



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:
Juan Manuel Górriz Sáez
Javier Ramírez Pérez de Inestrosa

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

Junio 2017

Francisco Jesus Martinez Murcia

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

Copyright © 2017

Titleback

This document was written with L^AT_EX on Linux using a modified ArsClassica, a reworking of the ClassicThesis style designed by André Miede, inspired to the masterpiece *The Elements of Typographic Style* by Robert Bringhurst.

Contacts

 fjesusmartinez@ugr.es

ABSTRACT

The rise of neuroimaging in the last years has provided physicians and radiologists 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 Computer Aided Diagnosis (**CAD**) pipeline: preprocessing, 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

Resumen de la tesis en español.

DECLARACIÓN

El doctorando Francisco Jesús Martínez Murcia y los directores de la tesis Juan Manuel Górriz Sáez y Javier Ramírez Pérez de Inestrosa 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:



Juan Manuel Górriz Sáez

Doctorando:



Francisco Jesús Martínez Murcia



Javier Ramírez Pérez de Inestrosa

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	6
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	11
2.1.2	Single Photon Emission Computed Tomography	13
2.1.3	Positron Emission Tomography	14
2.2	Medical Background	15
2.2.1	Alzheimer's Disease	15
2.2.2	Parkinsonism	16
2.2.3	Autism Spectrum Disorder	17
2.3	Voxelwise Analyses	17
2.3.1	Statistical Parametric Mapping	18
2.3.2	Voxel Based Morphometry	20
2.3.3	The Multiple Comparisons Problem	20
2.4	Machine Learning in Neuroimaging	23
2.4.1	Voxels as Features	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	35

II Reducing the Feature Space	37
4 Image Decomposition	39
4.1 Feature Selection	40
4.1.1 t-test	41
4.1.2 Kullback-Leibler Divergence	41
4.1.3 Mann-Whitney-Wilcoxon	41
4.2 Decomposition Algorithms	42
4.2.1 Factor Analysis	42
4.2.2 Independent Component Analysis	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	81
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	91
6.3 Sampling Paths via Hidden Markov Model (HMM)	93
6.3.1 Radial Texture Features	98
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	109
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	117
6.6.3	General Discussion of SBM	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	131
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 diseases	169
9.2	Conclusions	170
V	Appendix	171
A	Datasets	173
A.1	MRI	173

A.1.1	ADNI-MRI and the Alzheimer's Disease Neuroimaging Initiative	173
A.1.2	AIMS-MRI, MRC-AIMS Consortium	174
A.2	PET	176
A.2.1	ADNI-PET, Alzheimer's Disease Neuroimaging Initiative	176
A.3	SPECT	178
A.3.1	VDLN-HMPAO, Virgen de las Nieves	178
A.3.2	VDLN-DAT, Virgen de las Nieves	179
A.3.3	VDLV-DAT, Virgen de la Victoria Hospital	180
A.3.4	PPMI-DAT, Parkinson's Progression Markers Initiative	181
B	Background on Support Vector Machines	183
	Bibliography	185

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 T1 and T2-weighted MRI images.	12
Figure 2.2	Example of SPECT images.	14
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.	23
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 ADNI-PET.	63
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 (using the $I_{th} = 0.35I_{max}$) for PPMI-DAT, VDNL-DAT, VDLV-DAT database.	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.	92
Figure 6.7	SBM mapping of different cortical regions.	92
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) Grey Matter and b) White Matter.	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. 108
- 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 6.19 113
- Figure 6.20 Performance at the operation point for the different mappings over the Grey Matter and White Matter, compared with the performance of Voxels As Features (**VAF**). 115
- 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 **qT₁**, **qT₂**, **synT₁**, **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 **qT₁**, **qT₂**, **synT₁**, **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 **qT₁**, **qT₂**, **synT₁**, **GM** and **WM** maps before and after applying **SWPCA** to remove site effects. 139
- Figure 7.7 Location of hte singifcant 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 three 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.	157
Figure 8.5	Examples of some original and simulated subjects from the ADNI-PET dataset.	158
Figure 8.6	Examples of some original and simulated subjects from the PPMI dataset.	159
Figure 8.7	SPM Analysis of the ADNI dataset (AD vs CTL).	160
Figure 8.8	SPM Analysis of the ADNI dataset (CTL vs MCI).	161
Figure 8.9	SPM Analysis of the ADNI dataset (MCI vs AD).	162
Figure 8.10	SPM Analysis of the PPMI dataset.	163

LIST OF TABLES

Table 3.1	Confusion matrix and its parts	35
Table 4.1	Performance values for the Alzheimer's datasets	53
Table 4.2	Performance values for the Parkinson's datasets	60
Table 4.3	Percentage of overlap between the selected areas by each method and the AAL atlas regions.	62
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.	80
Table 5.4	Comparison between our proposed system and other Parkinson's Disease (PD) diagnosis systems in the literature.	84
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. 113
Table 6.4	Performance values (\pm SD) for each of the 10 texture features. 114
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 (Sen) 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 (Sen), 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. 173
Table A.2	Demographics of the ADNI-MRI dataset. 174
Table A.3	Demographics of the AIMS-MRI dataset. 175
Table A.4	Demographics of the ADNI-PET dataset. 177
Table A.5	Demographic of the VDLN-HMPAO dataset. 178
Table A.6	Demographics of the VDLN-DAT dataset. 179
Table A.7	Demographics of the VDLV-DAT dataset. 180
Table A.8	Demographics of the PPMI-DAT dataset. 181

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 these studies are common: a selection 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 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 a list of significant differences between structure or function that could be linked to the disease.

[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 [[102](#)]. 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* [[33](#)] which negatively affects the statistical power of any experiment performed using these datasets [[16](#)].

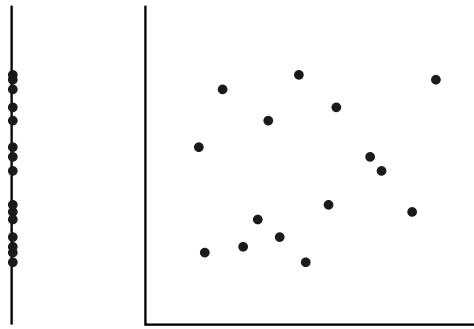


Figure 1.1: Illustration of the separation between points in one-dimensional and two-dimensional spaces.

