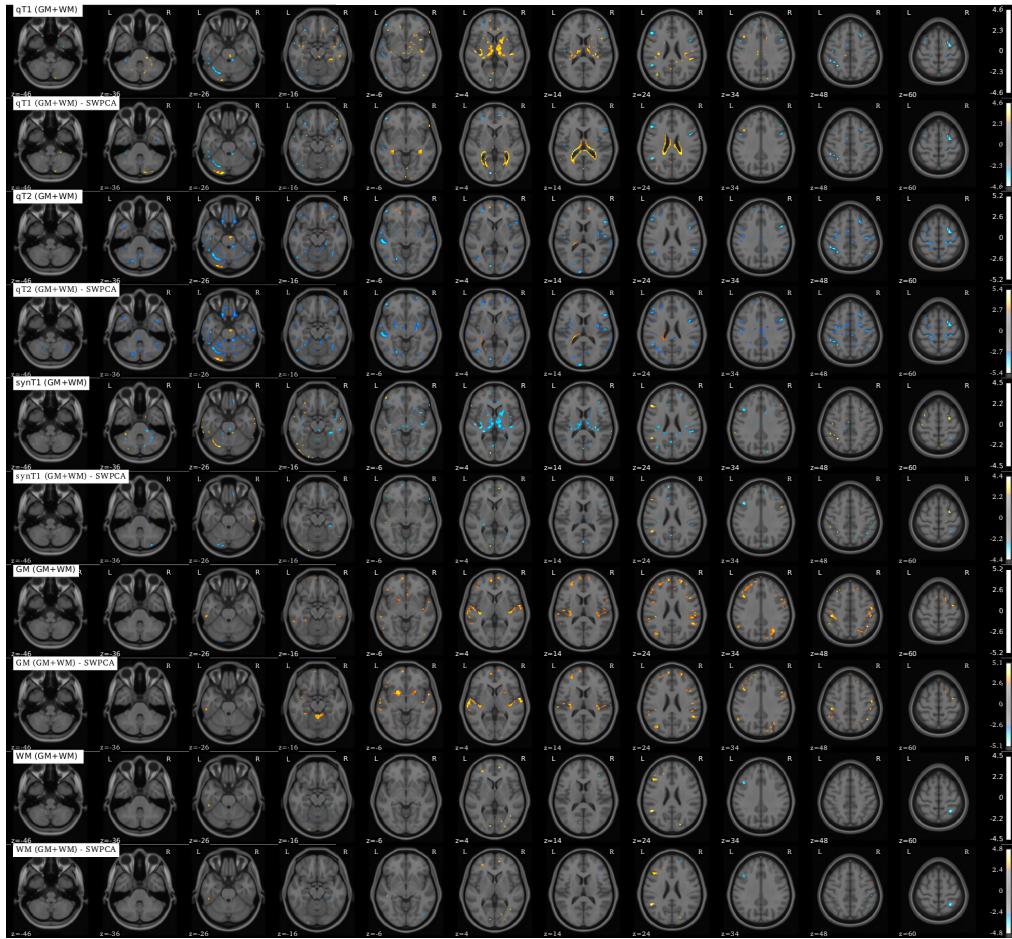


Figure 7.5: Brain t-map (VBM) of significant ($p < 0.01, |t| > 2.57$) GM and WM differences in ASD using qT1, qT2, synT1, GM and WM maps before and after applying SWPCA to remove site effects.

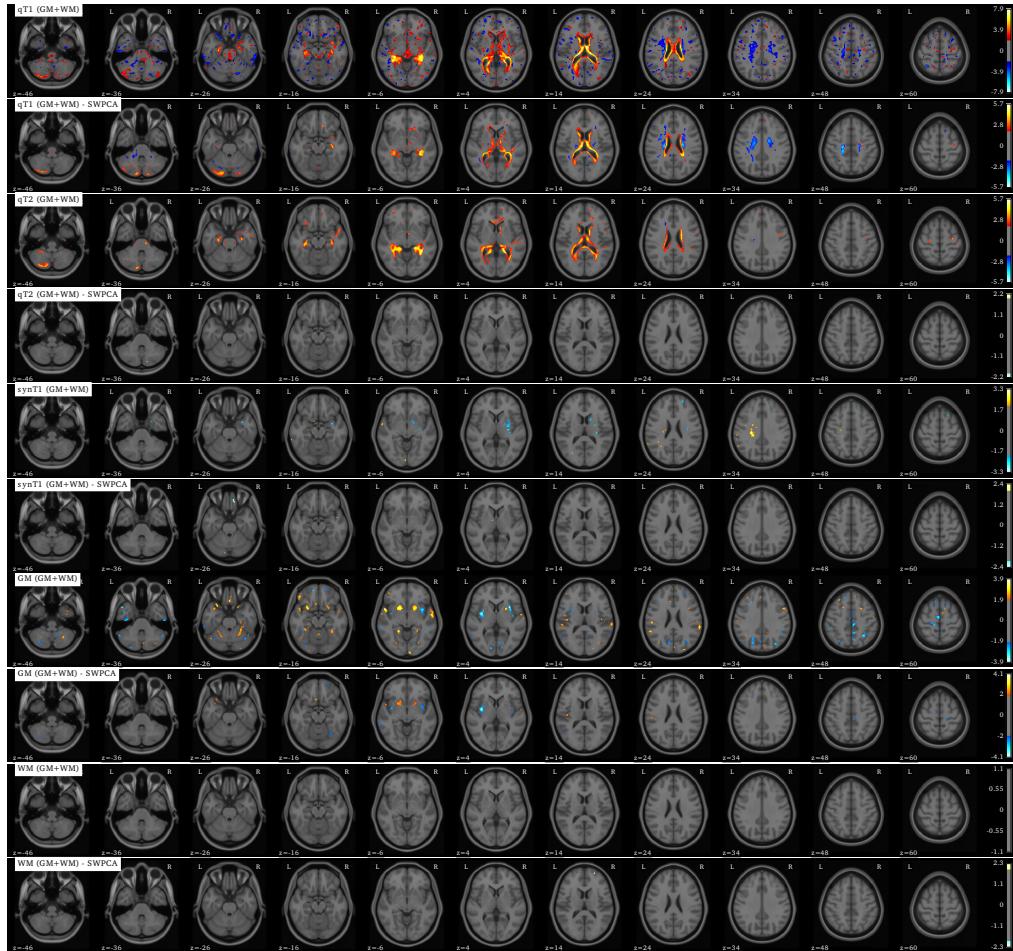


Figure 7.6: Brain Z-map (CBM) of significant ($p < 0.01$, $|t| > 2.57$) GM and WM differences using qT₁, qT₂, synT₁, GM and WM maps before and after applying SWPCA to remove site effects.

Modality	Mask	no-SWPcA			SWPcA		
		acc.	sens.	spec.	acc.	sens.	spec.
qT ₁	GM+WM	0.564 ± 0.123	0.503 ± 0.179	0.625 ± 0.177	0.435 ± 0.123	0.499 ± 0.181	0.371 ± 0.178
	GM	0.523 ± 0.112	0.468 ± 0.162	0.580 ± 0.192	0.458 ± 0.120	0.477 ± 0.187	0.441 ± 0.210
	WM	0.504 ± 0.131	0.475 ± 0.191	0.533 ± 0.194	0.484 ± 0.123	0.511 ± 0.179	0.456 ± 0.194
qT ₂	GM+WM	0.578 ± 0.115	0.487 ± 0.208	0.669 ± 0.178	0.593 ± 0.136	0.546 ± 0.206	0.640 ± 0.194
	GM	0.554 ± 0.135	0.492 ± 0.194	0.614 ± 0.181	0.526 ± 0.144	0.512 ± 0.209	0.543 ± 0.222
	WM	0.516 ± 0.138	0.508 ± 0.198	0.522 ± 0.216	0.499 ± 0.137	0.477 ± 0.209	0.521 ± 0.196
synT ₁	GM+WM	0.596 ± 0.132	0.509 ± 0.194	0.680 ± 0.172	0.577 ± 0.130	0.479 ± 0.208	0.676 ± 0.183
	GM	0.587 ± 0.139	0.509 ± 0.210	0.665 ± 0.169	0.483 ± 0.136	0.489 ± 0.218	0.480 ± 0.200
	WM	0.496 ± 0.139	0.500 ± 0.189	0.492 ± 0.194	0.487 ± 0.134	0.513 ± 0.189	0.461 ± 0.211
GM	GM+WM	0.498 ± 0.120	0.486 ± 0.197	0.507 ± 0.203	0.490 ± 0.123	0.514 ± 0.197	0.465 ± 0.182
	GM	0.574 ± 0.121	0.571 ± 0.189	0.579 ± 0.163	0.593 ± 0.127	0.602 ± 0.172	0.587 ± 0.190
WM	GM+WM	0.499 ± 0.132	0.506 ± 0.189	0.487 ± 0.181	0.521 ± 0.129	0.510 ± 0.209	0.532 ± 0.180
	WM	0.506 ± 0.143	0.488 ± 0.219	0.526 ± 0.197	0.507 ± 0.122	0.521 ± 0.165	0.492 ± 0.193

Table 7.3: Classification accuracy (acc), sensitivity (sens), and specificity (spec) ± STD for the different modalities and masks using ALL, before and after applying SWPCA.

7.4 Discussion

Brain anatomical and functional differences between ASD participants and controls have been explored by a number of previous studies [11, 12, 18, 19, 33]. Many affected structures have been proposed in each of these studies, however as a recent large-scale study points out [21], these are frequently inconsistent throughout the literature. Researchers argue that most of these structures are database-dependent, and since many studies use multi-site acquisition procedures, the variance introduced by each acquisition site is a probable source of Type I errors.

The technical and logistical drawbacks of multicentre studies are widely documented, including participant recruitment procedures [75] and technical effects that range from the usage of different equipment or acquisition parameters [79] to physical changes that affect the performance of MRI scanners across time [75]. There is general recognition that standardization is needed to ensure the uniformity of the acquired maps. Different approaches have been used in large-scale studies, such as ADNI where human “phantoms” were used to perform a preparatory optimisation of MRI scanning platforms [106].

There are two major types of site effects, regardless of their source: geometric distortions and intensity inhomogeneities. In this work, we focused on the latter, since much of the geometric distortion has been eliminated during acquisition (see Section A.1.2), and the DARTEL normalization and registration acts as a homogenizing step, reducing both between-site and between-subject geometric differences, substantially reducing their impact.

Regarding intensity correction, in the MRC-AIMS database used in this study [32, 44], a standardization procedure based on quantitative imaging [83] was used to minimize inter-site variance and improve the signal-to-noise contrast. However, as the between-site analysis in Section 7.3.1 suggests, this strategy still results in variance that makes it easier to distinguish scanning sites than diagnostic groups. For example, when using qT₁ the accuracy for LON vs. CAM classification was >80%, whilst when classifying ASD vs. CTL it was 52%. This marks the substantial effect of site variance on the maps’ intensity distribution, even when the multi-site study employs quantitative imaging protocol on the same model of scanner platform across sites. However, with the inclusion of GM and WM maps, we can observe that the inhomogeneities found on qT₁ or synT₁ barely affected the segmentation procedure.

In this work, the approach we have taken is to perform a multivariate decomposition of each dataset into a number of components that explain different portions of variance. The following step was to identify the components of variance that are due to multi-site acquisition and reduce them. Decomposi-

tion was completed using PCA and then, to identify which of the components were linked to acquisition site, we performed an ANOVA on the component scores. Finally, using the weighting function defined in Sec. 7.3.3, we reconstructed the original signal reducing the undesired variance, in what we called SWPCA. The method has proven its ability in reducing undesired variance, quantifiable by means of the accuracy obtained in a site vs. site classification. In this case, SWPCA reduced the accuracy from >0.8 to approximately ≈ 0.5 , a random classifier, suggesting that most site-related variance was eliminated.

A simpler approach such as applying a voxel-by-voxel ANOVA would also be useful to reduce the acquisition site effects [51]. However, SWPCA is a multivariate approach that still offers major advantages over this voxel-wise algorithm, and similar algorithms have found utility in text document searches [41, 86]. First, PCA models the different sources of variance of the dataset, whereas a simple voxel-wise ANOVA only removes mean site differences, which might result in less statistical power. Secondly, SWPCA is multivariate in nature, where each component contains information that potentially affects all voxels. Together, these two features allow SWPCA to identify the components linked to the undesired effects, and reduce their impact with a weighted reconstruction approach, reducing the general variance related to the acquisition site. However, this increased power reveals a major drawback: SWPCA needs at least a moderate number of participants to work properly. That is the reason why we cannot apply SWPCA to databases such as ADNI [106] or ABIDE [19], where the number of participants acquired at each site is small, or to the six travelling phantoms used in the calibration of the MRC-AIMS study.

There exist a number of similar multivariate methods that model the influence of categorical variables, such as the well-known PLS algorithm [69] or Surrogate Variable Analysis (SVA) [99]. In the first case, both PLS and SWPCA take categorical variables \mathbf{Y} along with the data \mathbf{X} as inputs to partition the influence of these into components. However, the most significant difference is the underlying model. Whilst SWPCA estimates the principal components blindly using their variance, which is what we aim to reduce, and performs an ANOVA afterwards, PLS uses the categorical variable in the computation of the covariance matrix and then estimates the components.

On the other hand, SVA, used for gene expression studies [99], is more comparable to SWPCA. The SVA algorithm uses a number of decomposition and significance estimation steps to construct a set of surrogate variables; that is, variables that account for the unmodeled variance and expression heterogeneity. While similar to SWPCA in the steps used (i.e. SVD decomposition and significance estimation), their approaches are fundamentally different. SVA constructs a higher complexity model that starts by eliminating the contribu-

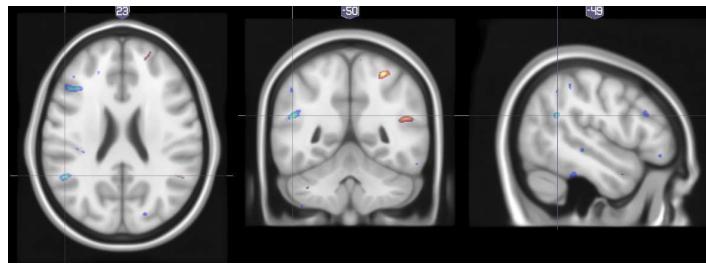


Figure 7.7: Location of the significant region labelled D (posterior part of the superior temporal gyrus) within the MNI template.

tion of primary variables to produce a number of unknown hidden (surrogate) variables, whereas SWPCA is intended to reduce complexity by producing variance-reduced maps to reduce the influence of previously known, but unconsidered, variables and facilitate a subsequent analysis focused only on the relevant variables.

Focusing on the VBM results, after performing the site-effects removal by SWPCA significant between-group differences were noted in five areas: A) the right superior frontal gyrus; B) the pars opercularis of the left inferior frontal gyrus; C) the pars triangularis of the left inferior frontal gyrus; D) the posterior part of the left middle temporal gyrus; and E) the left crus I of cerebellar hemisphere. The first three regions are within Brodmann areas 6, 44 and 45. However, when examining the projection of the region D onto the MNI template (see Figure 7.7), it is also located in the posterior part of the left superior temporal gyrus. Therefore, D corresponds closely with the region between Brodmann areas 22 and 39, the Temporo-Parietal Junction (TPJ), with negative t-value at the left side (containing Wernicke's area) and positive t-value at the right side.

The role of these regions in autism has received much attention. Brodmann areas 44 and 45, that together make the Broca's Area (of importance in speech production and a proposed part of the human mirror neuron system [112]), is a region where mirror neuron dysfunction has been consistently reported in ASD-affected children [105] and adults [28, 88, 107]. Wernicke's area, contained in the left TPJ, is also linked to language, and has been associated with ASD in several works [28, 107]. Additionally, the right TPJ has been proposed as related to mentalizing and has been repeatedly implicated in autism [114], including a fMRI study of a subsample of this same AIMS dataset [57]. The right superior frontal gyrus (region A) is more equivocal, with some studies [44, 65] reporting abnormalities in this area, while others [26, 107] report no

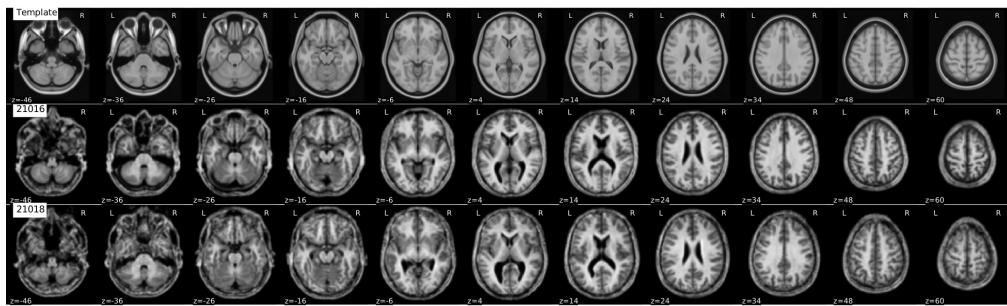


Figure 7.8: The template used in this work compared to two of the participants with abnormal ventricle size (21016 and 21018). Atrophy of the cerebellum in participant 21016 can also be appreciated, responsible for some of the ‘highlighted’ areas in qT₁, qT₂ and synT₁ t-maps (see Fig. 7.5).

significant differences. Our analyses reveal no differences in the insula and amygdala, brain structures frequently linked to autism.

Some regions, particularly in qT₂, synT₁ and segmented GM maps show potentially spurious significance peaks around the ventricles and especially in the left crus I of cerebellar hemisphere (region E). After examining the database, two individuals had appreciable structural abnormalities in the form of abnormal ventricle size and cerebellar atrophy, as can be seen in Figure 7.8. It is possible that these participants influenced the computation of the t-maps, and therefore are responsible for the significance in region E and areas surrounding the ventricles and, since they are part of the LON subdataset, could also be responsible for the increased classification accuracy of the quantitative T₁ and T₂, and the synthetic T₁ maps in this sub-dataset.