1.2 The Small Sample Size Problem

The *Small Sample Size Problem* refers to a problem that arises when the proportion between number of subjects and 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 farther than in the two-dimensional world. And the same would happen if we move to four, 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* [33, 143], 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 one of the fundamental sources of non-reproducibility in current neuroimaging studies [16].

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 [52]. 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 [28, 49, 65, 102, 167] with great success, and solutions using this approach will be treated in Chapters 4, 5 and 6.

The Small Sample Size problem is directly related to the *Curse of Dimensionality* [80], which proves that, in contrast to what could 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 [11], 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 provide 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 previous sections: 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 PCA [75, 152] or Partial Least Squares [132], among others, are widely studied. We have developed three different approaches to those:

- A combination of image decomposition algorithms and feature selection. In this approach we have used three criteria to select the most significant voxels from the images, and then applied Factor Analysis (FA) and

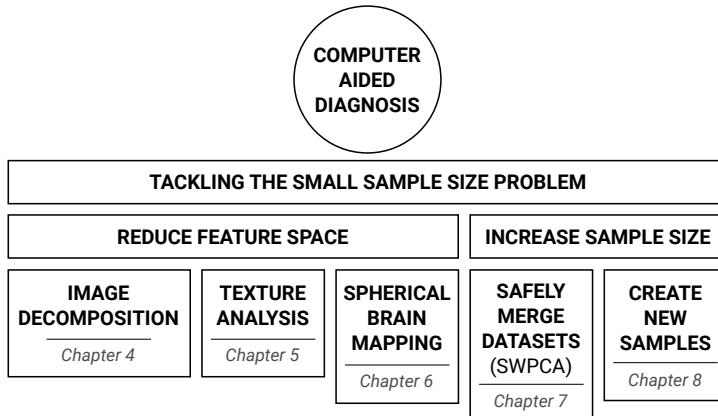


Figure 1.2: Structure and connexions between the different strategies proposed in this thesis, organized by chapters and parts.

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 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 analysis.
- 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 preprocessing 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 and conference presentations.

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. Martinez-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. Martinez-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 Magnetic Resonance Imaging ([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, linked to a receptor, can measure its distribution. This is the case of Positron Emission Tomography ([PET](#)) and Single Photon Emission Computed Tomography ([SPECT](#)).

2.1.1 Magnetic Resonance Imaging

Magnetic Resonance Imaging ([MRI](#)) is perhaps the most widespread 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, is considered non-invasive. [MRI](#) uses strong magnetic fields to excite certain atomic nuclei, that can absorb and emit this energy.

[MRI](#) combines the 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](#) use hydrogen atoms, since they are present in water (which adds up to around 70% of body mass) and

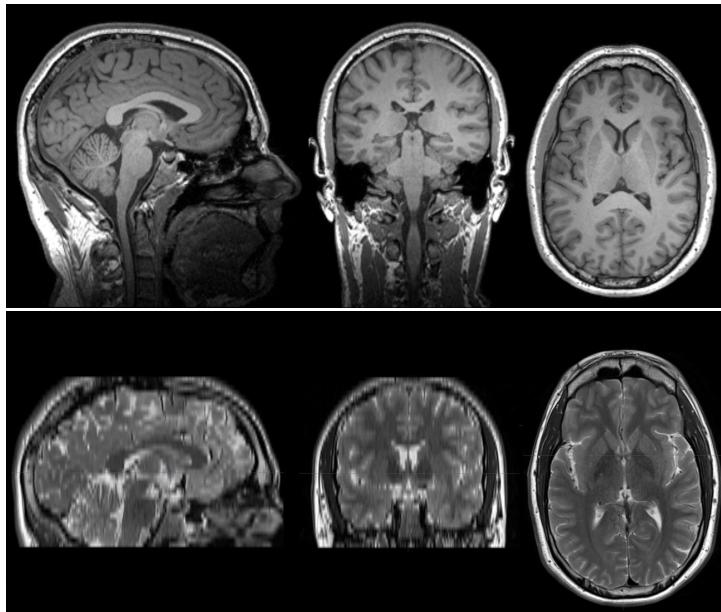


Figure 2.1: Example of T1 and T2-weighted MRI images of the same subject (me).

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, $\gamma = 42.6 \text{ MHz/T}$. 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

The Single Photon Emission Computed Tomography (SPECT) is based on the principles of Computed Tomography (CT), by which a series of signal acquisition 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 [71], 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 number of radiopharmaceutical used in clinical practice, and therefore, we will 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 (¹²³I), a cocaine analog with high binding affinity for Dopamine Transporters (DAT). It is

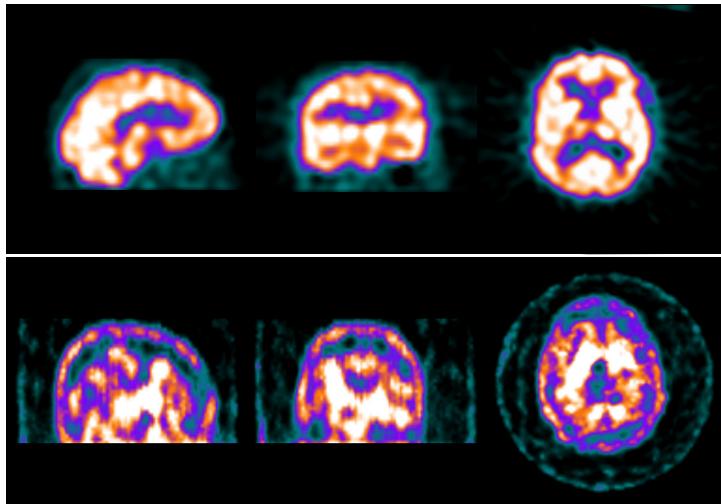


Figure 2.2: Example of [SPECT](#) images, a SPECT-HMPAO and a SPECT-DaTSCAN.

used fundamentally in the assessment of [PD](#), given that the disease is associated with a loss of dopaminergic neurons in the striatal region.

2.1.3 Positron Emission Tomography

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 PET-FDG, also known as Fludeoxyglucose (^{18}F). It is a glucose analogue that allows us to measure the glucose metabolism in the brain. It is widely used in neurology [128] and cancer detection [110], since it can be correlated with cellular activity.

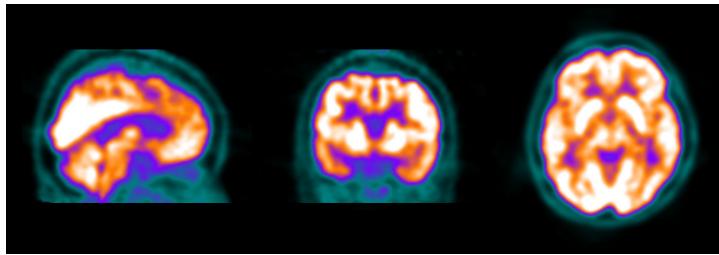


Figure 2.3: Example of a PET-FDG image.

2.2 Medical Background

2.2.1 Alzheimer's Disease

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 ($A\beta$) plaques or neurofibrillary tangles accompanied by synaptic dysfunction and neurodegeneration that eventually lead to cell death [[9, 51](#)].

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 diagnosis can only be addressed post-mortem, via a direct examination of the brain tissue [[51](#)]. 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, leading to a state 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 [[36, 73, 95, 132](#)] that neurodegeneration in Alzheimer's Disease mainly occurs in the [GM](#) tissue. Particularly grey matter loss has been described in the Hippocampus and Parahippocampal lobes, according to the NINCDS-ADRDA criteria for AD diagnosis [[95](#)], with further atrophy described in the medial temporal structures, the Posterior Cingulate gyrus and adjacent Precuneus [[132](#)]. 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 [[73](#)], and some structures in the striatum (putamen and

caudate nucleus) have shown important volume abnormalities [36]. 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 [159]. It also affects the frontal cortex and the whole brain in severe cases [139, 167].

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 [83]. 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

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)[96]. It is characterized by hypokinesia, rigidity, tremor and postural instability [96]. 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 [92, 114].

2.2.2.1 Parkinson's Disease

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, the ^{123}I -ioflupane (better known by its tradename DaTSCAN) is perhaps the most popular. DaTSCAN binds to the DAT at the striatum [96, 100, 125], allowing the estimation of DAT density by means of a SPECT scanner. In DaTSCAN images, a

reduced uptake at the striatum is a clear indication of DAT loss, and therefore, of PD [100].

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 [92]. 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 ¹⁸F-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

Autism Spectrum Disorder (ASD) is a neurodevelopmental syndrome characterized by social and communication impairment as well as restricted, repetitive patterns of behaviour, interests or activities. Its origins are still unknown, although research suggest [145] 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 Medical Research Council Autism Imaging Multicentre Study (MRC-AIMS) [31, 43] 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) [160] and its extension to structural imaging Voxel Based Morphometry (VBM) [136].

2.3.1 Statistical Parametric Mapping

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 [160]. 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 Alzheimer's Disease ([AD](#)) and [CTL](#)), 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 [97], 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 test 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. p-values are very extended in neuroimaging, representing the probability of a Z value being equal or more extreme than the reference value given. 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.

[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 active 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.

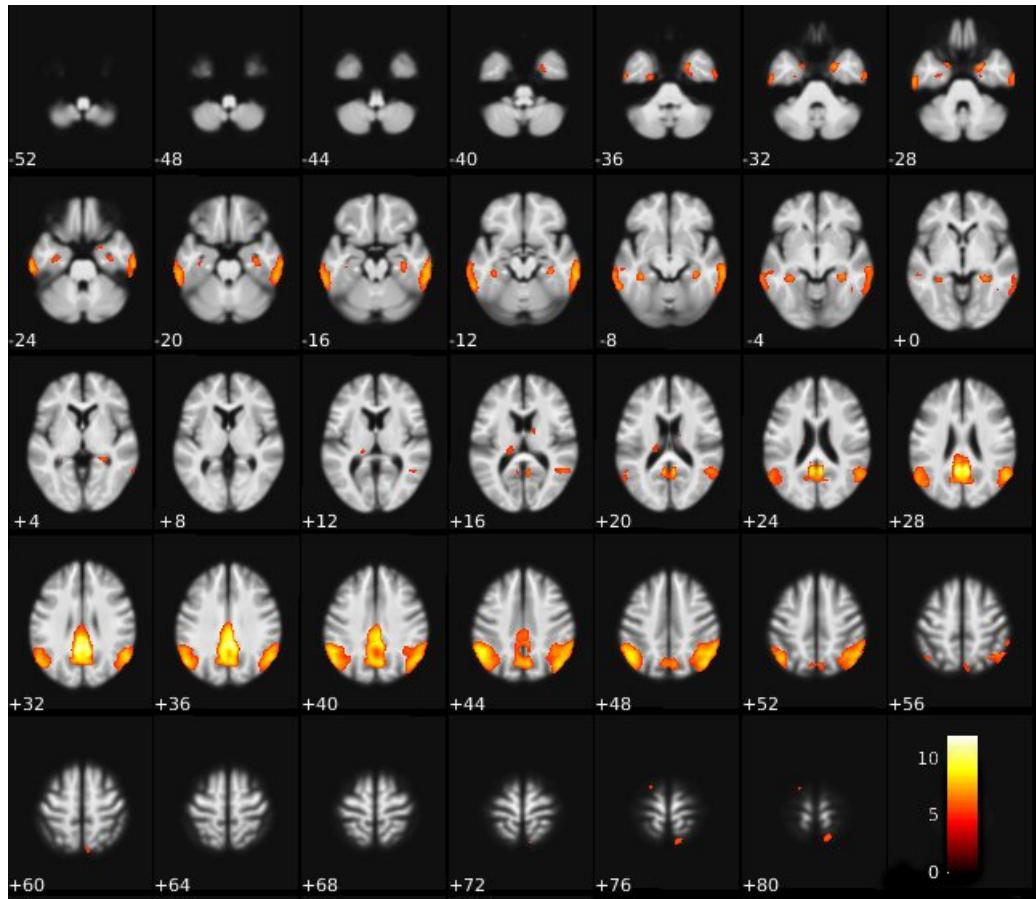


Figure 2.4: Example of a SPM analysis on a PET dataset displaying the differences between AD and CTL, using $p < 0.05$ and FWE correction.