After observing the influence of these participants on the computation of the t-maps, we can assume that most of the structural differences in ASD are so subtle that the influence of just one or two images can impact on the final results. This, along with the poor performance of the classification pipeline presented in Section 7.2, dramatically reduces the significance of the aforementioned t-maps. Therefore, the existing evidence leads to the conclusion that ASD presents as either undetectable structural differences or, more likely, with such heterogeneous differences that are difficult to establish a common pattern even after reducing the variance introduced by acquisition site.

It may be the case that cohorts of individuals examined at different sites are somehow systematically biased towards a specific type of patient (in ways that we cannot see simply based on phenotypic information), then site-related intensity variability is also enriched with important variability about nested autism subgroups. So with any technique trying to remove the site-related

inhomogeneity, the subgroup information could also be removed. Together, the evidence supports the claim that defining meaningful subgroups based on different measures, such as genetic profiling, clinical co-morbidities or sensory sensitivities, is the most urgent next step for ASD research [21].

8

SIMULATION OF FUNCTIONAL BRAIN IMAGES

In Chapter 7, we proposed a novel methodology for eliminating inhomogeneities due to acquisition site in multi-centre datasets. This is a good approach for MRI images, but still requires a large recruitment procedure and individual analyses and assessments. In this chapter, we propose a simulation procedure (see Figure 8.1) intended to synthesize completely new samples of an existing database that share the same properties as the original set of subjects belonging to a certain class.

Our system focuses on nuclear imaging techniques such as PET or SPECT, using PCA to project the dataset to the eigenbrain space and a statistical distribution estimator to model the distribution of the brains and classes in this new space. Once the statistical distribution of the classes is known, we can generate new coordinates on the eigenbrain space belonging to the same class, that can be finally projected to the brain space.

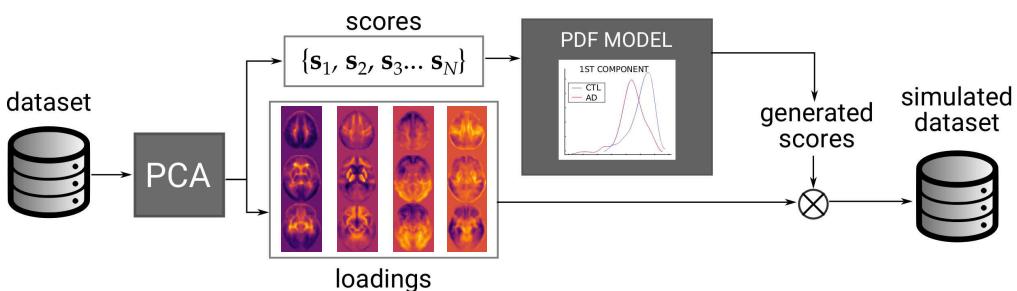


Figure 8.1: Schema of the brain image synthesis algorithm.

We have evaluated our proposed system on two modalities: FDG-PET images from the ADNI-PET dataset and SPECT-DaTSCAN from the PPMI-DAT database. We have tested the images under three different experiments that estimate the relation between the simulated and the real images, the ability of the simulated dataset to predict the outcome of real datasets, or the uncertainty introduced by the simulation procedure.

For our brain synthesis algorithm to work, the source datasets must share certain characteristics: to be spatially normalized (which we did by using the SPM8 software) and intensity normalized, using normalization to the maxi-

mum strategy, which proved to be very useful in most applications [25, 47] (see Chapter 3 for more details).

8.1 Simulation Methodology

8.1.1 PCA Decomposition

The first step of the simulation pipeline is the analysis of the dataset that we want to use as a reference. In order to do so, we project the dataset from its original image space to a lower dimensional space using PCA, a very extended tool for analysis and feature extraction in neuroimaging [13, 56], and that was already presented at Section 7.1.1.

In this chapter, we use the component score matrix \mathbf{S} , or the coordinates of the database samples in the eigenbrain space, to estimate per-class and per-component statistical distributions from which we can generate new coordinates. For efficiency, and without loss of generality, we will use only the first C components of the PCA decomposition.

8.1.2 Density Estimation

Once the images have been projected to the eigenbrain space, we want to generate new samples in this new space, in order to synthesize new images. For this purpose, we assume that the coordinates of the subjects of a certain class in the eigenbrain space are different realizations of a random process with a given PDF.

In order to estimate the PDF of the process that generates the coordinates of each class, we use two different density estimation procedures: the parametrical model of this distribution using a Multivariate Normal distribution (MVN), and the empirical Kernel Density Estimation (KDE).

8.1.2.1 *Multivariate Normal distribution (MVN)*

The MVN, also known as Multivariate Gaussian Distribution, is a generalization of the random normal distribution to n dimensions.

Let us note \mathbf{S}^c the matrix containing only the coordinates in the eigenbrain space of individuals belonging to class c. That way, $\mathbf{S}^{i,c}$ would be the vector containing the i^{th} coordinate of \mathbf{S}^c , a vector with mean $\mu^{i,c}$ and covariance matrix $\Sigma^{i,c}$. The PDF of that combination of coordinate and class would be:

$$\hat{f}_{mvn}^{i,c}(x) = \frac{1}{(2\pi)^{N/2} |\Sigma_{i,c}|^{1/2}} \exp\left(\frac{(x - \mu^{i,c})^\top \Sigma_{i,c}^{-1} (x - \mu^{i,c})}{2}\right) \quad (8.1)$$

8.1.2.2 Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) is an increasingly used method to estimate the empirical PDF of a set of data [50, 64]. The KDE performs the PDF estimation for the i^{th} coordinate of class c as:

$$\hat{f}_{kde}^{i,c}(x) = \frac{1}{N_c} \sum_{l=1}^{N_c} K_h(x - S_l^{i,c}) = \frac{1}{N_c h} \sum_{l=1}^{N_c} K\left(\frac{x - S_l^{i,c}}{h}\right), \quad (8.2)$$

where $l = 1, \dots, N_c$, with N_c the number of subjects belonging to class c , and $S^{i,c}$ contains the i^{th} coordinate of S^c , as in the latter section. The kernel $K(x)$ is a function of \mathbb{R}^n that must define a probability:

$$\int \dots \int_{\mathbb{R}^n} K(x) dx = 1 \quad (8.3)$$

It also must be centred:

$$\int \dots \int_{\mathbb{R}^n} x K(x) dx = \mathbf{0} \quad (8.4)$$

and its covariance matrix must be close to identity:

$$\forall \mathbf{u} \in \mathbb{R}^n, \|\mathbf{u}\| = 1 \quad \int_{\mathbb{R}} t^2 K(t\mathbf{u}) dt \approx 1 \quad (8.5)$$

In this chapter, we use a gaussian kernel $K(x) = 1/(2\pi) \exp(-\frac{1}{2}x^2)$ for the estimate. Estimation of the bandwidth $h > 0$ is performed using the diffusion approximation proposed by [64].

8.1.3 Brain Image Synthesis

After estimating the empirical PDF of the coordinates, we aim to generate a new set of coordinates for class c \hat{S}^c that match the distribution of the originals. To do so, we compute the CDF from the PDF that we estimated previously as:

$$F(x) = \int_{-\infty}^x f(t) dt \quad (8.6)$$

Afterwards, we can use a random number generator to provide uniformly distributed random numbers in the interval $[0, 1]$. These numbers are in the

range of the CDF (from 0 to 1), and therefore we can consider them as $F(x)$, from which we could obtain the value x . In practice, we perform a numerical approximation to the problem, in which we calculate the full CDF in a wide range of x , and then interpolate the value of x using the generated $F(x)$ as query point.

This procedure is repeated for all coordinates $i = 1, \dots, K$ G times, where G is the number of subjects of class c that we want to synthesize. Then, the new set of images can be reconstructed using the eigenbrain basis \mathbf{W} and the newly matrix of scores $\hat{\mathbf{S}}^c$.

$$\hat{\mathbf{X}}^c = \hat{\mathbf{S}}^c \mathbf{W}^{-1} \quad (8.7)$$

The code used for this brain image synthesis can be freely downloaded at <https://github.com/SiPBA/brainSimulator>.

8.2 Experimental Setup

Validation of a simulated dataset is still a matter of discussion. Most works exploring that possibility [144, 164] use visual analysis to validate the results. In addition to this, we chose an empirical approach to validation, based on the classification of samples, which will give us quantitative results. The following experiments have been performed:

- **Baseline:** First of all, we estimate the performance of both the original and the simulated datasets using a VAF [120] classification analysis, under different scenarios (depending on the classes available in each dataset). This will be used as a baseline in the following experiments. An additional SPM [98] analysis is performed on both the original and a simulated dataset.
- **Experiment 1: Predictive power of the simulated images:** To test whether the simulated images can be considered similar to the original images, we try to predict real-world samples with a set of simulated images. This is done within a cross-validation loop, where the original training set is used to generate a new training set of completely new images, and after the SVC training with this simulated set, it is tested against the original test set.
- **Experiment 2: Independence of the simulated images:** To test the independence of a generated dataset, we use the opposite approach. Within the cross-validation loop, we first establish the training set of the original images, and the classifier is trained using these. Then, we test the

classifier against either this same training set or a simulated dataset, generated using the same training set. If our simulated images are independent from the originals, the performance of the system should decrease substantially.

In the classification analysis, we use a SVC with linear kernel. To estimate C , we perform a grid search in an inner cross-validation loop using the training set. Finally, the following performance values are provided: accuracy (acc), sensitivity (sens), specificity (spec) and their standard deviation (SD).

Additionally, a SPM analysis [98] has been performed on both the original and the simulated datasets. Mass-univariate two-sample t-maps are provided in Figures 8.7, 8.8, 8.9 and 8.10, using a FWE correction t-threshold for a $p < 0.05$, with no masking applied.

8.3 Results

8.3.1 Baseline

In this experiment, we compute the baseline performance of the original and the simulated images, by applying it to the ADNI-PET and the PPMI-DAT dataset. We have used all subjects in the computation of these values, and 200 simulated images from each class in their evaluation. We show these results at Table 8.1.

Regarding the AD dataset, we can see that the maximum accuracy can be obtained when differentiating between AD and CTL subjects, as it could be expected. Since MCI is an intermediate class where there exist cognitive impairment but not enough to be considered AD, this class should be closer to both AD and CTL. That is what we see when under the MCI vs CTL and MCI vs AD scenarios, achieving a lower (but still not bad), similar performance.

As for the simulated images derived from the ADNI-PET dataset, there exist two very different behaviours. The MVN estimator produces highly differentiable images under all scenarios, even in those including MCI. There is even an almost perfect accuracy obtained in AD vs CTL. Still, the behaviour of the original dataset, in which the AD vs CTL achieves higher performance than the others is repeated. For its part, the KDE estimation of density produces more overlapping classes. In the best case (AD vs CTL), it only achieves up to 80% accuracy. However, the scenarios including MCI perform similarly to the original dataset.

Database	Est.	Scenario	acc (\pm SD)	sens (\pm SD)	spec (\pm SD)
ADNI-PET	Orig.	AD vs CTL	0.882 ± 0.012	0.865 ± 0.091	0.901 ± 0.118
		MCI vs CTL	0.698 ± 0.042	0.791 ± 0.064	0.504 ± 0.179
		MCI vs AD	0.702 ± 0.117	0.444 ± 0.219	0.822 ± 0.258
	MVN	AD vs CTL	0.997 ± 0.007	0.995 ± 0.015	1.000 ± 0.011
		MCI as CTL	0.985 ± 0.016	0.969 ± 0.033	1.000 ± 0.027
		MCI as AD	0.979 ± 0.022	1.000 ± 0.000	0.959 ± 0.037
PPMI-DAT	KDE	AD vs CTL	0.800 ± 0.037	0.800 ± 0.067	0.800 ± 0.089
		MCI vs CTL	0.767 ± 0.061	0.770 ± 0.110	0.765 ± 0.093
		MCI vs AD	0.742 ± 0.060	0.755 ± 0.096	0.730 ± 0.105
	MVN	Orig. PD vs CTL	0.923 ± 0.057	0.929 ± 0.090	0.918 ± 0.088
		MVN PD vs CTL	0.997 ± 0.007	0.995 ± 0.015	1.000 ± 0.011
		KDE PD vs CTL	0.842 ± 0.038	0.840 ± 0.083	0.845 ± 0.069

Table 8.1: Baseline experiment, in which we evaluate the performance of a VAF system under the different scenarios of the two datasets.

On the other hand, we take a look at the Parkinson’s Progression Markers Initiative (PPMI) dataset. In this case, we have highly specific functional imaging, in which the dopaminergic level (cause of PD) is assessed. Therefore, the application of VAF to the original dataset yields high performance ($>90\%$ accuracy, sensitivity and specificity). The images synthesized using MVN modelling achieve almost perfect accuracy, whereas the KDE images, given their performance, might be more heterogeneously distributed. We will comment on this at the discussion.

8.3.2 Experiment 1: Predictive Power of the Simulated Images

Results for the first of the experiments are shown at Table 8.2. We have applied this experiment to the ADNI-PET and the PPMI-DAT datasets, using the multivariate normal estimator and the KDE. For each cross-validation iteration we synthesize 200 samples of each class.

Database	Est.	Scenario	acc (\pm SD)	sens (\pm SD)	spec (\pm SD)
ADNI-PET	MVN	AD vs CTL	0.852 ± 0.078	0.804 ± 0.151	0.900 ± 0.137
		MCI vs CTL	0.688 ± 0.067	0.743 ± 0.108	0.572 ± 0.169
		MCI vs AD	0.675 ± 0.121	0.468 ± 0.192	0.774 ± 0.224
	KDE	AD vs CTL	0.770 ± 0.114	0.747 ± 0.153	0.790 ± 0.169
		MCI vs CTL	0.642 ± 0.044	0.726 ± 0.098	0.472 ± 0.188
		MCI vs AD	0.672 ± 0.081	0.484 ± 0.142	0.760 ± 0.187
PPMI-DAT	MVN	PD vs CTL	0.948 ± 0.041	0.948 ± 0.055	0.946 ± 0.064
	KDE	PD vs CTL	0.914 ± 0.082	0.916 ± 0.113	0.909 ± 0.121

Table 8.2: Performance of Experiment 1, demonstrating the predictive ability of the simulated images over the real dataset.

In general, it can be seen that for both datasets, using the MVN estimator yields a better performance in predicting the real-world images, with higher accuracy, sensitivity and specificity. This holds for both ADNI and PPMI datasets.

As for the ADNI-PET dataset, we tested three different scenarios: AD vs CTL, MCI vs CTL and MCI vs AD. The higher performance is obtained in the AD vs CTL, but when including MCI subjects, the results vary. In the case of the MVN

estimator, the MCI vs CTL scenario performs slightly better than the MCI vs AD. On the other hand, when using the KDE, MCI vs AD obtains better results than MCI vs CTL. However, the big difference among them is that, whereas the MVN estimator achieves similar performance to the baseline, the images simulated with the KDE estimator lead to smaller predictive power.

In the PPMI dataset, the differences are subtle. While the MVN estimator achieves even higher performance than the baseline for the original dataset, the images synthesized using the KDE estimator have a predictive power similar to the original set.

8.3.3 Experiment 2: Independence of the Simulated Images

With this experiment, we aim to check whether the simulated images are independent enough from the original set so that a trained classifier can be confound by them. To do so, at each cross-validation iteration, we use the original training set as test set, assuming that it will yield almost perfect performance. Afterwards, to test whether the simulated images are different from these or not, we generate an equivalent number of subjects of both classes, and evaluate how the test performance loss. Results are provided at Table 8.3.

With this experiment, we can see that the MVN estimator produces no performance loss, when compared with the original test set. Conversely, with the KDE estimator, the differences are far more significant, with a performance loss of more than 0.15 in accuracy in both datasets. That means that the images simulated using a KDE estimator are far more independent from the original set than those simulated using the MVN. This occurs in both datasets, which gives us an idea of which method generates more independent images.

8.4 Discussion

In this chapter we propose a brain image synthesis algorithm that can analyse a dataset, extract its most relevant characteristics and then simulate new images that share the same properties. The algorithm is based upon PCA, which defines a new common space for each dataset, in which the individual images are represented as points. Therefore, by analysing the distribution of these points and generate new points belonging to each distribution, we can do the inverse transformation of these points to images.

The use of PCA for feature extraction neuroimaging in components is widely documented [4, 13, 56]. In AD, the computed eigenbrains model different

Database	Est	Scenario	Test set	acc ($\pm SD$)	sens ($\pm SD$)	spec ($\pm SD$)
MVN	AD vs CTL	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000
		Simulated	0.994 \pm 0.003	0.988 \pm 0.006	1.000 \pm 0.007	
ADNI-PET	MCI vs CTL	Original	0.999 \pm 0.001	0.998 \pm 0.002	1.000 \pm 0.001	
		Simulated	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
KDE	MCI vs AD	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
		Simulated	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
PPMI-DAT	AD vs CTL	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
		Simulated	0.811 \pm 0.014	0.791 \pm 0.021	0.832 \pm 0.028	
MVN	MCI vs CTL	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
		Simulated	0.823 \pm 0.018	0.843 \pm 0.024	0.801 \pm 0.033	
KDE	MCI vs AD	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
		Simulated	0.820 \pm 0.014	0.799 \pm 0.022	0.841 \pm 0.030	
PPMI-DAT	PD vs CTL	Original	1.000 \pm 0.000	1.000 \pm 0.000	1.000 \pm 0.000	
		Simulated	0.998 \pm 0.002	0.997 \pm 0.003	1.000 \pm 0.003	

Table 8.3: Performance of Experiment 2, proving the independence of the simulated images with respect to the originals.

features that have been associated with the progression of the disease, and the contribution of each eigenbrain (or the coordinates of each subject in the eigenbrain space) can effectively model the advance of the disease in many works. Therefore, it was the most straightforward tool to be used in this chapter.