2.3.2 Voxel Based Morphometry

Voxel Based Morphometry ([VBM](#)) can be considered an extention of [SPM](#) applied to structural [MRI](#) images [136]. The procedure involves preprocessing (see Chapter 3), where smoothing is applied to reduce 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 [80], or multidimensional Tensor Based Morphometry [62].

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 , 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 [122].

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 ($T_i > T_{\text{th}}$), 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:

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

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 [[157](#)] rewrites eq. [2.2](#) setting $\alpha = \frac{\alpha}{N}$ so that:

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

That way, using the Boole's inequality:

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

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** [97]).

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 neuroimaging can be found at [97].

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 [155]. The Benjamini and Hochberg method start with calculating the p-values of all voxels and ranking them so that:

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

2.3.3.3 FDR Controlling Procedures

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, 131]. 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. Th 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.

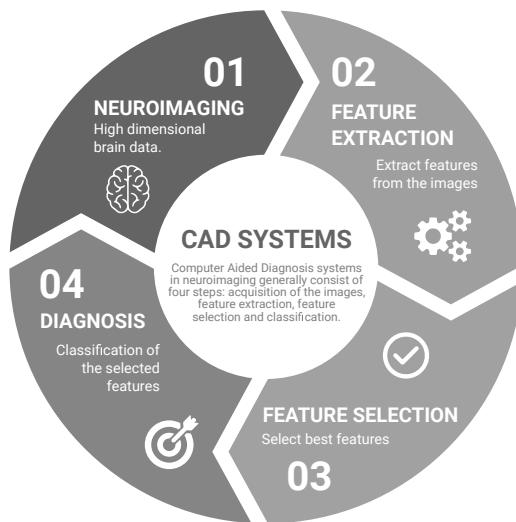


Figure 2.5: Illustration of a typical neuroimaging [CAD](#) system.

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 [119]. However, many more advances can be made in this field by exploring different types of

2.4.1 Voxels as Features

Voxels As Features (VAF) [119] 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, 76, 77], since it is comparable to the performance achieved by expert physicians using visual analysis [119]. 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.

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 [99, 103], but also the widespread Cortical Thickness [141] provided by *FreeSurfer*. 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 hypotrophy. Other pieces of software, such as SPM, also include toolboxes to compute cortical thickness, providing outputs such as the one that can be seen in Figure 2.6.

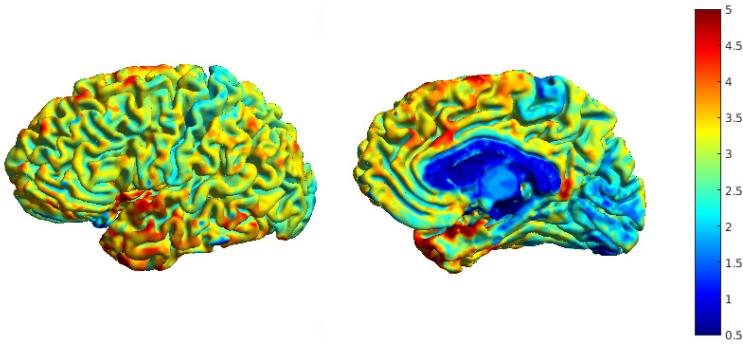


Figure 2.6: Example of the cortical thickness of a subject, obtained with the toolbox CAT12 in SPM12.

Other more advanced algorithms are image decomposition techniques such as PCA or Partial Least Squares (PLS), which have been extensively used in neuroimaging CAD systems [13, 39, 55, 59, 77]. 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 [123]. 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, 34]. The output of these tests has been already used as a tool in voxelwise analyses, as we commented in Section 2.3.

Wrappers are similar to filtering methods, given that 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 [156]. The most obvious measure here is accuracy, although other techniques such as Forward selection, backward elimination [123], genetic algorithms [156], or the expectation-maximization algorithm [69] 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 analysis techniques and [CAD](#) systems. We will apply them to many experiments, and use similar data and techniques in them. In this chapter, we will focus on the methodology that is common to most of these experiments, particularly focusing on 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, they can be classified 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 commonly 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, generally 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](#).

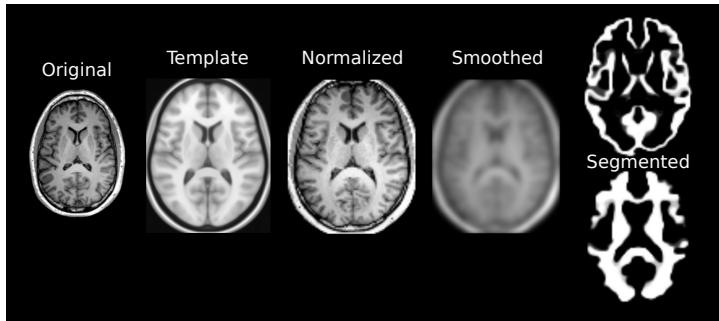


Figure 3.1: Typical pre-processing pipeline in [MRI](#).

3.1.1 Spatial Normalization or Registration

Spatial Normalization, also known as Registration, is the procedure that 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 [67] or FSL (in the FLIRT and FNIRT package) [118], most of them perform linear, non-rigid and elastic transformations or a combination of these. In this work we have used the software SPM8 [97] to perform registration of all the datasets, including [MRI](#), [SPECT](#) and [PET](#) images. So, from this moment, we will focus on the registration as performed in the Statistical Parametric Mapping Software, version 8 ([SPM8](#)).

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 that we found 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 [97] in the case of within-modality registration, or the mutual information in between-modality registration. These functions are also used in FLIRT [133], whereas FreeSurfer uses the Tukey's biweight function (in `mri_robust_template`) [47].