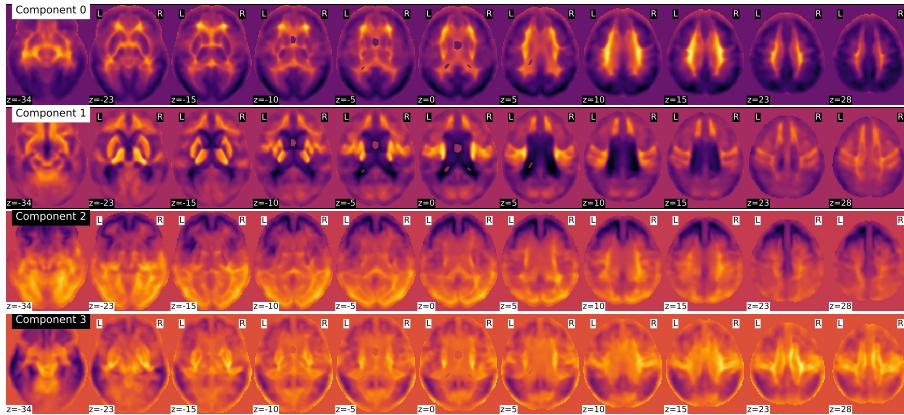


Figure 8.2: Illustration of the four first eigenbrains in PCA.

In Figure 8.2 we show the first four eigenbrains for the ADNI-PET dataset. We can see interesting patterns such as the contribution and size of ventricles in Component 0, the contrast in the cingulate gyri and precunei in Component 1, that have been widely documented [56, 120]. We can also recognise the contrast in glucose metabolism at the occipital lobe and some temporal and parietal negative influence encoded in Components 2 and 3. These features prove that the computed eigenbrains are representative of independent structures and activities that together can positively influence the synthesis of new brain images via a correct parametrization and estimation of the component scores.

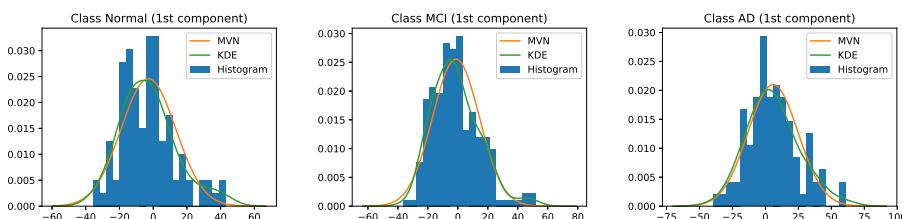


Figure 8.3: Comparison between the MVN and KDE PDF estimation methods for the first component and the three groups, setting the histogram as reference.

Here we compare the two different estimation methods proposed: a Multivariate Normal distribution (MVN) and the Kernel Density Estimation (KDE) via diffusion. An accurate estimation of the distribution of the scores belonging to a certain class is paramount to obtain a reliable brain image synthesis. In Figure 8.3, we compare the different estimation methods for the three AD classes (in the first eigenbrain), using the histogram as a reference. We can see that, while the KDE adapts more to the data, the MVN is more general. This is for the first eigenbrain, but in Figure 8.4, we can see the distribution of the first four scores in all groups, in which we can see that some of them (especially for eigenbrain 1 and 3), the distributions of some groups are less gaussian, and therefore, KDE adapts better to this. Conversely, KDE models each component independently, whereas the MVN creates a parametric distribution of the scores of all eigenbrains, which could be considered an advantage.

A visual analysis of the synthesized images in Figures 8.5 and 8.6 reveals that the new images preserve similar characteristics of the original datasets. For example, it is easy to appreciate differences in glucose metabolism in Figure 8.5 typically associated with AD, such as a smaller activity at the temporal lobe or the striatum [56, 120]. In the simulated DaTSCAN images (Fig. 8.6) differences in shape and intensity of the striatum, and bilateral differences [25, 45, 60] can also be noticed, using both MVN and KDE modelling.

To deepen the analysis of the synthesized figures, we will perform a SPM analysis [98], using the SPM12 software. In Figure 8.7, we can look at the differences (*t*-maps, FWE corrected, $p < 0.05$) between AD and CTL images with the original, and two synthesized datasets with 200 samples per class, using either MVN or KDE modelling. In these maps, the differences are located in similar places in both synthesized databases, that are as well, although less intensely, represented in the SPM analysis of the original dataset. The aforementioned regions such as the precuneus, angular, mid-temporal gyrus, hippocampus, amygdala, among others, are represented in both the original and synthesized datasets, in both the CTL vs AD (fig. 8.7) and MCI vs AD (fig. 8.9) scenarios, with a special mention to the cingulum, which also is the main difference in the MCI vs CTL (fig. 8.8).

In the PPMI dataset, the SPM analysis revealed (Figure 8.10) that the main differences are located in the posterior part of the striatum, specifically at the posterior part of the putamen and globus pallidus. This behaviour is consistent in both original and synthesized images, although with more statistical significance in the case of the MVN model, and a more homogeneous distribution of the negative differences under the KDE model.

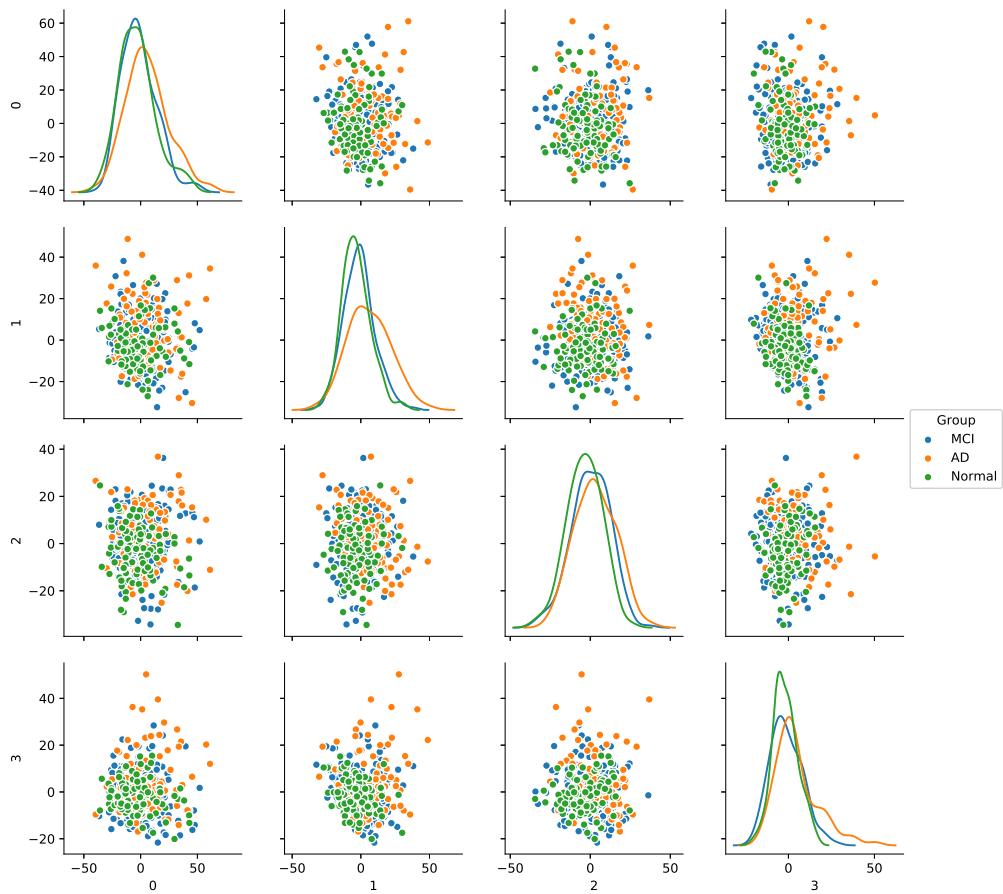


Figure 8.4: Pairplot plotting the coordinates of the subjects in the ADNI-PET database over the four first eigenbrains. In the diagonal, distribution of the classes estimated using KDE for each eigenbrain.

As for the classification analysis, first we will compare the baseline performance of the simulated dataset with the original. From Table 8.1, we can infer that MVN produces minimally overlapping classes, whereas KDE yields more similar images. This might be an indication of the more realistic distribution of classes when using KDE modeling. Experiments 1 and 2 will give us more information on this.

First, in Experiment 1, we tested the predictive power of the simulated images. In this case, the SVC is trained with a set of simulated images, and then real images from the original dataset are predicted. In this case, we obtain higher accuracy when using MVN modelling. This could point to either a more similar distribution of the synthesized images on the eigenbrain space,

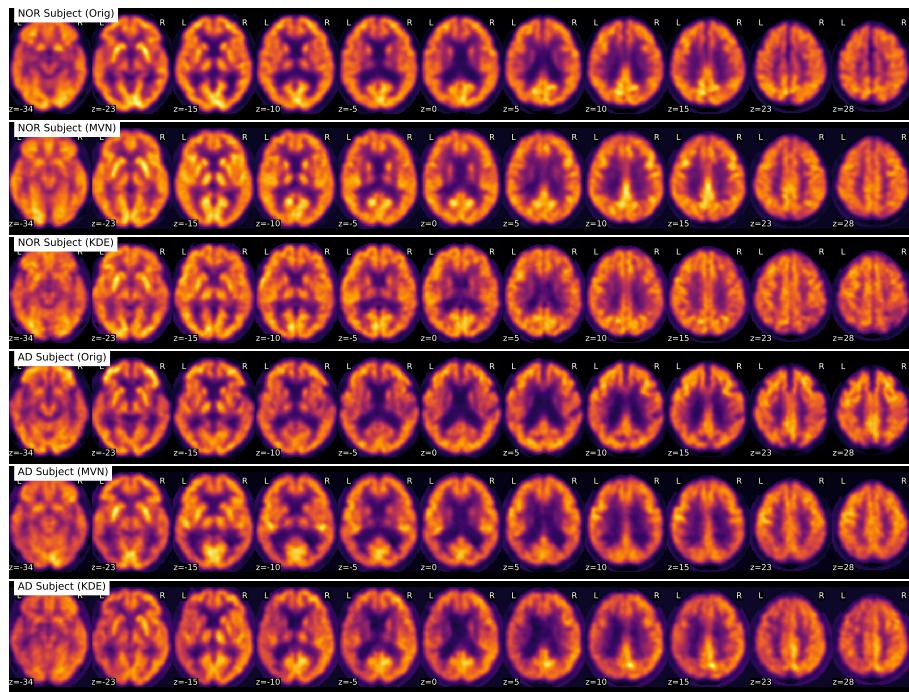


Figure 8.5: Examples of some original and simulated subjects from the ADNI-PET dataset.

or a better definition of the support vectors of the decision plane of the classifier. In the first case, that would mean that the images synthesized using MVN would be more similar to the original dataset. On the other hand, the decision plane could be better defined due to more concentrated classes, as hinted in the baseline experiment. For its part, the KDE synthesis produces more disperse classes, which encompass the original dataset, degrading its performance in predicting the original dataset.

Finally, Experiment 2 demonstrates the degree of similarity between the original and the synthesized images. To this purpose we train and test the classifier with the same subset of the original database. Then, we use that very subset to synthesize a new test set. The performance of this synthesized test set will give us an idea of how similar to the original dataset are our simulated images. From this experiment, we can see that the images simulated using MVN achieved almost perfect accuracy when derived from the original training set, which could mean that there is almost no independence between the original and the simulated set. Conversely, the KDE modelling produces significantly different images, in which the test performance decreases more than 15% with respect to the original. This may indicate that these images are

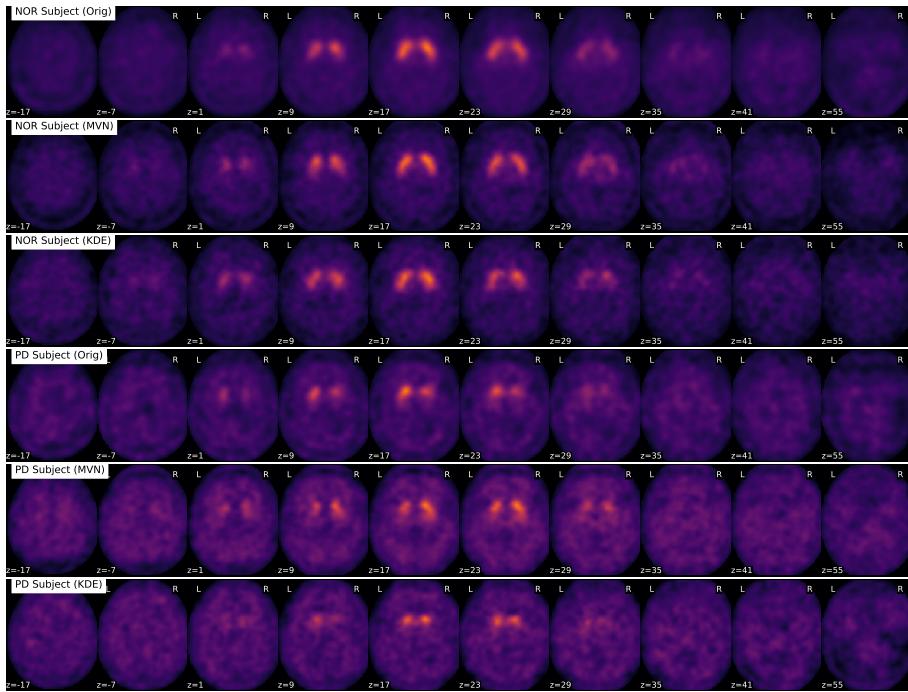


Figure 8.6: Examples of some original and simulated subjects from the PPMI dataset.

highly independent from the original dataset, although belonging to the same classes, as they achieved performances similar to the real-world baseline.

In conclusion, we can assume that the methodology producing the most real world images is the KDE modelling. The images synthesized using this method are more similar to the real world images, in which we have more heterogeneous classes, and are in fact, less dependent on the dataset used as reference. However, all the results provided here prove that the simulated images can be safely used to increase sample size without losing predictive power, and the ones using KDE can even introduce new behaviours that could compensate the false positives due to the small sample size problem. And these images could even be used in training environments for educational purposes, which is an inviting possibility.

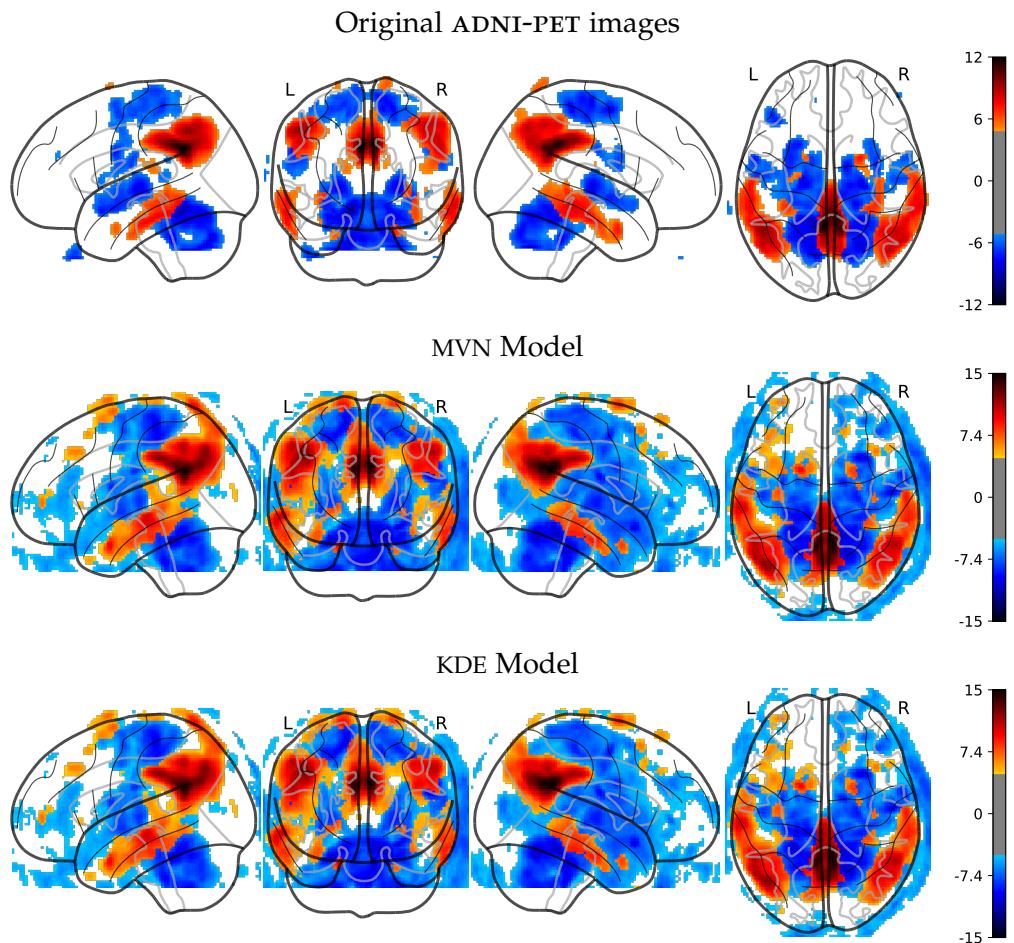


Figure 8.7: SPM Analysis of the ADNI dataset (AD vs CTL), FWE corrected, with $p = 0.05$, for the original and the simulated images.