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 anatomically

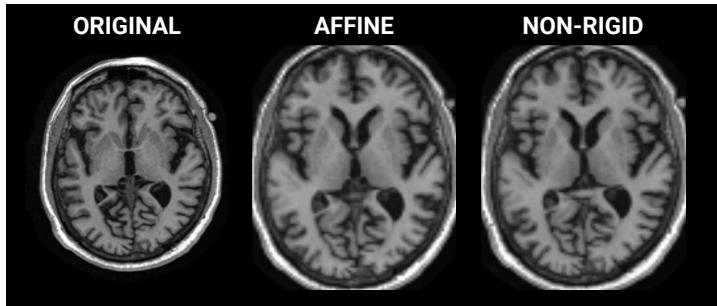
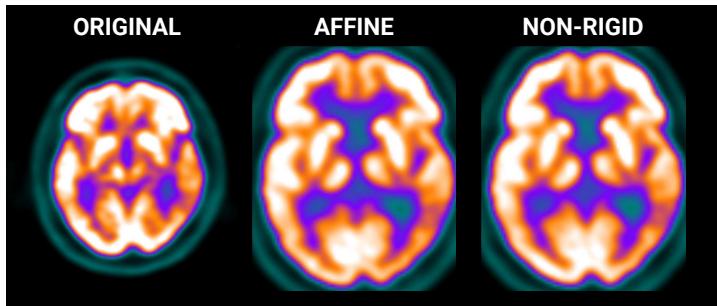
(a) Comparison in [MRI](#).(b) Comparison in [PET](#).

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.

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 [135]. 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 Tilarach space, a system based on a dissected brain, that was used to compose an atlas by Tilarach and Tournoux [165]. The current template is known as ICBM152, and features the average of 152 normal [MRI](#) scans matched to an older [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 lately some software can output data for bone, soft tissue or very detailed functional regions and subregions [127].

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: contrast uptake, radiotracer 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 are comparable.

In the case of intensity normalization, the method acts as a linear transformation of the image, 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 that is unique for each image. After this division, the new intensities would be directly comparable. The technique used to compute the normalization parameter varies, ranging from the simplest normalization to the maximum [46, 76] 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 [112].

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 [153], the gaussian distribution or the alpha-stable distribution [38]. 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 ([qT1](#)) and quantitative T2 - weighted ([qT2](#)) images [42] 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 images.

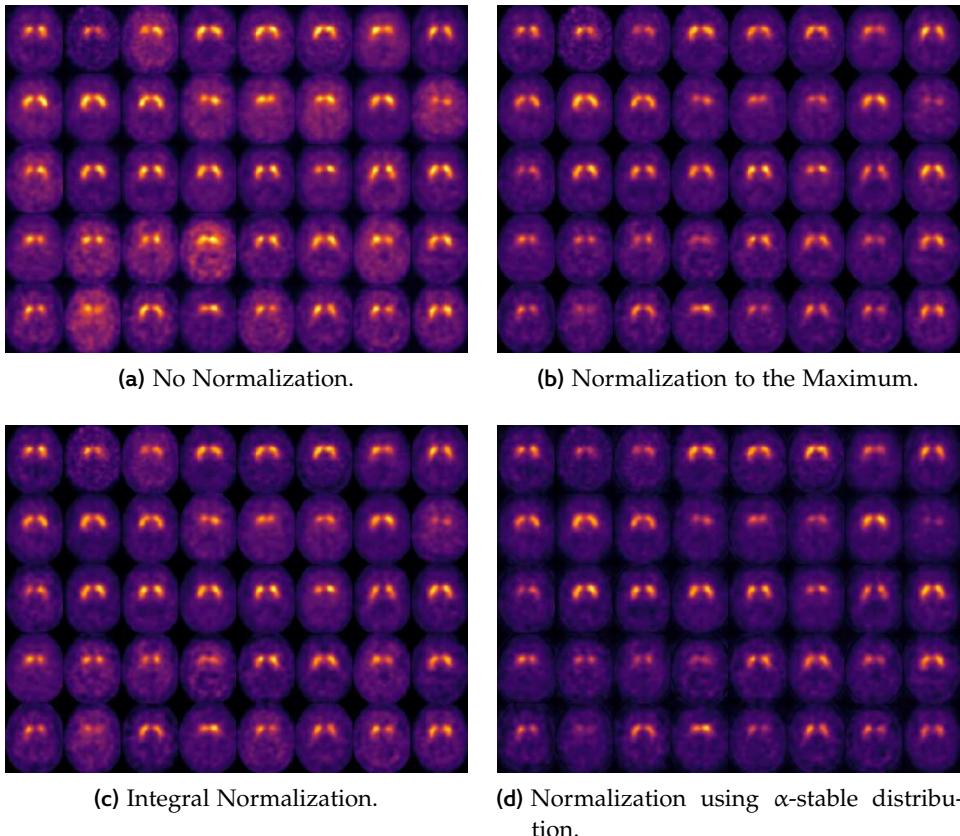


Figure 3.3: Comparison between different types of intensity normalization, applied to the VDLN-DAT dataset (see Appendix A).

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 [150]. 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 testing.

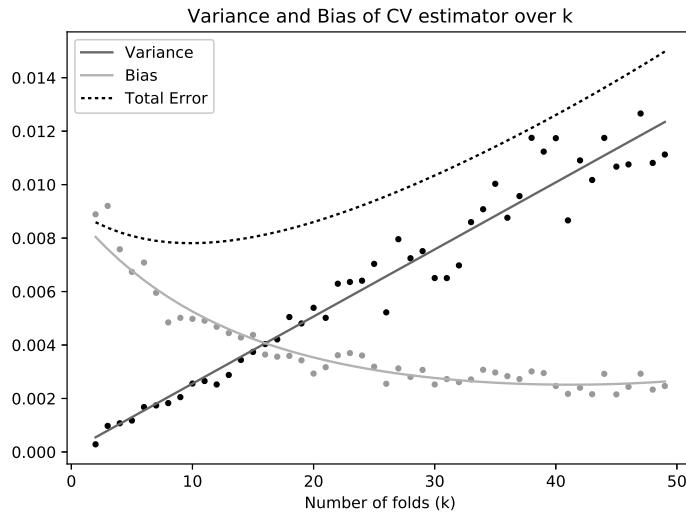


Figure 3.4: Evolution of bias and variance when increasing the number of folds in a k-fold [CV](#).

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 [70]. 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.

Using $k = 10$ is assumed as a good compromise between variance and bias in many works [70, 156]. In this thesis, when referring to k-fold, we often use stratified cross validation, which is a subclass of k-fold where the distribution of classes within each fold is similar to the distribution of classes in the whole dataset, making the estimates more accurate [156].

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

Table 3.1: Confusion matrix and its parts

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.

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 the positive).

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, 34, 46, 58].

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 vector of length c . The i -th neuroimage in our dataset can be therefore decomposed as:

$$\mathbf{x}_i = s_0 \mathbf{w}_0 + s_1 \mathbf{w}_1 + \cdots + s_c \mathbf{w}_c + \epsilon = \mathbf{s} \mathbf{W} + \epsilon \quad (4.1)$$

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.

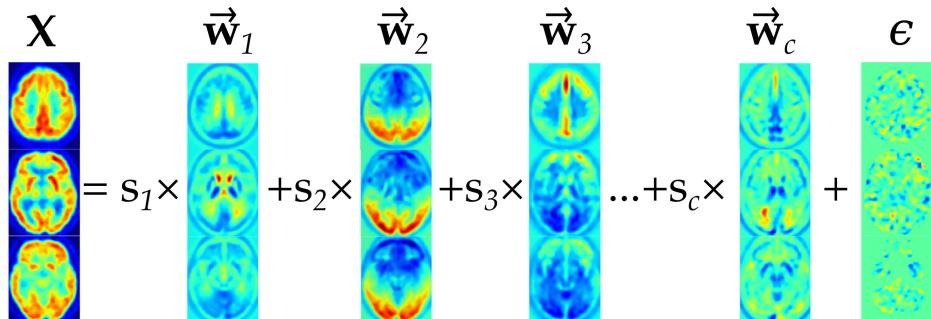


Figure 4.1: Illustration of how decomposition algorithms such as **FA** and **ICA** work on a **PET-FDG** brain image.

Many signal decomposition techniques are used in the literature, for example **PCA** or **PLS** [13, 39, 55, 59, 77]. 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.

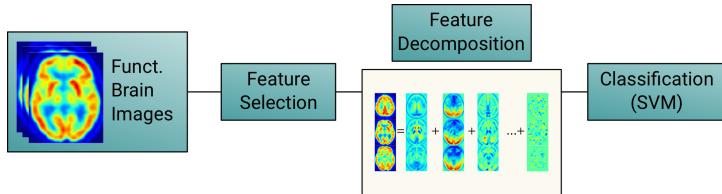


Figure 4.2: Illustration of the system used in Chapter 4.

4.1 Feature Selection

Feature selection is the first strategy used for feature reduction [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 [123]. 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 [121].

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 [101].

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 [46, 58]. This also removes noise in some cases where the decomposition algorithm cannot compute correctly the variance.

Three feature selection algorithms have been used in this thesis, not only in the **CAD** systems proposed in this chapter, but 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 [65]. It quantifies the differences between two classes using an assumption of 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 [97]. See figure 4.13a for an example of the t-test computed on the ADNI-PET database.

4.1.2 Kullback-Leibler Divergence

Another alternative is the Kullback-Leibler ([KL](#)) divergence, also known as Relative Entropy. 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 [KL](#) divergence can be calculated with equation 4.3 [146].

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

using the same notation than in t-test. See figure 4.13b for an example of the computed [KL](#) divergence on the ADNI-PET database.

4.1.3 Mann-Whitney-Wilcoxon

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 the 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 $\min U_1^f, U_2^f$ [65], but the usual approach nowadays is to assign $U^f = U_2^f$. Unlike t-test, MW test does not assume any prior distribution, and therefore is less likely than it to spuriously indicate significance because of the presence of outliers. Under the normal distribution, it performs relatively similar [65]. See figure 4.13c for an example of the MWU 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 ??).

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

Factor Analysis (FA) was used in [46, 58] 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 (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

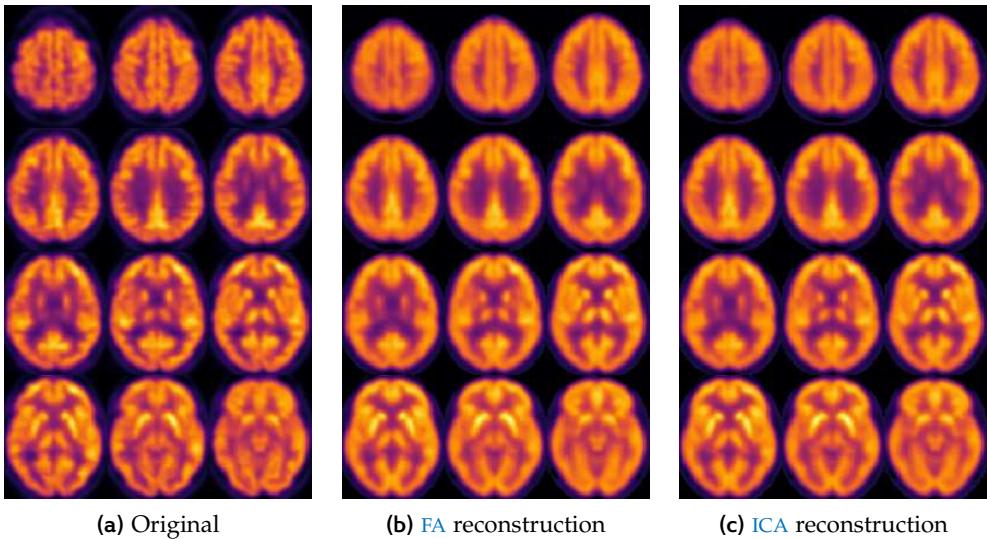


Figure 4.3: Original PET image from the ADNI-PET dataset, and examples of reconstruction using FA or ICA, with 10 components.

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:

$$X - \mu = SW + \epsilon \quad (4.7)$$

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

Now we can rewrite Eq. 4.7 as:

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

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

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

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.