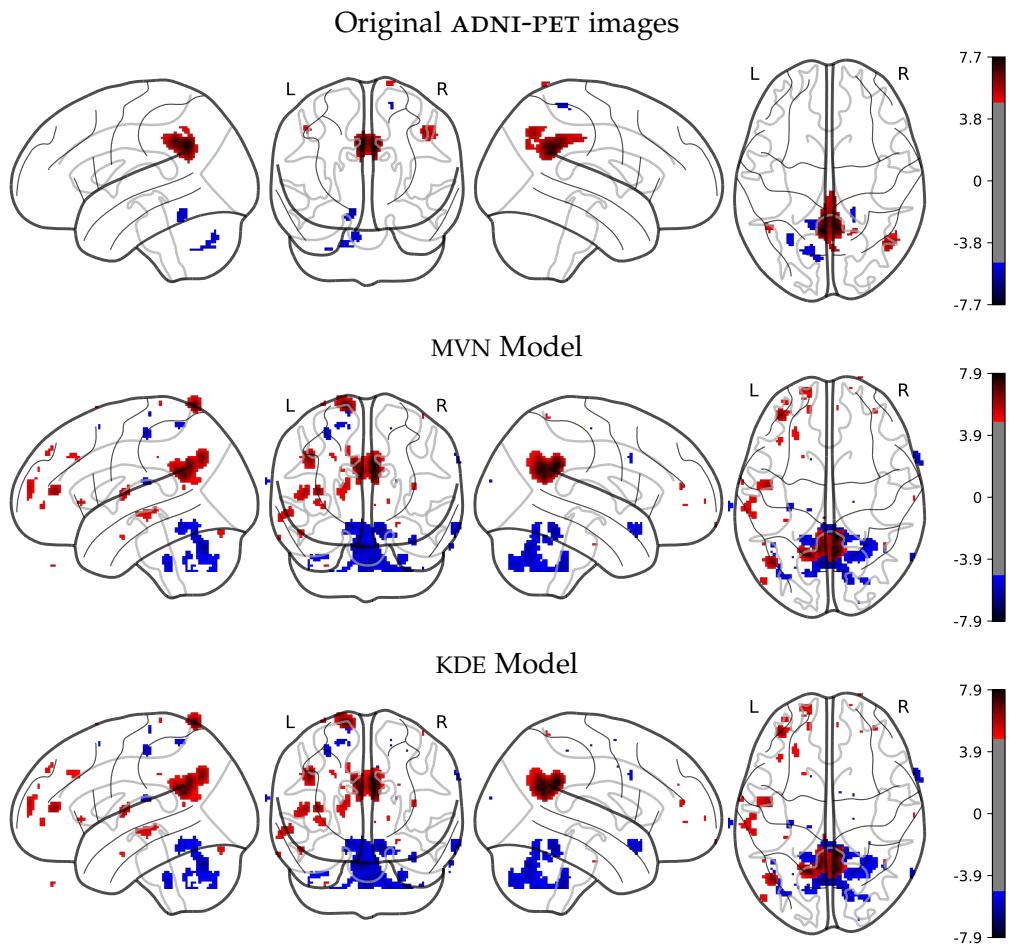


Figure 8.8: SPM Analysis of the ADNI dataset (CTL vs MCI), FWE corrected, with $p = 0.05$, for the original and the simulated images.

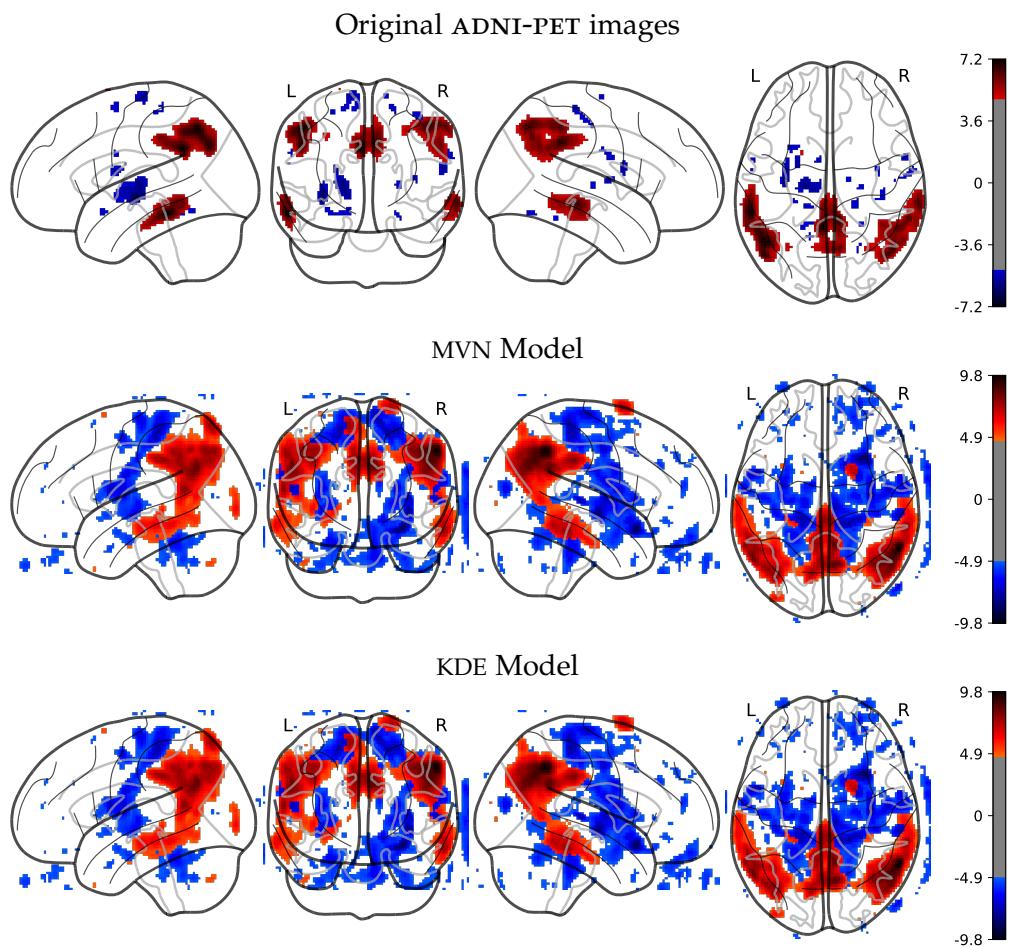


Figure 8.9: SPM Analysis of the ADNI dataset (MCI vs AD), FWE corrected, with $p = 0.05$, for the original and the simulated images.

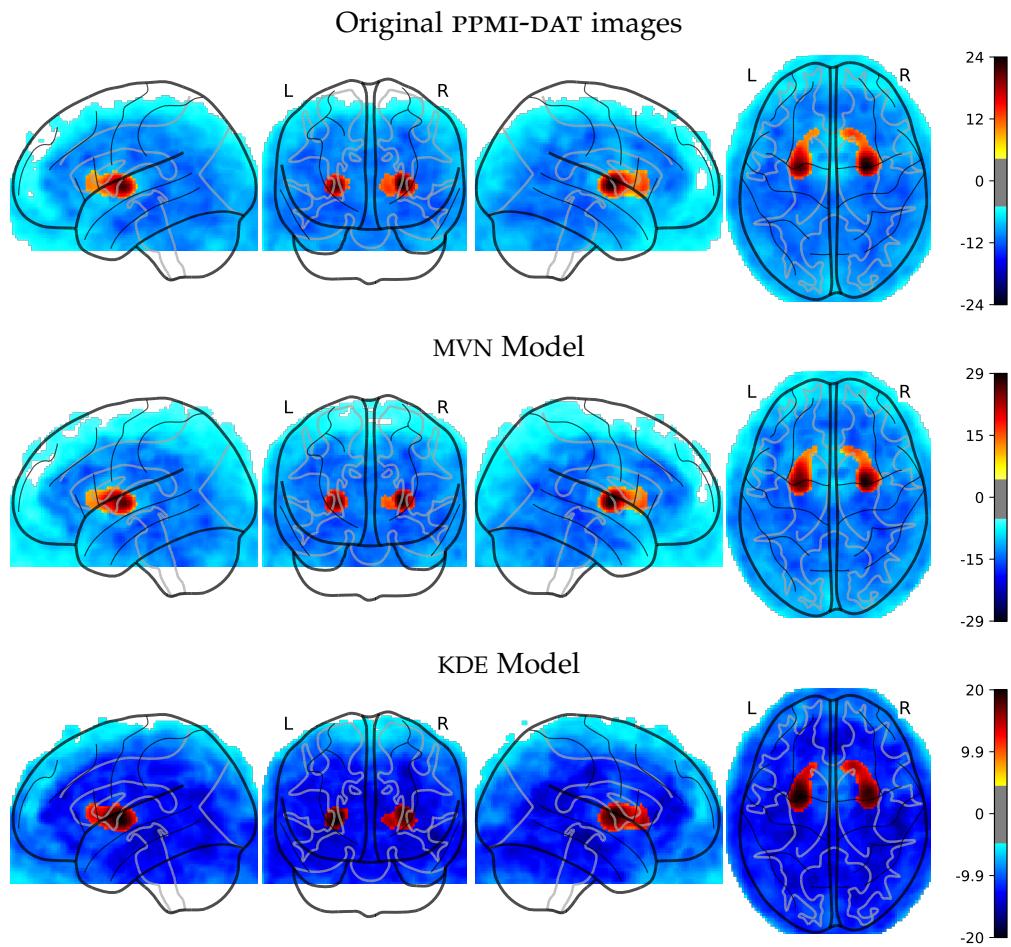


Figure 8.10: SPM Analysis of the PPMI dataset (PD vs CTL), FWE corrected, with $p = 0.05$, for the original and the simulated images.

Part IV

GENERAL DISCUSSION AND CONCLUSIONS

9

GENERAL DISCUSSION AND CONCLUSIONS

9.1 General Discussion

The different contributions that make up this thesis have already been discussed in detail at each chapter. However, in this last chapter, we will discuss the contributions that this work makes to neuroimaging and the CAD field itself, and what discoveries or confirmations we have done on the different diseases that have been analysed.

9.1.1 Discussion on the algorithms

As commented at the introduction, in this thesis we proposed different strategies for tackling the small sample size problem. The first three approaches (chapters 4, 5 and 6) are feature extraction algorithms that perform a significant reduction of the number of features used in neuroimaging.

In the decomposition approach (chapter 4), we obtained a very significant feature reduction, from hundreds of thousands of voxels to between 2 and 25 features, which were the coordinates of each sample in the space defined by the components. Both FA and ICA were able to detect similar regions in the AD and PKS functional datasets, obtaining an accuracy higher than 90% in the first and higher than 95% in the latter.

The decomposition also makes our CAD system better generalizable, since the features are no longer subject to the small sample size problem (the number of subjects is several times higher than the number of features). The samples are projected to a dense space, where the SVC are able to perform a reliable classification, and it also implies a more accurate model of the diseases studied.

Thanks to the smoothness of nuclear imaging techniques such as PET and SPECT, the decomposition techniques are a very useful tool to characterize and predict the stage of a disease. However, due to their resolution, they could not perform as well in structural MRI datasets unless they have been previously smoothed.

As for the texture-based CAD proposed in chapter 5, we have already seen its potential when applied to DaTSCAN images. In these images, the cluster

tendency and homogeneity were the texture features that achieved best performance, which is closely related to the characteristics of the images. Since the distribution of intensities is highly concentrated around the striatum, differences in shape (cluster tendency) and in the distribution of the radiopharmaceutical (homogeneity) are usually regarded by physicians when working with this modality. A combination of these and other texture features yielded the best outcome, achieving up to 97% accuracy in the PPMI-DAT dataset (excluding SWEDD subjects), when using the relative entropy (or KL divergence) as selection criterion.

Despite their ability in detecting PD related DAT deficit in DaTSCAN imaging, these texture features are an inviting possibility to explore in conjunction with other modalities such as MRI. In structural images, textural information in high resolution is available, which could be exploited to characterize textural changes in longitudinal datasets, and associate these with the progression of neurodegeneration.

Finally, we have proposed SBM in chapter 6, a novel technique that maps structural (and possibly functional) images to two-dimensional maps representing various measures. The SBM establishes a framework that has been expanded with divisions on the mapping vector, extensions to the type of sampling and even a path tracing algorithm based on HMM. Of these, the most powerful approaches were the original SBM measures (specifically, the average of the the intensities selected by the mapping vector) and the VRLBP approach, that characterized the texture of and around the mapping vector by means of an helical sampling. The HMM paths, for their part, were not as powerful as a feature selection algorithm, although their ability to adapt to the intensity changes on the images make them a perfect candidate for testing new possibilities such as morphological measures, or even image segmentation.

The whole SBM framework is still to be developed, but it shows very promising results in the classification of structural images. In some preliminary results that we are currently testing, the application of SBM to predict the conversion of MCI to dementia versus MCI stable has reached a remarkable 77.6% accuracy. The MCI conversion is today a major challenge, and results like these open a whole range of possibilities.

On the other hand, chapters 7 and 8 aimed at increasing the number of samples available, and solve problems that usually appear when working with large datasets. The SWPCA (chapter 7), developed by the PhD candidate at the University of Cambridge, was intended to correct many inhomogeneities that were identified in multicentre datasets like the AIMS-MRI. These inhomogeneities caused a SVC to be able to distinguish between centres with higher accuracy than between subjects affected by ASD. Our intention was to cor-

rect this behaviour by decomposing the images using PCA and parsing out the components whose contribution was more related to the acquisition site. After applying SWPCA to the dataset we found that many recent claims about ASD heterogeneity [21] were well founded, given that the differences between individuals affected and not-affected by the disorder were almost null. This makes it a very useful tool for pooling structural datasets acquired at different centres, with the reported limitation that it needs a large sample size at each centre to work out these site differences.

Finally, the functional brain synthesis algorithm proposed in chapter 8 offers an inviting possibility to every neuroimaging scientist: generate hundreds of new images that share characteristics with a certain datasets. It only needs a large enough dataset (of hundreds of images) that has been previously normalized in both space and intensity. Our algorithm transform the original dataset to the ‘eigenbrain’ space, where the statistical distribution of each class is estimated, and then, new coordinates of these can be generated. The synthesized images closely resemble the original ones, and physicians were unable to realize which one was real and which one synthesized in a preliminary test. The experiments proposed in that chapter prove that the synthesized images can effectively predict real world examples and, at the same time, be independent from the original ones. A SPM analysis revealed that the significant class differences are in the same location in both original and simulated images.

We have proved the ability of all these algorithms in the differential diagnosis of two diseases, AD and PD, and the utility of the SWPCA to reduce the site-related inhomogeneities in multicentre datasets. All these algorithms perform either a reduction of the feature space or a safe increase of the sample size in order to reduce the amount of false positives currently found in neuroimaging studies. Other improvements, such as the computational load reduction thanks to feature extraction or the SBM visualization tools may also be acknowledged.

9.1.2 Discussion on the Disorders

All the aforementioned CAD tools have been applied to different neuroimaging datasets that comprise AD (chapters 4, 6 and 8), PKS (chapters 4, 5 and 8) and ASD (chapter 7). Several structural (chapters 6 and 7) and functional (chapter 4, 5 and 8) modalities have been used to analyse these neurological and psychiatric disorders as well.

Regarding AD, we found consistent hypometabolism and hypoperfusion patterns, using the ADNI-PET and VDNL-HMPAO datasets. In chapter 4 we

reported general patterns that could be found in both modalities, mainly at the occipital and temporal lobes, with strong focus on the angular gyrus [96, 160]. These patterns were afterwards confirmed at chapter 8.

Apart from these common regions, the most relevant differences in the ADNI-PET dataset were located at the angular gyrus and the cingulum, although fundamental structures linked to AD such as the hippocampus or the parahippocampal gyrus were also highlighted by some of the selection criteria [56, 120]. However, when using the VDLN-HMPAO, the selected regions were more diffuse (probably due to less resolution images), and mainly located at the angular gyrus, the occipital lobe and parts of the temporal lobe [96, 160]. When using these selection criteria (and consequently, the intensity of these regions), our system achieved the best performance, which gives us an idea of the most relevant differences between AD and CTL in functional datasets.

AD was also analysed at chapter 6, by means of the SBM of MRI datasets. We already mentioned the best performing measures: the average and the VRLBP, which may be linked to anatomical properties such as tissue density and texture. When overimposing the reference image to the SBM t-maps over the average measure, fundamental differences were found at the middle temporal lobe, amygdala, hippocampus, parahippocampal gyrus or some structures of the basal ganglia, such as caudate nucleus, globus pallidus or putamen [37, 96]. The best performing VRLBP approach, however, was more precise, and focused mainly on small areas at the temporal lobe, and the space between the amygdala and hippocampus. It was in these parts where most of the texture changes were detected, and its higher performance gives a hint about the relevance of this report.

A more complex study of the MCI progression is currently being performed using SBM analysis of a longitudinal ADNI dataset, and the preliminary results are very promising. However, there is still much to be done, and we cannot still report significant differences that allow us to predict MCI conversion at the moment.

Now, we will focus on the analysis of PKS. In this thesis, we have always used SPECT-DaTSCAN imaging, and the reported differences only correspond to this. Since DaTSCAN is a highly specific drug that binds to the DATs at the striatum, we can easily observe dopaminergic deficit due to neurodegeneration in PD, and distinguish this from other extrapyramidal symptoms.

In chapter 4 we performed the analysis of the three DaTSCAN datasets: PPMI-DAT, VDLN-DAT and VDLV-DAT. Although they yielded differing performance (mainly due to the inclusion or not of SWEDD subjects in the analysis of chapter 4 and but never at chapter 5), the affected regions were obviously located at the main ROI. The voxels selected by different criteria were located

in areas covering the whole caudate, putamen and globus pallidus, but also, in the case of t-test or wilcoxon, some external structures. The relative entropy, however, focused only on the striatum, and achieved the best performance, from which we can conclude that all the other areas introduced mostly noise.