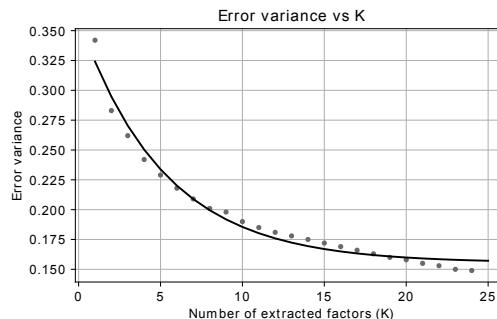


Figure 4.4: Specific variance of reconstruction error $\boldsymbol{\Psi}$ using FA, in function of number of factors extracted (K) for ADNI-PET database (the behaviour is similar in other datasets).

4.2.2 Independent Component Analysis

Independent Component Analysis ([ICA](#)) [[138](#)] is an algorithm that performs decomposition imposing that the resulting components must be independent. It was used in [[46](#), [58](#), [72](#)] as part of a [CAD](#) system, and it had been used in other medical imaging applications such as segmentation [[94](#)].

[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 [[138](#)]. To do so, we assume the source signals to be non-gaussian, in addition to the independence assumption that we mentioned before. That is why their authors consider [ICA](#) to be a non-gaussian version of [FA](#) [[124](#)], 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 remains:

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

where \mathbf{S} are the component scores and \mathbf{W} are the component loading, ‘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 [[138](#)]. 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:

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

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

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

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} .

Thanks to the central limit theorem, we assume that the sum of a large number of independent random variables tends will be approximately normally distributed, regardless of the individual statistical distributions [109]. This property is used to maximize non-gaussianity and independence in the sources using any independence criteria such as the kurtosis or negentropy in any of the proposed algorithms. In this work, we will use the FastICA algorithm.

4.2.2.1 FastICA

FastICA is a block fixed-point iteration algorithm [142, 151] based on negentropy as a non-gaussianity measure. Fixed-point algorithms are converge faster than adaptive algorithms [142]. The FastICA algorithm can be considered a neural algorithm [138], where the weight vector \mathbf{w} can be updated using a learning rule. FastICA defines a learning rule that finds a direction \mathbf{w} , a unit vector such that the projection $\mathbf{w}^T \mathbf{x}_i$ maximizes non-gaussianity [142].

The non-gaussianity measure used here is the negative entropy, or negentropy. The negentropy is a form of differential entropy, which for a random vector \mathbf{y} is defined as:

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

where $\mathbf{y}_{\text{gauss}}$ and \mathbf{y} share the same covariance matrix, although \mathbf{y} is not a gaussian random variable, and $\mathbf{y}_{\text{gauss}}$ 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 $G(x)$ is any non-quadratic function. Many functions have been proposed, but in the FastICA algorithm we use either $G(x)_1 = (1/\alpha_1) \log \cosh \alpha_1 x$ with $1 < \alpha_1 < 2$ or $G(x)_2 = \exp(-x^2/2)$ [142].

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 [142] as:

1. Choose an initial (e.g. random) weight vector \mathbf{w} .
2. Let $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$
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 in the FastICA algorithm. Additional modifications were presented in [138], in which step 2 is converted to a Newton iteration and further simplification is performed.

This is the algorithm for one computational unit, or neuron, which computes one component. However, the procedure can be extended to c components by defining c neurons with weight vectors $\mathbf{w}_1, \dots, \mathbf{w}_c$ so that $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$. The outputs $\mathbf{w}_1^T \mathbf{x}, \dots, \mathbf{w}_n^T \mathbf{x}$ must be decorrelated to prevent them from converging to the same maxima, using three methods proposed in [138].

The method used in this work uses a two-step iterative algorithm [138] to decorrelate the outputs after each iteration:

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

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 using either FA or ICA. Finally, the feature vectors are classified using a SVC with linear kernel, and performance values are obtained via cross-validation (see Section 3.3 for more information).

We vary the number of selected voxels and the number of factors or components depending on the algorithm and the dataset used and evaluate 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: 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 use [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 it 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 is 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).