The texture analysis of chapter 5 was more focused on the differences between dopaminergic deficit caused by PD, and therefore, we excluded SWEDD subjects from the analysis. The texture analysis, especially using texture measures such as cluster tendency and homogeneity, achieved remarkably high accuracy in all three PKS datasets, which correspond to abnormalities in the shape and intensity of the striatum, along with bilateral differences. These patterns have been widely reported in the literature [25, 45, 60] and were confirmed as well in the SPM analysis of both the original and synthesized datasets at chapter 8.

Finally, the analysis of the AIMS-MRI dataset gave us first an idea of the problems to which multi-centre datasets are subject to. In these datasets, the differences between acquisition sites were far more significant than the differences between ASD affected subjects and CTL. That helped us to devise the SWPCA method, which, once it had reduced the influence of the acquisition site on the images, yielded a surprising result: that either there was no difference between classes, or these differences were so heterogeneous that the analysis could not establish any common patterns. Patterns reported in other works could be probably linked to the dataset used and these acquisition site effects, as it had been anticipated by a groundbreaking paper in 2014 [21]. We proposed that defining meaningful subgroups based on different measures, such as genetic profiling, clinical co-morbidities or sensory sensitivities, is the most urgent next step for ASD research, in contrast to the recent pooling of ASD and other disorders such as Asperger's Syndrome in the same category of the DSM-V manual [30].

9.2 Conclusions

Finally, as a summary of this thesis, we provide the following conclusions of the contributions of this thesis:

- We have proposed two major approaches to overcome the small sample size problem in neuroimaging: feature extraction and increasing of sample size. The two approaches proposed are indeed complementary, and can be used together to construct new, more reliable neuroimaging analyses, and by extension, CAD systems.
- The feature extraction algorithms proposed in chapters 4, 5 and 6 have reached high performance in the differential diagnosis of diseased subjects and controls, particularly when applied to the AD and PKS databases. The major challenge here is to use these algorithms in the study of the progression of the different neurodegenerative diseases, particularly the conversion of MCI affected subjects to AD.
- The image decomposition techniques such as the used in chapter 4 offer a significant feature and computational load reduction when analysing high-dimensional data in neuroimaging. These provide fast and more reliable CAD systems that have proven their ability in the analysis of functional imaging modalities such as SPECT and PET.
- Texture analysis is a very powerful tool to analyse DaTSCAN imaging, as it has been proven in chapter 5. Features such as cluster tendency or homogeneity accurately model the variability of shapes and intensities found at the striatum of these modalities, which can be easily linked to PD.
- The SBM framework proposed at chapter 6 reveals itself as a very powerful and extendable tool for the analysis of structural images. Statistical and texture maps derived from this analysis are a further visual aid for detecting changes, at the same time that they provide a significant feature reduction. They have proven their notable ability to diagnose AD from structural patterns, mainly due to texture and tissue density changes that are comprised in the different SBM maps.
- The heterogeneity found in multi-centre structural databases could be substantially reduced using the SWPCA algorithm (chapter 7). This issue was detected in the AIMS-MRI ASD dataset, where a SVC achieved higher performance when classifying sites than classes. After applying SWPCA,

the differences between sites were almost null, and the performance in the differential diagnosis remained similar.

- The analysis of significant differences between ASD and CTL subjects of chapter 7 yielded no detectable differences, or that there was indeed no difference between classes. This might point to the acquisition site differences as one possible cause of variable results when analyzing ASD, as it was reported in [21].
- Finally, the neuroimaging synthesis algorithm proposed in chapter 8 provides a fast option for safely increasing the number of samples in a database, while keeping the individual characteristics. We proved that the simulated images belong to the same statistical distribution of the original and can effectively predict real world samples, at the same time that they could be used to train future professionals.

All these algorithms and frameworks are indeed complementary. One can use the synthesis algorithm to double the size of an existing dataset, and then use decomposition to reduce the feature space. Or purge a multi-centre database with SWPCA and then analyse and visualize the differences using SBM. All these are just drops that contribute to the huge sea of the neuroimaging processing field, in a humble approach to improve the speed, information and reliability of current diagnoses.

Part V
APPENDIX

A | DATASETS

Seven datasets have been used in this thesis, covering three imaging modalities and three different disorders. A summary of these can be found on Table A.1, followed by a longer description of each one.

Acronym	Entity	Disease	Modality	Drug
ADNI-MRI	ADNI	AD	MRI	-
AIMS-MRI	MRC-AIMS	ASD	MRI	-
ADNI-PET	ADNI	AD	PET	FDG
VDLN-HMPAO	VDLN	AD	SPECT	HMPAO
VDLN-DAT	VDLN	PKS	SPECT	DaTSCAN
VDLV-DAT	VDLV	PKS	SPECT	DaTSCAN
PPMI-DAT	PPMI	PKS	SPECT	DaTSCAN

Table A.1: Summary of the datasets used in this thesis.

A.1 MRI

A.1.1 ADNI-MRI and the Alzheimer's Disease Neuroimaging Initiative

The Alzheimer's Disease Neuroimaging Initiative (ADNI) was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. Its primary goal has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and up to 1500 adults (ages 55 to 90) were recruited from over 50 sites across the U.S. and Canada in ADNI and its following initiatives ADNI-GO and ADNI-2. Subjects had completed

at least 6 years of education, and were fluent in Spanish or English. For up-to-date information on inclusion/exclusion criteria and other topics, see www.adni-info.org.

In this thesis we will use data belonging to the ADNI-1 initiative. In particular, the database that we call ADNI-MRI correspond to the MRI volumes from the 'ADNI1: Screening 1.5T' collection (subjects who have a screening data). It contains 818 T1-weighted MRI images from CTL subjects (229), MCI (398) and AD (191) (see demographic details at Table A.2). To avoid prevalence in some of our experiments, we have created two subsets by randomly selecting 180 subjects from the AD and CTL classes.

Group	Sex	N	Age ($\mu \pm \sigma$ years)	MMSE ($\mu \pm \sigma$)
AD	F	91	74.75 ± 7.63	22.98 ± 2.65
	M	100	75.72 ± 2.35	23.15 ± 2.35
CTL	F	110	76.04 ± 0.92	29.21 ± 0.92
	M	119	75.70 ± 1.03	28.95 ± 1.03
MCI	F	141	73.66 ± 2.32	26.73 ± 2.32
	M	257	75.31 ± 2.11	26.97 ± 2.11

Table A.2: Demographics of the ADNI-MRI dataset.

Depending on the experiment, we may use spatially normalized (or registered) T1-weighted MRI images, using the SPM8 software (see Section 3.1) [98], or segmented GM and WM maps. Segmentation was performed using the VBM8 toolbox for SPM [117].

A.1.2 AIMS-MRI, MRC-AIMS Consortium

The MRC-AIMS consortium created this database to study ASD. A number of adult, right-handed males with no significant mean differences in age and full-scale IQ were recruited by advertisement. Participants were excluded from the study if they had a history of major psychiatric disorder or medical illness affecting brain function (e.g. psychosis or epilepsy), or current drug misuse (including alcohol), or were taking antipsychotic medication, mood stabilizers or benzodiazepines.

All participants with ASD were diagnosed according to International Classification of Diseases, 10th Revision (ICD-10) research criteria, and confirmed using the Autism Diagnostic Interview-Revised (ADI-R) [162]. Autism Diagnostic Observation Schedule (ADOS) [141] was performed, but the score was

not considered as an inclusion criteria. ASD participants, to be included, must have scored above the ADI-R cut-off in the three domains of impaired reciprocal social interaction, communication and repetitive behaviours and stereotyped patterns, although failure to reach cut-off in one of the domains by one point was permitted. Intellectual ability was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI) [148], ensuring the participants fell within the high-functioning range on the spectrum defined by a full-scale IQ > 70 .

In this work, only structural MRI from participants recruited at the Institute of Psychiatry, King's College London (LON) and the Autism Research Centre, University of Cambridge (CAM) were included, where an equivalent set of images were acquired from each participant. This makes for a total number of 136 adult, right-handed males (68 with ASD and 68 matched controls). The demographics of the participants are shown in detail in Table A.3.

Database	Group	N	Age ($\mu \pm \sigma$ years)	IQ ($\mu \pm \sigma$)
LON	ASD	39	28.74 ± 6.52	111.28 ± 13.13
	CTL	40	25.30 ± 6.62	104.67 ± 11.16
CAM	ASD	29	26.83 ± 4.64	115.83 ± 11.88
	CTL	28	26.75 ± 7.32	115.25 ± 13.67
ALL	ASD	68	25.90 ± 6.95	109.03 ± 13.31
	CTL	68	27.93 ± 5.87	113.22 ± 12.81

Table A.3: Demographics of the AIMS-MRI dataset.

Structural MRI were obtained using Driven Equilibrium Single Pulse Observation of T₁ and T₂ (DESPOT₁, DESPOT₂) [83] at King's College London and University of Cambridge, both with 3T GE Medical Systems HDx scanners. Using multiple Spoilt Gradient Recall (SPGR) acquisitions in the DESPOT₁ sequence and Steady State Free Procession (SSPF) acquisitions in the DESPOT₂ sequence, with different flip angles and repetition times, qT₁ and qT₂ maps were calculated with a custom ImageJ plug-in package. Correction of main and transmit magnetic field (B₀ and B₁) inhomogeneity effects was performed during the estimation of T₁ and T₂.

For accurate registration to the standard stereotatic space of the MNI, a synT₁ images were created based on the qT₁ maps [32, 44, 46]. The synT₁ images were then segmented using New Segment into GM and WM maps, and normalized to the MNI space using DARTEL in SPM8 [98], with modulation (preserve volume) to retain information of regional/local GM and WM volumes, and

smoothed with a 3mm FWHM Gaussian Kernel to account for inter-subject mis-registration. The synT_1 , qT_1 and qT_2 maps were also registered to the standard MNI space using the same DARTEL flow fields, but without modulation (preserve concentration) to retain information of regional/local T_1 contrast, T_1 relaxation time, and T_2 relaxation times, and smoothed with a 3mm FWHM Gaussian kernel. Therefore, there were five different modalities: qT_1 , qT_2 , synT_1 map, GM and WM maps, for each every participant, which allows us to observe the impact of our SWPCA correction of site-related undesired variance on quantitative (qT_1 and qT_2), simulated (synT_1) images and probability maps (GM and WM).

During the pre-processing of the images, several procedures targeted the reduction of inter-subject and inter-site geometric distortion, amongst them the correction of B_0 and B_1 field inhomogeneity effects and the registration to MNI space. Many other algorithms have been proposed to help in this task. However, the study of their relative performance lies beyond the scope of this article. Following image registration, it was assumed that only the intensity of the maps was affected between sites.

A.2 PET

A.2.1 ADNI-PET, Alzheimer's Disease Neuroimaging Initiative

The ADNI-MRI database also refers from images acquired at the ADNI. However, in this case we will use FDG PET images, used to estimate the metabolic activity of the brain. This radiopharmaceutical is a glucose analog, and its distribution of the brain can be used to trace glucose metabolism, and by extension, brain function. The acquisition procedure is detailed at their website (http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml). In brief, PET images were acquired at a variety of scanners nationwide using either a 30-min six frame scan acquired 30–60 min post-injection or a static 30-min single-frame scan acquired 30–60 min post-injection. The resulting images were aligned along the AC-PC line and a subject-specific intensity normalization was applied, so that the sum of all voxels in the cerebral mask of each subject summed to one. Finally, images were smoothed using the smallest resolution across all scanners, with a uniform isotropic filter of 8mm FWHM.

We used 403 images from the ADNI1 screening database: 95 PET images from AD affected subjects, 207 images from MCI affected subjects and 101 images from CTL. Demographic details of this population can be found at Table A.4.

Group	N	Age ($\mu \pm \sigma$ years)	MMSE ($\mu \pm \sigma$)
AD	95	77.2 ± 7.5	23.4 ± 2.1
MCI	207	76.6 ± 7.2	27.2 ± 1.7
CTL	101	77.8 ± 4.6	29.0 ± 1.1

Table A.4: Demographics of the ADNI-MRI dataset.

Further preprocessing of these images included spatial normalization, via the SPM8 software and its PET template, and intensity normalization. The type of intensity normalization will be specified in each chapter, because it depends on the application.

A.3 SPECT

A.3.1 VDLN-HMPAO, Virgen de las Nieves

The database is built up of imaging studies of subjects following the protocol of a hospital-based service. First, the neurologist evaluated the cognitive function, and those patients with findings of memory loss or dementia were referred to the nuclear medicine department in the Virgen de las Nieves Hospital (VDLN) (Granada, Spain), in order to acquire complementary screening information for diagnosis¹.

The images were visually evaluated by experienced physicians, using 4 different labels: CTL for subjects without scintigraphic abnormalities and mild perfusion deficit (AD-1), moderate deficit (AD-2) and severe deficit (AD-3), to differentiate between levels of hypo-perfusion patterns compatible with AD.

In total, the database consists of 97 subjects: 41 CTL, 30 AD-1, 22 AD-2 and 4 AD-3 (see table A.5 for demographic details). Since the patients are not pathologically confirmed, the subject's labels has some degree of uncertainty, as the pattern of hypo-perfusion may not reflect the underlying pathology of AD, nor the different classification of scans necessarily reflect the severity of the patients symptoms. However, when pathological information is available, visual assessments by experts have been shown to be very sensitive and specific labelling methods, in contrast to neuropsychological tests [118, 149]. Given that this is an inherent limitation of 'in vivo' studies, our working-assumption

¹ Clinical information is unfortunately not available for privacy reasons, but only demographic information

	N	Sex(M/F)(%)	Age ($\mu \pm \sigma$ years)
CTL	41	32.95/12.19	71.51 ± 7.99
AD-1	29	10.97/18.29	65.29 ± 13.36
AD-2	22	13.41/9.76	65.73 ± 8.25
AD-3	4	0/2.43	76 ± 9.90

Table A.5: Demographic details of the VDLN-HMPAO dataset. CTL = Normal Controls, AD-1 = possible AD, AD-2 = probable AD, AD-3 = certain AD. μ and σ stands for population mean and standard deviation respectively.

is that the labels are true, considering the subject label positive when belonging to any of the AD classes, and negative otherwise.

Images were subsequently registered to a custom SPECT template, and a posterior intensity normalization could have been applied, which will be reported in each particular case.

A.3.2 VDLN-DAT, Virgen de las Nieves

This database was supplied by the Virgen de las Nieves Hospital (VDLN) (Granada). It contains patients that were derived to the nuclear medicine service between 2007 and 2012, after a PKS clinical diagnosis, to perform a confirmatory DaTSCAN analysis of the nigrostriatal system. SPECT images were acquired after the injection of 185 MBq of DATSCAN on previously thyroid-blocked patients, and acquired using a three head Picker Prism 3000 gamma camera.

Labels were established by both visual interpretation and exploration quantification, using a delimitation of ROI in the striatum. The visual interpretation was subjective by the physicians, mainly by analyzing the total intensity in both striata, possible asymmetries between them and the existent relationship between the striatal and the background brain activity. For its part, the exploration quantification was performed by computing the radiopharmaceutical activity in counts per pixel for each area, from which measures such as total and differential activity between the striatum and the occipital lobe were obtained.

This database contains 148 subjects: 45 CTL, 73PD and 30 SWEDD, from which only CTL and PD were used. For more details on the demographics of this dataset, see Table A.6.

Group	Sex	N	Age ($\mu \pm \sigma$ years)
CTL	F	20	73.20 ± 9.24
	M	25	69.24 ± 10.80
SWEDD	F	14	73.16 ± 4.83
	M	16	71.50 ± 8.13
PD	F	28	69.57 ± 8.81
	M	45	69.34 ± 9.38

Table A.6: Demographics of the VDLN-DAT dataset.

After the image acquisition, images were preprocessed by means of a spatial normalization, using a template developed by the SiPBA Research Group [15]. A posterior intensity normalization could have been performed, depending on the experiment, and it has specifically been stated in each chapter.

A.3.3 VDLV-DAT, Virgen de la Victoria Hospital

This database was supplied by the Virgen de la Victoria Hospital (VDLV) (Málaga). The images were obtained after a period of between 3 and 4 hours after the intravenous injection of 185 MBq (5 mCi) of DaTSCAN, with prior thyroid blocking with Lugol's solution. The tomographic study (SPECT) with Ioflupane/FP-CIT-I-123 was performed using a General Electric gamma camera, Millennium model, equipped with a dual head and general purpose collimator. A 360-degree circular orbit was made around the cranium, at 3-degree intervals, 60 images with a duration of 35 seconds per interval, 128 × 128 matrix. Image reconstruction was carried out using filtered back-projection algorithms without attenuation correction [167, 171], application of a Hanning filter (frequency 0.7) and images were obtained with transaxial cuts, following the method proposed in [76].

The images were interpreted by three Nuclear Medicine specialists, with masking of the clinical orientation. Visual assessment was established by exclusively considering the normal/abnormal criterion and after arriving at a consensus report between the three specialists, i. e. whether the FP-CIT SPECT allowed differentiation of a group of conditions with presynaptic involvement from others in which their integrity is assumed, without trying to assign them to different clinical groups within the set of pathological studies. A study was considered to be normal when bilateral, symmetrical uptake appeared in cau-

date and putamen nuclei, and abnormal when there were areas of qualitatively reduced uptake in any of the striatal structures.

Group	Sex	N	Age ($\mu \pm \sigma$ years)
CTL	F	54	68.51 ± 10.54
	M	54	69.58 ± 10.01
PD	F	47	67.61 ± 10.24
	M	53	69.52 ± 8.78

Table A.7: Demographics of the VDLV-DAT dataset.

A total of 208 subjects (100 PD and 108 CTL), randomly selected from the total studies performed in this center until December 2008 and referred to it because of a movement disorder, were included in the study. Clinical diagnosis, a parameter used as ‘gold Standard’ to establish the existence of PKS, was made using the diagnostic criteria established previously, with an established minimum follow-up period of 18 months. Those patients who were receiving treatment with drugs that had known or suspected effect on the level of the dopaminergic transporters through direct competitive mechanism were excluded. A more detailed description of the database can be found in [100].

Images were registered to our custom DaTSCAN template, and then, an intensity normalization was usually applied (which will be specified at each experiment).

A.3.4 PPMI-DAT, Parkinson’s Progression Markers Initiative

PPMI –a public-private partnership– is funded by The Michael J. Fox Foundation for Parkinson’s Research and funding partners, including Abbott, Biogen Idec, F. Hoffman-La Roche Ltd., GE Healthcare, Genentech and Pfizer Inc. It is a landmark study launched in 2010 aimed at finding biomarkers for PD diagnosis and treatments.

The images in this dataset were acquired after the injection of between 111 and 185 MBq of DaTSCAN, in subjects that had been pretreated with saturated iodine solution. To facilitate image processing and preserve lateality, a ^{57}Co line marker was affixed along the canthomeatal line, which does not affect the ROIs [58, 67]. For up-to-date information on the study, visit www.ppmi-info.org.

The PPMI-DAT refers to a subset of the DaTSCAN screening data of the PPMI initiative, containing 301 subjects: 111 CTL, 31 SWEDD and 159 PD. These

images were later registered to the custom DaTSCAN template defined previously, and depending on the experiment, an intensity normalization was applied. Demographic details of these subjects can be found at Table A.8.

Group	Sex	N	Age ($\mu \pm \sigma$ years)
CTL	F	45	55.37 ± 10.97
	M	66	59.68 ± 11.48
SWEDD	F	15	56.26 ± 10.25
	M	16	60.18 ± 11.46
PD	F	45	61.20 ± 10.18
	M	114	62.94 ± 8.70

Table A.8: Demographics of the PPMI-DAT dataset.

B

BACKGROUND ON Support Vector Machines (SVMs)

Support Vector Machines (SVMs), introduced in the late 70s [150], are a set of related supervised learning methods widely used in classification (SVCs), regression analysis and, more generally, in any field of machine learning. They have been extensively used in neuroimaging applications such as those presented in this thesis [1, 6, 7, 24, 25, 35] and many others [8, 36, 45, 56, 60, 120].

The simplest SVC assumes the data to be linearly separable. Each sample of a given dataset is considered to be a p-dimensional vector. Therefore, the objective is to separate a set of data, with binary labels, using a hyperplane (multidimensional plane) that maximizes the distance between the two classes. Therefore, we must build a function $f : \Re^p \rightarrow \{\pm 1\}$ that is able to assign a binary value (+1 or -1) to a new sample. This function is built using a set of training data X , which contains N p-dimensional vectors (samples) x_i and their corresponding class y_i :

$$(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \in \Re^p \times \{\pm 1\} \quad (B.1)$$

The linear SVC defines a separation hyperplane in a multidimensional space using the following function:

$$g(x) = \mathbf{w}^T x + \omega_0 = 0, \quad (B.2)$$

where \mathbf{w} is a weight vector, and ω_0 is the threshold. Finding weight vector orthogonal to the maximum separation hyperplane and the unknown thresholds $\omega_i, i = 1, \dots, n$ is known as “training” a SVC. Our training data can be described therefore as:

$$x_i \mathbf{w} + \omega_0 \geqslant +1 \quad \text{for } y_i = +1 \quad (B.3)$$

$$x_i \mathbf{w} + \omega_0 \leqslant -1 \quad \text{for } y_i = -1 \quad (B.4)$$

$$(B.5)$$

which can be combined to a simpler equation:

$$y_i(\mathbf{x}_i \mathbf{w}) - 1 \geqslant 0 \quad \forall i \quad (B.6)$$

Maximizing the margin of the hyperplane is equivalent to minimize $\|\mathbf{w}\|$ subject to Eq. B.6. This can be translated, if we want to use Quadratic Programming, to minimize $\frac{1}{2}\|\mathbf{w}\|^2$.

Suppose now that the data is not linearly separable (as it could be expected from real data). In this case, we introduce the slack variable ξ_i , and Eq. B.6 is converted into:

$$y_i(\mathbf{x}_i \mathbf{w}) - 1 + \xi_i \geq 0 \quad \text{where} \quad \xi_i \geq 0 \quad \forall i \quad (\text{B.7})$$

This is known as soft margin SVM. In this case, we penalize the data points on the incorrect side of the separation hyperplane. We can adapt then the maximization of the margin with:

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_i^L \xi_i \quad \text{s.t.} \quad y_i(\mathbf{x}_i \mathbf{w}) - 1 \geq 0 \quad \forall i \quad (\text{B.8})$$

This procedure is known as regularization, and parameter C controls the trade-off between the slack variable penalty and the size of the margin.

Throughout this thesis we have used the SVM implementation contained in LIBLINEAR [53] and the LinearSVC contained within scikit-learn in python, which is based upon LIBLINEAR itself. We used the L2-regularized L2-loss SVC solver, which summarizes Eq. B.8 in the equation:

$$\text{minimize}_{\mathbf{w}} \quad \frac{\mathbf{w}^T \mathbf{w}}{2} + C \sum \max(0, 1 - y_i \mathbf{w}^T \mathbf{x}_i)^2 \quad (\text{B.9})$$

where we can see that the regularization parameter controls the sum of squared losses, as it corresponds to the L2 solver. The minimization is performed via the Trust Region Newton method (TRON), described in detail at [87].

BIBLIOGRAPHY

- [1] F.J. Martínez-Murcia et al. "On the Brain Structure Heterogeneity of Autism: Parsing out Acquisition Site Effects With Significance-Weighted Principal Component Analysis". In: *Human Brain Mapping* 38.3 (Mar. 2017), pp. 1208–1223.
DOI: [10.1002/hbm.23449](https://doi.org/10.1002/hbm.23449).
- [2] Alzheimer's Association et al. "2016 Alzheimer's disease facts and figures". In: *Alzheimer's & Dementia* 12.4 (2016), pp. 459–509.
- [3] Neil R Carlson. *Physiology of behavior*. Pearson, 2016.
- [4] F.J. Martínez-Murcia, J. Górriz, and J. Ramírez. "Computer Aided Diagnosis in Neuroimaging". In: *Computer-aided Technologies - Applications in Engineering and Medicine*. Ed. by Razvan Udroiu. 1st ed. InTech, Dec. 2016. Chap. 7, pp. 137–160. ISBN: 978-953-51-4895-1.
DOI: [10.5772/64980](https://doi.org/10.5772/64980).
- [5] F.J. Martínez-Murcia, J.M. Górriz, and J. Ramírez. "Feature Extraction". In: *Wiley Encyclopedia of Electrical and Electronics Engineering*. Wiley, 2016, Accepted.
DOI: [10.1002/047134608X](https://doi.org/10.1002/047134608X).
- [6] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, and A. Ortiz. "A Spherical Brain Mapping of MR Images for the Detection of Alzheimer's Disease". In: *Current Alzheimer Research* 13.5 (Mar. 2016), pp. 575–588.
DOI: [10.2174/1567205013666160314145158](https://doi.org/10.2174/1567205013666160314145158).
- [7] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, and A. Ortiz. "A Structural Parametrization of the Brain Using Hidden Markov Models-Based Paths in Alzheimer's Disease". In: *International Journal of Neural Systems* 26.7 (Nov. 2016), p. 1650024.
DOI: [10.1142/S0129065716500246](https://doi.org/10.1142/S0129065716500246).
- [8] F Segovia, JM Gorriz, J Ramírez, and D Salas-Gonzalez. "Multiclass classification of 18 F-DMFP-PET data to assist the diagnosis of parkinsonism". In: *Pattern Recognition in Neuroimaging (PRNI), 2016 International Workshop on*. IEEE. 2016, pp. 1–4.

- [9] Jeff Sevigny, Ping Chiao, Thierry Bussière, Paul H Weinreb, Leslie Williams, Marcel Maier, Robert Dunstan, Stephen Salloway, Tianle Chen, Yan Ling, et al. "The antibody aducanumab reduces A β plaques in Alzheimer's disease". In: *Nature* 537.7618 (2016), pp. 50–56.
- [10] Brett A Clementz, John A Sweeney, Jordan P Hamm, Elena I Ivleva, Lauren E Ethridge, Godfrey D Pearson, Matcheri S Keshavan, and Carol A Tamminga. "Identification of distinct psychosis biotypes using brain-based biomarkers". In: *American Journal of Psychiatry* 173.4 (2015), pp. 373–384.
- [11] Christine Ecker, Susan Y Bookheimer, and Declan GM Murphy. "Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan". In: *The Lancet Neurology* 14.11 (2015), pp. 1121–1134.
- [12] Leanna M Hernandez, Jeffrey D Rudie, Shulamite A Green, Susan Bookheimer, and Mirella Dapretto. "Neural signatures of autism spectrum disorders: insights into brain network dynamics". In: *Neuropsychopharmacology* 40.1 (2015), pp. 171–189.
- [13] L. Khedher, J. Ramírez, J.M. Górriz, A. Brahim, and F. Segovia. "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images". In: *Neurocomputing* 151 (Mar. 2015), pp. 139–150. ISSN: 0925-2312.
DOI: [10.1016/j.neucom.2014.09.072](https://doi.org/10.1016/j.neucom.2014.09.072).
- [14] F.J. Martínez-Murcia, A. Ortiz, J. Manuel Górriz, J. Ramírez, and I.A. Illán. "A volumetric radial LBP projection of MRI brain images for the diagnosis of alzheimer's disease". In: *Artificial Computation in Biology and Medicine*. Vol. 9107. Springer Science + Business Media, 2015, pp. 19–28.
DOI: [10.1007/978-3-319-18914-7_3](https://doi.org/10.1007/978-3-319-18914-7_3).
- [15] Diego Salas-Gonzalez, Juan M Górriz, Javier Ramírez, Ignacio A Illán, Pablo Padilla, Francisco J Martínez-Murcia, and Elmar W Lang. "Building a FP-CIT SPECT Brain Template Using a Posterization Approach". In: *Neuroinformatics* 13.4 (2015), pp. 391–402.
- [16] Minna Sikiö, Kirsi K Holli-Helenius, Lara CV Harrison, Pertti Ryymin, Hanna Ruottinen, Tiia Saunamäki, Hannu J Eskola, Irina Elovaara, and Prasun Dastidar. "MR image texture in Parkinson's disease: a longitudinal study". In: *Acta Radiologica* 56.1 (2015), pp. 97–104.

- [17] Zheng Wang, Shashwath A Meda, Matcheri S Keshavan, Carol A Tamminga, John A Sweeney, Brett A Clementz, David J Schretlen, Vince D Calhoun, Su Lui, and Godfrey D Pearlson. "Large-scale fusion of gray matter and resting-state functional MRI reveals common and distinct biological markers across the psychosis spectrum in the B-SNIP cohort". In: *Frontiers in psychiatry* 6 (2015), p. 174.
- [18] Nicole R Zürcher, Anisha Bhanot, Christopher J McDougle, and Jacob M Hooker. "A systematic review of molecular imaging (PET and SPECT) in autism spectrum disorder: current state and future research opportunities". In: *Neuroscience & Biobehavioral Reviews* 52 (2015), pp. 56–73.
- [19] A Di Martino et al. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism." In: *Molecular psychiatry* 19 (6 June 2014), pp. 659–667. ISSN: 1476-5578. DOI: [10.1038/mp.2013.78](https://doi.org/10.1038/mp.2013.78).
- [20] Christine Ecker and Declan Murphy. "Neuroimaging in autism—from basic science to translational research." In: *Nature reviews. Neurology* 10 (2 Feb. 2014), pp. 82–91. ISSN: 1759-4766. DOI: [10.1038/nrneurol.2013.276](https://doi.org/10.1038/nrneurol.2013.276).
- [21] Shlomi Haar, Sigal Berman, Marlene Behrmann, and Ilan Dinstein. "Anatomical Abnormalities in Autism?" In: *Cerebral Cortex* (2014). DOI: [10.1093/cercor/bhu242](https://doi.org/10.1093/cercor/bhu242).
- [22] Ignacio A. Illán, Juan M. Górriz, Javier Ramírez, and Anke Meyer-Base. "Spatial component analysis of MRI data for Alzheimer's disease diagnosis: a Bayesian network approach." eng. In: *Frontiers in Computational Neuroscience* 8 (2014), p. 156. DOI: [10.3389/fncom.2014.00156](https://doi.org/10.3389/fncom.2014.00156).
- [23] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I. Álvarez Illán, D. Salas-González, and F. Segovia. "Projecting MRI brain images for the detection of Alzheimer's Disease". In: *Studies in Health Technology and Informatics*. Vol. 207. 2014, pp. 225–233. DOI: [10.3233/978-1-61499-474-9-225](https://doi.org/10.3233/978-1-61499-474-9-225).
- [24] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, and A. Ortiz. "Automatic detection of Parkinsonism using significance measures and component analysis in DaTSCAN imaging". In: *Neurocomputing* 126 (Feb. 2014), pp. 58–70. DOI: [10.1016/j.neucom.2013.01.054](https://doi.org/10.1016/j.neucom.2013.01.054).

- [25] F.J. Martinez-Murcia, J.M. Górriz, J. Ramírez, M. Moreno-Caballero, and M. Gómez-Río. "Parametrization of textural patterns in ^{123}I -ioflupane imaging for the automatic detection of Parkinsonism". In: *Medical Physics* 41.1 (2014), p. 012502.
DOI: [10.1118/1.4845115](https://doi.org/10.1118/1.4845115).
- [26] Fermín Segovia, Rosemary Holt, Michael Spencer, Juan M Górriz, Javier Ramírez, Carlos G Puntonet, Christophe Phillips, Lindsay Chura, Simon Baron-Cohen, and John Suckling. "Identifying endophenotypes of autism: a multivariate approach". In: *Frontiers in computational neuroscience* 8 (2014), p. 60.
- [27] John Suckling, Julian Henty, Christine Ecker, Sean C Deoni, Michael V Lombardo, Simon Baron-Cohen, Peter Jezzard, Anna Barnes, Bhismadev Chakrabarti, Cinly Ooi, et al. "Are power calculations useful? A multicentre neuroimaging study". In: *Human brain mapping* 35.8 (2014), pp. 3569–3577.
- [28] Marjolein Verly, Judith Verhoeven, Inge Zink, Dante Mantini, Ronald Peeters, Sabine Deprez, Louise Emsell, Bart Boets, Ilse Noens, Jean Steyaert, et al. "Altered functional connectivity of the language network in ASD: role of classical language areas and cerebellum". In: *NeuroImage: Clinical* 4 (2014), pp. 374–382.
- [29] Anderson M Winkler, Gerard R Ridgway, Matthew A Webster, Stephen M Smith, and Thomas E Nichols. "Permutation inference for the general linear model". In: *Neuroimage* 92 (2014), pp. 381–397.
- [30] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.
- [31] Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. "Power failure: why small sample size undermines the reliability of neuroscience". In: *Nature Reviews Neuroscience* 14.5 (Apr. 2013), pp. 365–376.
DOI: [10.1038/nrn3475](https://doi.org/10.1038/nrn3475).
- [32] Christine Ecker et al. "Brain surface anatomy in adults with autism: the relationship between surface area, cortical thickness, and autistic symptoms." eng. In: *JAMA Psychiatry* 70.1 (Jan. 2013), pp. 59–70.
DOI: [10.1001/jamapsychiatry.2013.265](https://doi.org/10.1001/jamapsychiatry.2013.265).
- [33] Rhoshel K Lenroot and Pui KaYeung. "Heterogeneity within autism spectrum disorders: what have we learned from neuroimaging studies?" In: (2013).

- [34] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, I.A. Illán, and C.G. Puntonet. "Texture Features Based Detection of Parkinson's Disease on DaTSCAN Images". In: *Natural and Artificial Computation in Engineering and Medical Applications*. Vol. 7931 LNCS. PART 2. Springer Science + Business Media, 2013, pp. 266–277.
DOI: [10.1007/978-3-642-38622-0_28](https://doi.org/10.1007/978-3-642-38622-0_28).
- [35] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, C.G. Puntonet, and I.A. Illán. "Functional activity maps based on significance measures and Independent Component Analysis". In: *Computer Methods and Programs in Biomedicine* 111.1 (July 2013), pp. 255–268.
DOI: [10.1016/j.cmpb.2013.03.015](https://doi.org/10.1016/j.cmpb.2013.03.015).
- [36] Andrés Ortiz, Juan M Górriz, Javier Ramírez, and Francisco Jesús Martínez-Murcia. "LVQ-SVM Based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease". In: *Pattern Recognition Letters* 34.14 (Oct. 2013), pp. 1725–1733. ISSN: 0167-8655.
DOI: [10.1016/j.patrec.2013.04.014](https://doi.org/10.1016/j.patrec.2013.04.014).
- [37] Michela Pievani, Martina Bocchetta, Marina Boccardi, Enrica Cavedo, Matteo Bonetti, Paul M. Thompson, and Giovanni B. Frisoni. "Striatal morphology in early-onset and late-onset Alzheimer's disease: a preliminary study". In: *Neurobiology of Aging* 34.7 (July 2013), pp. 1728–1739. ISSN: 0197-4580.
DOI: [10.1016/j.neurobiolaging.2013.01.016](https://doi.org/10.1016/j.neurobiolaging.2013.01.016).
- [38] A. Rojas, J.M. Górriz, J. Ramírez, I.A. Illán, F.J. Martínez-Murcia, A. Ortiz, M. Gómez Río, and M. Moreno-Caballero. "Application of Empirical Mode Decomposition (EMD) on DaTSCAN SPECT images to explore Parkinson Disease". In: *Expert Systems with Applications* 40.7 (2013), pp. 2756–2766. ISSN: 0957-4174.
DOI: [10.1016/j.eswa.2012.11.017](https://doi.org/10.1016/j.eswa.2012.11.017).
- [39] Diego Salas-Gonzalez, Juan M Górriz, Javier Ramírez, Ignacio A Illán, and Elmar W Lang. "Linear intensity normalization of FP-CIT SPECT brain images using the α -stable distribution." In: *NeuroImage* 65 (Jan. 2013), pp. 449–455. ISSN: 1095-9572.
DOI: [10.1016/j.neuroimage.2012.10.005](https://doi.org/10.1016/j.neuroimage.2012.10.005).
- [40] Fermín Segovia, JM Górriz, Javier Ramírez, Diego Salas-Gonzalez, and Ignacio Álvarez. "Early diagnosis of Alzheimer's disease based on partial least squares and support vector machine". In: *Expert Systems with Applications* 40.2 (2013), pp. 677–683.

- [41] Reza Tavoli, Ehsan Kozegar, Mohammad Shojafar, Hossein Soleimani, and Zahra Pooranian. "Weighted PCA for improving Document Image Retrieval System based on keyword spotting accuracy". In: *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*. IEEE. 2013, pp. 773–777.
- [42] Jessica A Turner, Eswar Damaraju, Theo GM Van Erp, Daniel H Mathalon, Judith M Ford, James Voyvodic, Bryon A Mueller, Aysenil Belger, Juan Bustillo, Sarah Christine McEwen, et al. "A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia". In: *Frontiers in neuroscience* 7 (2013), p. 137.
- [43] Nikolaus Weiskopf, John Suckling, Guy Williams, Marta Morgado Correia, Becky Inkster, Roger Tait, Cinly Ooi, Edward T Bullmore, and Antoine Lutti. "Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation". In: *Frontiers in neuroscience* 7 (2013), p. 95.
- [44] Christine Ecker, John Suckling, Sean C Deoni, Michael V Lombardo, Ed T Bullmore, Simon Baron-Cohen, Marco Catani, Peter Jezzard, Anna Barnes, Anthony J Bailey, et al. "Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study". In: *Archives of general psychiatry* 69.2 (2012), pp. 195–209.
- [45] I.A.a Illán, J.M.a Górriz, J.a Ramírez, F.a Segovia, J.M.b Jiménez-Hoyuela, and S.J.b Ortega Lozano. "Automatic assistance to Parkinsons disease diagnosis in DaTSCAN SPECT imaging". In: *Medical Physics* 39.10 (2012), pp. 5971–5980.
- [46] Meng-Chuan Lai, Michael V Lombardo, Bhismadev Chakrabarti, Christine Ecker, Susan A Sadek, Sally J Wheelwright, Declan GM Murphy, John Suckling, Edward T Bullmore, Simon Baron-Cohen, et al. "Individual differences in brain structure underpin empathizing–systemizing cognitive styles in male adults". In: *Neuroimage* 61.4 (2012), pp. 1347–1354.
- [47] F.J. Martínez-Murcia, J.M. Górriz, J. Ramírez, C.G. Puntonet, and D. Salas-González. "Computer Aided Diagnosis tool for Alzheimer's Disease based on Mann-Whitney-Wilcoxon U-Test". In: *Expert Systems with Applications* 39.10 (Aug. 2012), pp. 9676–9685.
DOI: 10.1016/j.eswa.2012.02.153.

- [48] Martin Reuter, Nicholas J Schmansky, H Diana Rosas, and Bruce Fischl. "Within-subject template estimation for unbiased longitudinal image analysis." In: *NeuroImage* 61 (4 July 2012), pp. 1402–1418. ISSN: 1095-9572.
DOI: [10.1016/j.neuroimage.2012.02.084](https://doi.org/10.1016/j.neuroimage.2012.02.084).
- [49] F. Segovia, J. M. Górriz, J. Ramírez, I. Álvarez, J. M. Jiménez-Hoyuela, and S. J. Ortega. "Improved Parkinsonism diagnosis using a partial least squares based approach". In: *Medical physics* 39.7 (2012), pp. 4395–4403.
- [50] Jeffrey S Simonoff. *Smoothing methods in statistics*. Springer Science & Business Media, 2012.
- [51] John Suckling, Anna Barnes, Dominic Job, David Brennan, Katherine Lymer, Paola Dazzan, Tiago Reis Marques, Clare MacKay, Shane McKie, Steve R Williams, et al. "The neuro/PsyGRID calibration experiment". In: *Human brain mapping* 33.2 (2012), pp. 373–386.
- [52] Clive Ballard, Serge Gauthier, Anne Corbett, Carol Brayne, Dag Aarsland, and Emma Jones. "Alzheimer's disease." In: *Lancet (London, England)* 377 (9770 Mar. 2011), pp. 1019–1031. ISSN: 1474-547X.
DOI: [10.1016/S0140-6736\(10\)61349-9](https://doi.org/10.1016/S0140-6736(10)61349-9).
- [53] Chih-Chung Chang and Chih-Jen Lin. "LIBSVM: a library for support vector machines". In: *ACM Transactions on Intelligent Systems and Technology* 2.3 (Apr. 2011), pp. 1–27. ISSN: 2157-6904.
DOI: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199).
- [54] J. M. Górriz, F. Segovia, J. Ramírez, A. Lassl, and D. Salas-Gonzalez. "GMM based SPECT image classification for the diagnosis of Alzheimer's disease". In: *Applied Soft Computing* 11.2 (2011), pp. 2313–2325. ISSN: 1568-4946.
DOI: [10.1016/j.asoc.2010.08.012](https://doi.org/10.1016/j.asoc.2010.08.012).
- [55] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, and A. Besga. "Computer Aided Diagnosis system for Alzheimer Disease using brain Diffusion Tensor Imaging features selected by Pearson's correlation". In: *Neuroscience Letters* 502.3 (Sept. 2011), pp. 225–229. ISSN: 0304-3940.
DOI: [10.1016/j.neulet.2011.07.049](https://doi.org/10.1016/j.neulet.2011.07.049).
- [56] I. A. Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. M. López, F. Segovia, R. Chaves, M. Gómez-Rio, and C. G. Puntonet. "¹⁸F-FDG PET imaging analysis for computer aided Alzheimer's diagnosis". In:

- Information Sciences* 181 (4 Feb. 2011), pp. 903–916. ISSN: 0020-0255.
DOI: 10.1016/j.ins.2010.10.027.
- [57] Michael V Lombardo, Bhismadev Chakrabarti, Edward T Bullmore, Simon Baron-Cohen, MRC AIMS Consortium, et al. “Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism”. In: *Neuroimage* 56.3 (2011), pp. 1832–1838.
- [58] Kenneth Marek et al. “The Parkinson Progression Marker Initiative (PPMI)”. In: *Progress in Neurobiology* 95.4 (Dec. 2011), pp. 629–635. ISSN: 0301-0082.
DOI: 10.1016/j.pneurobio.2011.09.005.
- [59] F.J. Martínez, D. Salas-González, J.M. Górriz, J. Ramírez, C.G. Puntonet, and M. Gómez-Río. “Analysis of Spect Brain Images Using Wilcoxon and Relative Entropy Criteria and Quadratic Multivariate Classifiers for the Diagnosis of Alzheimer’s Disease”. In: *New Challenges on Bioinspired Applications*. Vol. 6687. Lecture Notes in Computer Science PART 2. Springer Science + Business Media, 2011, pp. 41–48.
DOI: 10.1007/978-3-642-21326-7_5.
- [60] David J Towey, Peter G Bain, and Kuldip S Nijran. “Automatic classification of ^{123}I -FP-CIT (DaTSCAN) SPECT images”. In: *Nuclear Medicine Communications* 32.8 (Aug. 2011). PMID: 21659911, pp. 699–707. ISSN: 1473-5628.
DOI: 10.1097/MNM.0b013e328347cd09.
- [61] Brian B Avants, Paul Yushkevich, John Pluta, David Minkoff, Marc Korczykowski, John Detre, and James C Gee. “The optimal template effect in hippocampus studies of diseased populations”. In: *Neuroimage* 49.3 (2010), pp. 2457–2466.
- [62] Yoav Benjamini. “Simultaneous and selective inference: current successes and future challenges”. In: *Biometrical Journal* 52.6 (2010), pp. 708–721.
- [63] Matias Bossa, Ernesto Zacur, Salvador Olmos, Alzheimer’s Disease Neuroimaging Initiative, et al. “Tensor-based morphometry with stationary velocity field diffeomorphic registration: application to ADNI”. In: *Neuroimage* 51.3 (2010), pp. 956–969.
- [64] ZI Botev, JF Grotowski, DP Kroese, et al. “Kernel density estimation via diffusion”. In: *The Annals of Statistics* 38.5 (2010), pp. 2916–2957.

- [65] Christine Ecker, Vanessa Rocha-Rego, Patrick Johnston, Janaina Mourao-Miranda, Andre Marquand, Eileen M Daly, Michael J Brammer, Clodagh Murphy, Declan G Murphy, MRC AIMS Consortium, et al. "Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach". In: *Neuroimage* 49.1 (2010), pp. 44–56.
- [66] Michael P. Fay and Michael A. Proschan. "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules". In: *Statistics Surveys* 4 (2010), pp. 1–39.
- [67] The Parkinson Progression Markers Initiative. *PPMI. Imaging Technical Operations Manual*. 2nd ed. June 2010.
- [68] Martin Reuter, H Diana Rosas, and Bruce Fischl. "Highly accurate inverse consistent registration: a robust approach." In: *NeuroImage* 53 (4 Dec. 2010), pp. 1181–1196. ISSN: 1095-9572.
DOI: 10.1016/j.neuroimage.2010.07.020.
- [69] V Vinzi, Wynne W Chin, Jörg Henseler, Huiwen Wang, et al. *Handbook of partial least squares*. Springer, 2010.
- [70] J. M. Górriz, A. Lassl, J. Ramírez, D. Salas-Gonzalez, C. G. Puntonet, and E. W. Lang. "Automatic selection of ROIs in functional imaging using Gaussian mixture models". In: *Neuroscience Letters* 460.2 (2009), pp. 108–111.
- [71] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "The Elements of Statistical Learning". In: Springer New York, Aug. 26, 2009. Chap. 7. ISBN: 978-0-387-84858-7.
- [72] Gabor T. Herman. "Filtered Backprojection for Parallel Beams". In: *Fundamentals of Computerized Tomography*. Springer, Jan. 2009. ISBN: 978-1-85233-617-2.
DOI: 10.1007/978-1-84628-723-7_8.
- [73] I. Álvarez Illán, J. M. Górriz, J. Ramírez, D. Salas-Gonzalez, M. López, C. G. Puntonet, and F. Segovia. "Independent Component Analysis of SPECT Images to Assist the Alzheimer's Disease Diagnosis". In: *Advances in Neural Networks Research, ISNN 2009. Advances in Intelligent and Soft Computing*. Wuhan (China). May 2009.
- [74] Chandan Misra, Yong Fan, and Christos Davatzikos. "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI". In: *Neuroimage* 44.4 (Feb. 2009), pp. 1415–1422. ISSN: 1053-8119.
DOI: 10.1016/j.neuroimage.2008.10.031.

- [75] Godfrey Pearlson. *Multisite collaborations and large databases in psychiatric neuroimaging: advantages, problems, and challenges*. 2009.
- [76] J. Ramírez, J. M. Górriz, R. Chaves, M. López, D. Salas-González, I. Álvarez, and F. Segovia. "SPECT image classification using random forests". In: *Electronics Letters* 45 (2009), pp. 604–605.
doi: 10.1049/el.2009.1111.
- [77] Diego Salas-Gonzalez, Juan M Górriz, Javier Ramírez, Miriam López, Ignacio A Illan, Fermín Segovia, Carlos G Puntonet, and Manuel Gómez-Río. "Analysis of SPECT brain images for the diagnosis of Alzheimer's disease using moments and support vector machines." In: *Neuroscience letters* 461 (1 Sept. 2009), pp. 60–64. ISSN: 1872-7972.
doi: 10.1016/j.neulet.2009.05.056.
- [78] Phoebe G. Spetsieris, Yilong Ma, Vijay Dhawan, and David Eidelberg. "Differential diagnosis of parkinsonian syndromes using functional PCA-based imaging features". In: *Neuroimage* 45.4 (May 2009), pp. 1241–52.
ISSN: 1053-8119.
- [79] John Darrell Van Horn and Arthur W Toga. "Multi-site neuroimaging trials". In: *Current opinion in neurology* 22.4 (2009), p. 370.
- [80] Li Wang, Chunming Li, Quansen Sun, Deshen Xia, and Chiu-Yen Kao. "Active contours driven by local and global intensity fitting energy with application to brain MR image segmentation". In: *Computerized Medical Imaging and Graphics* 33.7 (2009), pp. 520–531.
- [81] Lai Xu, Karyn M Groth, Godfrey Pearlson, David J Schretlen, and Vince D Calhoun. "Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia". In: *Human brain mapping* 30.3 (2009), pp. 711–724.
- [82] H.C. Chen, M. Goldberg, M. Magdon-Ismail, and W.A. Wallace. "Reverse Engineering a Social Agent-Based Hidden Markov Model - ViSAGE". In: *International Journal of Neural Systems* 18.06 (2008), pp. 491–526.
- [83] Sean CL Deoni, Steven CR Williams, Peter Jezzard, John Suckling, Declan GM Murphy, and Derek K Jones. "Standardized structural magnetic resonance imaging in multicentre studies using quantitative T₁ and T₂ imaging at 1.5 T". In: *Neuroimage* 40.2 (2008), pp. 662–671.

- [84] Milos D Ikonomovic et al. "Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease." In: *Brain : a journal of neurology* 131 (Pt 6 June 2008), pp. 1630–1645. ISSN: 1460-2156. DOI: 10.1093/brain/awn016.
- [85] L. W. de Jong, K. van der Hiele, I. M. Veer, J. J. Houwing, R. G. J. Westendorp, E. L. E. M. Bollen, P. W. de Bruin, H. A. M. Middelkoop, M. A. van Buchem, and J. van der Grond. "Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study". eng. In: *Brain* 131.12 (Nov. 2008), pp. 3277–3285. ISSN: 1460-2156. DOI: 10.1093/brain/awn278.
- [86] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. "A general framework for increasing the robustness of PCA-based correlation clustering algorithms". In: *International Conference on Scientific and Statistical Database Management*. Springer. 2008, pp. 418–435.
- [87] Chih-Jen Lin, Ruby C Weng, and S Sathiya Keerthi. "Trust region newton method for logistic regression". In: *Journal of Machine Learning Research* 9.Apr (2008), pp. 627–650.
- [88] Edith Lopez-Hurtado and Jorge J Prieto. "A microscopic study of language-related cortex in autism". In: *Am J Biochem Biotechnol* 4 (2008), pp. 130–145.
- [89] David A. Medina and Moises Gaviria. "Diffusion tensor imaging investigations in Alzheimer's disease: the resurgence of white matter compromise in the cortical dysfunction of the aging brain." eng. In: *Neuropsychiatric Disease and Treatment* 4.4 (Aug. 2008), pp. 737–742.
- [90] Carl Philips, Daniel Li, Daniela Raicu, and Jacob Furst. "Directional Invariance of Co-occurrence Matrices within the Liver". In: *International Conference on Biocomputation, Bioinformatics, and Biomedical Technologies* o (2008), pp. 29–34.
DOI: 10.1109/BIOTECHNO.2008.24.
- [91] GN Srinivasan and G Shobha. "Statistical texture analysis". In: *Proceedings of world academy of science, engineering and technology*. Vol. 36. 2008, pp. 1264–1269.
- [92] Cynthia M Stonnington, Geoffrey Tan, Stefan Klöppel, Carlton Chu, Bogdan Draganski, Clifford R Jack, Kewei Chen, John Ashburner, and Richard SJ Frackowiak. "Interpreting scan data acquired from multiple scanners: a study with Alzheimer's disease". In: *Neuroimage* 39.3 (2008), pp. 1180–1185.

- [93] Klaus Tatsch. "Extrapyramidal syndromes: PET and SPECT". In: *Diseases of the Brain, Head & Neck, Spine*. Springer, 2008, pp. 234–239.
- [94] John Ashburner. "A fast diffeomorphic image registration algorithm". In: *Neuroimage* 38.1 (2007), pp. 95–113.
- [95] Federico De Martino, Francesco Gentile, Fabrizio Esposito, Marco Balsi, Francesco Di Salle, Rainer Goebel, and Elia Formisano. "Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers". In: *Neuroimage* 34.1 (Jan. 2007), pp. 177–194. ISSN: 1053-8119.
- [96] Bruno Dubois et al. "Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria". In: *The Lancet Neurology* 6.8 (Aug. 2007), pp. 734–746. ISSN: 1474-4422. doi: 10.1016/s1474-4422(07)70178-3.
- [97] Thomas Eckert and Christine Edwards. "The application of network mapping in differential diagnosis of parkinsonian disorders". In: *Clinical Neuroscience Research* 6.6 (2007). Neural Networks in the Imaging of Neuropsychiatric Diseases, pp. 359–366. ISSN: 1566-2772. doi: 10.1016/j.cnr.2007.05.001.
- [98] K.J. Friston, J. Ashburner, S.J. Kiebel, T.E. Nichols, and W.D. Penny. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press, 2007.
- [99] Jeffrey T Leek and John D Storey. "Capturing heterogeneity in gene expression studies by surrogate variable analysis". In: *PLoS Genet* 3.9 (2007), e161.
- [100] S.J. Ortega Lozano, M.D. Martínez del Valle Torres, J.M. Jiménez-Hoyuela García, A.L. Gutiérrez Cardo, and V. Campos Arillo. "Diagnostic accuracy of FP-CIT SPECT in patients with Parkinsonism". In: *Revista Española de Medicina Nuclear (English Edition)* 26.5 (2007), pp. 277–285. ISSN: 1578-200X. doi: 10.1016/S1578-200X(07)70062-1.
- [101] J Puñal Riobóo, L Varela Lema, A Serena Puig, and A Ruano-Ravina. "Effectiveness of ^{123}I -ioflupane (DaTSCAN) in the diagnosis of Parkinsonian syndromes. A systematic review". In: *Revista Española De Medicina Nuclear* 26.6 (Dec. 2007). PMID: 18021694, pp. 375–384. ISSN: 0212-6982.

- [102] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. "A review of feature selection techniques in bioinformatics." eng. In: *Bioinformatics* 23.19 (Oct. 2007), pp. 2507–2517.
DOI: [10.1093/bioinformatics/btm344](https://doi.org/10.1093/bioinformatics/btm344).
- [103] Guoying Zhao and Matti Pietikainen. "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.6 (June 2007), pp. 915–928. ISSN: 0162-8828.
DOI: [10.1109/TPAMI.2007.1110](https://doi.org/10.1109/TPAMI.2007.1110).
- [104] Yun Zhou, Susan M. Resnick, Weiguo Ye, Hong Fan, Daniel P. Holt, William E. Klunk, Chester A. Mathis, Robert Dannals, and Dean F. Wong. "Using a reference tissue model with spatial constraint to quantify [¹¹C]Pittsburgh compound B PET for early diagnosis of Alzheimer's disease". In: *Neuroimage* 36.2 (June 2007), pp. 298–312. ISSN: 1053-8119.
- [105] Mirella Dapretto, Mari S. Davies, Jennifer H. Pfeifer, Ashley A. Scott, Marian Sigman, Susan Y. Bookheimer, and Marco Iacoboni. "Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders." eng. In: *Nature Neuroscience* 9.1 (Jan. 2006), pp. 28–30.
DOI: [10.1038/nn1611](https://doi.org/10.1038/nn1611).
- [106] Lee Friedman and Gary H Glover. "Report on a multicenter fMRI quality assurance protocol". In: *Journal of Magnetic Resonance Imaging* 23.6 (2006), pp. 827–839.
- [107] Nouchine Hadjikhani, Robert M Joseph, Josh Snyder, and Helen Tager-Flusberg. "Anatomical differences in the mirror neuron system and social cognition network in autism". In: *Cerebral cortex* 16.9 (2006), pp. 1276–1282.
- [108] Xiao Han, Jorge Jovicich, David Salat, Andre van der Kouwe, Brian Quinn, Sylvester Czanner, Evelina Busa, Jenni Pacheco, Marilyn Albert, Ronald Killiany, et al. "Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer". In: *Neuroimage* 32.1 (2006), pp. 180–194.
- [109] Jorge Jovicich, Sylvester Czanner, Douglas Greve, Elizabeth Haley, Andre van der Kouwe, Randy Gollub, David Kennedy, Franz Schmitt, Gregory Brown, James MacFall, et al. "Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data". In: *Neuroimage* 30.2 (2006), pp. 436–443.

- [110] John Rice. *Mathematical statistics and data analysis*. Nelson Education, 2006.
- [111] Gary J Kelloff, John M Hoffman, Bruce Johnson, Howard I Scher, Barry A Siegel, Edward Y Cheng, Bruce D Cheson, Joyce O'Shaughnessy, Kathryn Z Guyton, David A Mankoff, et al. "Progress and promise of FDG-PET imaging for cancer patient management and oncologic drug development". In: *Clinical Cancer Research* 11.8 (2005), pp. 2785–2808.
- [112] Nobuyuki Nishitani, Martin Schürmann, Katrin Amunts, and Riitta Hari. "Broca's region: from action to language". In: *Physiology* 20.1 (2005), pp. 60–69.
- [113] Christoph Scherfler, Klaus Seppi, Eveline Donnemiller, Georg Goebel, Christian Brenneis, Irene Virgolini, Gregor K Wenning, and Werner Poewe. "Voxel-wise analysis of [¹²³I]beta-CIT SPECT differentiates the Parkinson variant of multiple system atrophy from idiopathic Parkinson's disease." In: *Brain* 128 (Pt 7 July 2005), pp. 1605–1612. ISSN: 1460-2156.
DOI: [10.1093/brain/awh485](https://doi.org/10.1093/brain/awh485).
- [114] Naama Barnea-Goraly, Hower Kwon, Vinod Menon, Stephan Eliez, Linda Lotspeich, and Allan L Reiss. "White matter structure in autism: preliminary evidence from diffusion tensor imaging". In: *Biological psychiatry* 55.3 (2004), pp. 323–326.
- [115] Chadwick W. Christine and Michael J. Aminoff. "Clinical differentiation of parkinsonian syndromes: Prognostic and therapeutic relevance". In: *The American Journal of Medicine* 117.6 (2004), pp. 412–419. ISSN: 0002-9343.
DOI: [10.1016/j.amjmed.2004.03.032](https://doi.org/10.1016/j.amjmed.2004.03.032).
- [116] Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, et al. "Automatically parcellating the human cerebral cortex". In: *Cerebral cortex* 14.1 (2004), pp. 11–22.
- [117] E Luders, C Gaser, L Jancke, and G Schlaug. "A voxel-based approach to gray matter asymmetries". In: *Neuroimage* 22.2 (June 2004), pp. 656–664. ISSN: 1053-8119.
DOI: [10.1016/j.neuroimage.2004.01.032](https://doi.org/10.1016/j.neuroimage.2004.01.032).
- [118] S. Bruggink K. P. Ebmeier N. J. Dougall. "Systematic review of the diagnostic accuracy of ^{99m}Tc-HMPAO-SPECT in dementia". In: *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry* 12.6 (2004), pp. 554–570.

- [119] Stephen M Smith et al. "Advances in functional and structural MR image analysis and implementation as FSL." In: *NeuroImage* 23 Suppl 1 (2004), S208–S219. ISSN: 1053-8119.
DOI: [10.1016/j.neuroimage.2004.07.051](https://doi.org/10.1016/j.neuroimage.2004.07.051).
- [120] J. Stoeckel, N. Ayache, G. Malandain, P. M. Koulibaly, K. P. Ebmeier, and J. Darcourt. "Automatic Classification of SPECT Images of Alzheimer's Disease Patients and Control Subjects". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI*. Vol. 3217. Lecture Notes in Computer Science. Springer, 2004, pp. 654–662.
- [121] Rigoberto Pérez de Alejo, Jesús Ruiz-Cabello, Manuel Cortijo, Ignacio Rodriguez, Imanol Echave, Javier Regadera, Juan Arrazola, Pablo Avilés, Pilar Barreiro, Domingo Gargallo, et al. "Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks". In: *Magnetic resonance imaging* 21.8 (2003), pp. 901–912.
- [122] Walter Daelemans, Véronique Hoste, Fien Meulder, and Bart Naudts. "Combined Optimization of Feature Selection and Algorithm Parameters in Machine Learning of Language". In: *Machine Learning: ECML 2003* (Jan. 1, 2003).
DOI: [10.1007/978-3-540-39857-8_10](https://doi.org/10.1007/978-3-540-39857-8_10).
- [123] Peter Dixon. "The p-value fallacy and how to avoid it." In: *Canadian journal of experimental psychology = Revue canadienne de psychologie expérimentale* 57 (3 Sept. 2003), pp. 189–202. ISSN: 1196-1961.
- [124] I. Guyon and A. Elisseeff. "An Introduction to Variable and Feature Selection". In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182.
- [125] Aapo Hyvärinen and Yutaka Kano. "Independent component analysis for non-normal factor analysis". In: *New developments in psychometrics*. Springer, 2003, pp. 649–656.
- [126] A Winogrodzka, P Bergmans, J Booij, E A van Royen, J C Stoof, and E C Wolters. "[¹²³I]beta-CIT SPECT is a useful method for monitoring dopaminergic degeneration in early stage Parkinson's disease". In: *Journal of Neurology, Neurosurgery & Psychiatry* 74.3 (2003), pp. 294–298.
DOI: [10.1136/jnnp.74.3.294](https://doi.org/10.1136/jnnp.74.3.294).
- [127] David A Clausi. "An analysis of co-occurrence texture statistics as a function of grey level quantization". In: *Canadian Journal of remote sensing* 28.1 (2002), pp. 45–62.

- [128] Bruce Fischl et al. "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain." In: *Neuron* 33 (3 Jan. 2002), pp. 341–355. ISSN: 0896-6273.
- [129] Andrew Newberg, Abass Alavi, and Martin Reivich. "Determination of regional cerebral function with FDG-PET imaging in neuropsychiatric disorders". In: *Seminars in nuclear medicine*. Vol. 32. 1. Elsevier. 2002, pp. 13–34.
- [130] N Saeed and BK Puri. "Cerebellum segmentation employing texture properties and knowledge based image processing: applied to normal adult controls and patients". In: *Magnetic resonance imaging* 20.5 (2002), pp. 425–429.
- [131] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain." In: *Neuroimage* 15.1 (Jan. 2002), pp. 273–289.
DOI: [10.1006/nimg.2001.0978](https://doi.org/10.1006/nimg.2001.0978).
- [132] Marti J Anderson and John Robinson. "Permutation tests for linear models". In: *Australian & New Zealand Journal of Statistics* 43.1 (2001), pp. 75–88.
- [133] J. C. Baron, G. Chételat, B. Desgranges, G. Perchey, B. Landeau, V. de la Sayette, and F. Eustache. "In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease." eng. In: *Neuroimage* 14.2 (Aug. 2001), pp. 298–309.
DOI: [10.1006/nimg.2001.0848](https://doi.org/10.1006/nimg.2001.0848).
- [134] M Jenkinson and S Smith. "A global optimisation method for robust affine registration of brain images." In: *Medical image analysis* 5 (2 June 2001), pp. 143–156. ISSN: 1361-8415.
- [135] Vassili A Kovalev, Frithjof Kruggel, H-J Gertz, and D Yves von Cramon. "Three-dimensional texture analysis of MRI brain datasets". In: *Medical Imaging, IEEE Transactions on* 20.5 (2001), pp. 424–433.
- [136] J Mazziotta et al. "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)." In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 356 (1412 Aug. 2001), pp. 1293–1322. ISSN: 0962-8436.
DOI: [10.1098/rstb.2001.0915](https://doi.org/10.1098/rstb.2001.0915).

- [137] John Ashburner and Karl J Friston. "Voxel-based morphometry—the methods". In: *Neuroimage* 11.6 (2000), pp. 805–821.
- [138] R. P. W. Duin. "Classifiers in almost empty spaces". In: *Proceedings of the 15th International Pattern Recognition Conference*. Vol. 2. 2000, pp. 1–7. DOI: 10.1109/ICPR.2000.906006.
- [139] A. Hyvärinen and E. Oja. "Independent component analysis: algorithms and applications". In: *Neural Networks* 13.4-5 (2000), pp. 411–430. ISSN: 0893-6080.
DOI: 10.1016/S0893-6080(00)00026-5.
- [140] D. Kogure, H. Matsuda, T. Ohnishi, T. Asada, M. Uno, T. Kunihiro, S. Nakano, and M. Takasaki. "Longitudinal Evaluation of Early Alzheimer Disease Using Brain Perfusion SPECT". In: *The Journal of Nuclear Medicine* 41.7 (2000), pp. 1155–1162. URL: <http://jnm.snmjournals.org/cgi/content/abstract/41/7/1155>.
- [141] Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H Cook, Bennett L Leventhal, Pamela C DiLavore, Andrew Pickles, and Michael Rutter. "The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism". In: *Journal of autism and developmental disorders* 30.3 (2000), pp. 205–223.
- [142] A M Dale, B Fischl, and M I Sereno. "Cortical surface-based analysis. I. Segmentation and surface reconstruction." In: *NeuroImage* 9 (2 Feb. 1999), pp. 179–194. ISSN: 1053-8119.
DOI: 10.1006/nimg.1998.0395.
- [143] A. Hyvärinen. "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis". In: *IEEE Transactions on Neural Networks* 10.3 (1999), pp. 626–634.
- [144] Remi KS Kwan, Alan C Evans, and G Bruce Pike. "MRI simulation-based evaluation of image-processing and classification methods". In: *Medical Imaging, IEEE Transactions on* 18.11 (1999), pp. 1085–1097.
- [145] L-K Soh and Costas Tsatsoulis. "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices". In: *Geoscience and Remote Sensing, IEEE Transactions on* 37.2 (1999), pp. 780–795.
- [146] P Szatmari. "Heterogeneity and the genetics of autism." In: *Journal of psychiatry & neuroscience : JPN* 24 (2 Mar. 1999), pp. 159–165. ISSN: 1180-4882.

- [147] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. New York: Academic Press, 1999. ISBN: 9781597492720.
- [148] David Wechsler. *Wechsler abbreviated scale of intelligence*. Psychological Corporation, 1999.
- [149] L. P. Barnetson B. J. Shepstone K. A. Jobst. "Accurate prediction of histologically confirmed alzheimer's disease and the differential diagnosis of dementia: the use of NINCDS-ADRDA and DSM-III-R criteria, SPECT, x-ray CT, and apo e4 in medial temporal lobe dementias. Oxford project to investigate memory and aging," in: *International Psychogeriatrics* 10.3 (Sept. 1998), pp. 271–302.
- [150] V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., New York, 1998.
- [151] Andrew P Bradley. "The use of the area under the ROC curve in the evaluation of machine learning algorithms". In: *Pattern recognition* 30.7 (1997), pp. 1145–1159.
- [152] A. Hyvärinen and E. Oja. "A fast fixed-point algorithm for independent component analysis". In: *Neural Computation* 9 (1997), pp. 1483–1492.
- [153] Vladimir Vapnik, Steven E Golowich, Alex Smola, et al. "Support vector method for function approximation, regression estimation, and signal processing". In: *Advances in neural information processing systems* (1997), pp. 281–287.
- [154] S Arndt, T Cizadlo, D O'Leary, S Gold, and N C Andreasen. "Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain." In: *NeuroImage* 3 (3 Pt 1 June 1996), pp. 175–184. ISSN: 1053-8119.
DOI: 10.1006/nimg.1996.0019.
- [155] Timo Ojala, Matti Pietikäinen, and David Harwood. "A comparative study of texture measures with classification based on featured distributions". In: *Pattern Recognition* 29.1 (Jan. 1996), pp. 51–59. ISSN: 00313203.
DOI: 10.1016/0031-3203(95)00067-4.
- [156] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the royal statistical society. Series B (Methodological)* (1995), pp. 289–300.
- [157] R Kohavi. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In: *Proceedings of International Joint Conference on AI*. 1995, pp. 1137–1145. URL: <http://citeseer.ist.psu.edu/kohavi95study.html>.

- [158] Juliet Popper Shaffer. "Multiple hypothesis testing". In: *Annual review of psychology* 46.1 (1995), pp. 561–584.
- [159] Jui-Cheng Yen, Fu-Juay Chang, and Shyang Chang. "A new criterion for automatic multilevel thresholding". In: *Image Processing, IEEE Transactions on* 4.3 (Mar. 1995), pp. 370–378. ISSN: 1057-7149.
DOI: 10.1109/83.366472.
- [160] J. J. Claus, F. van Harskamp, M. M. B. Breteler, E. P. Krenning, I. de Koning abd J. M. van der Cammen, A. Hofman, and D. Hasan. "The diagnostic value of SPECT with Tc 99m HMPAO in Alzheimer's disease. A population-based study". In: *Neurology* 44.3 (1994), pp. 454–461.
DOI: 10.1159/000051283.
- [161] Karl J Friston, Andrew P Holmes, Keith J Worsley, J-P Poline, Chris D Frith, and Richard SJ Frackowiak. "Statistical parametric maps in functional imaging: a general linear approach". In: *Human brain mapping* 2.4 (1994), pp. 189–210.
- [162] Catherine Lord, Michael Rutter, and Ann Le Couteur. "Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders". English. In: *Journal of Autism and Developmental Disorders* 24.5 (1994), pp. 659–685. ISSN: 0162-3257.
DOI: 10.1007/BF02172145.
- [163] Alan C Evans, D Louis Collins, SR Mills, ED Brown, RL Kelly, and Terry M Peters. "3D statistical neuroanatomical models from 305 MRI volumes". In: *Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record*. IEEE. 1993, pp. 1813–1817.
- [164] Y Ma, M Kamber, and AC Evans. "3D simulation of PET brain images using segmented MRI data and positron tomograph characteristics". In: *Computerized medical imaging and graphics* 17.4-5 (1993), pp. 365–371.
- [165] RM Haralick and LG Shapiro. *Robot and Computer Vision (vols. 1 and 2)*. 1992.
- [166] J Talairach and P Tournoux. *Co-Planar Stereotactic Atlas of the Human Brain*. 1988.
- [167] Y. Vardi, L. A. Shepp, and L. Kaufman. "A Statistical Model for Positron Emission Tomography". In: *Journal of the American Statistical Association* 80.389 (1985), pp. 8–20.

- [168] M J de Leon, S H Ferris, A E George, B Reisberg, D R Christman, I I Kricheff, and A P Wolf. "Computed tomography and positron emission transaxial tomography evaluations of normal aging and Alzheimer's disease". In: *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism* 3.3 (Sept. 1983). PMID: 6603463, pp. 391–4.
- [169] David Freedman and David Lane. "A nonstochastic interpretation of reported significance levels". In: *Journal of Business & Economic Statistics* 1.4 (1983), pp. 292–298.
- [170] "Handbook of Statistics". In: ed. by R. R. Krishnaiah and L. N. Kanal. North-Holland, 1982. Chap. Dimensionality and sample size considerations in pattern recognition practice, pp. 825–855.
- [171] L. A. Shepp and Y. Vardi. "Maximum Likelihood Reconstruction for Emission Tomography". In: *IEEE Transactions on Medical Imaging* 1.2 (1982), pp. 113–122.
- [172] R. Haralick, K. Shanmugam, and I. Dinstein. "Textural features for image classification". In: *IEEE Transactions on Systems, Man and Cybernetics* 3.6 (1973), pp. 610–621.