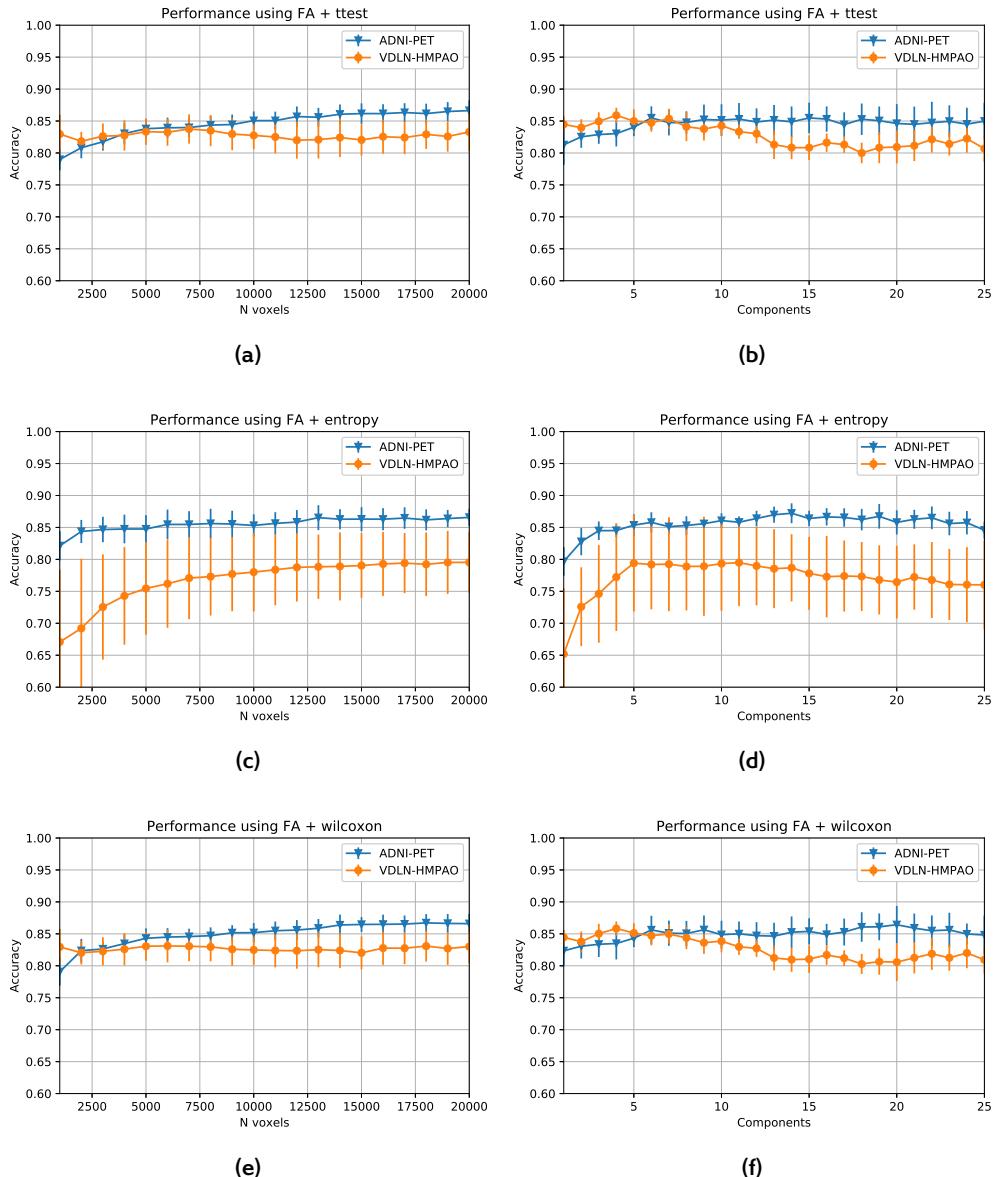


Figure 4.5: Average performance and standard deviation of the proposed system using the two AD datasets, FA and the three feature selection criteria: t-test ((a) and (b)), relative entropy ((c) and (d)) and wilcoxon ((e) and (f)).

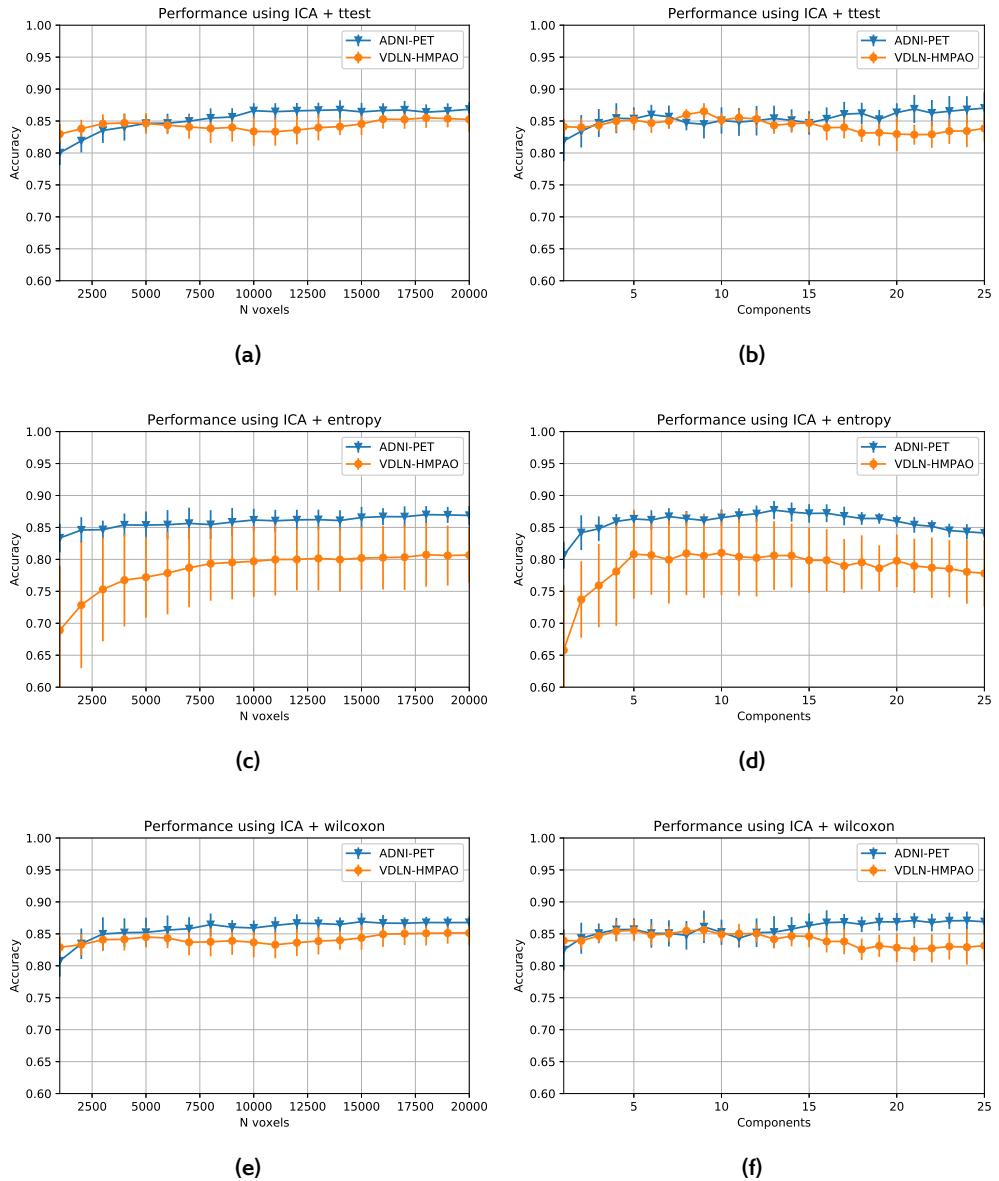


Figure 4.6: Average performance and standard deviation of the proposed system using the three AD datasets, ICA and the three feature selection criteria: t-test ((a) and (b)), relative entropy ((c) and (d)) and wilcoxon ((e) and (f)).

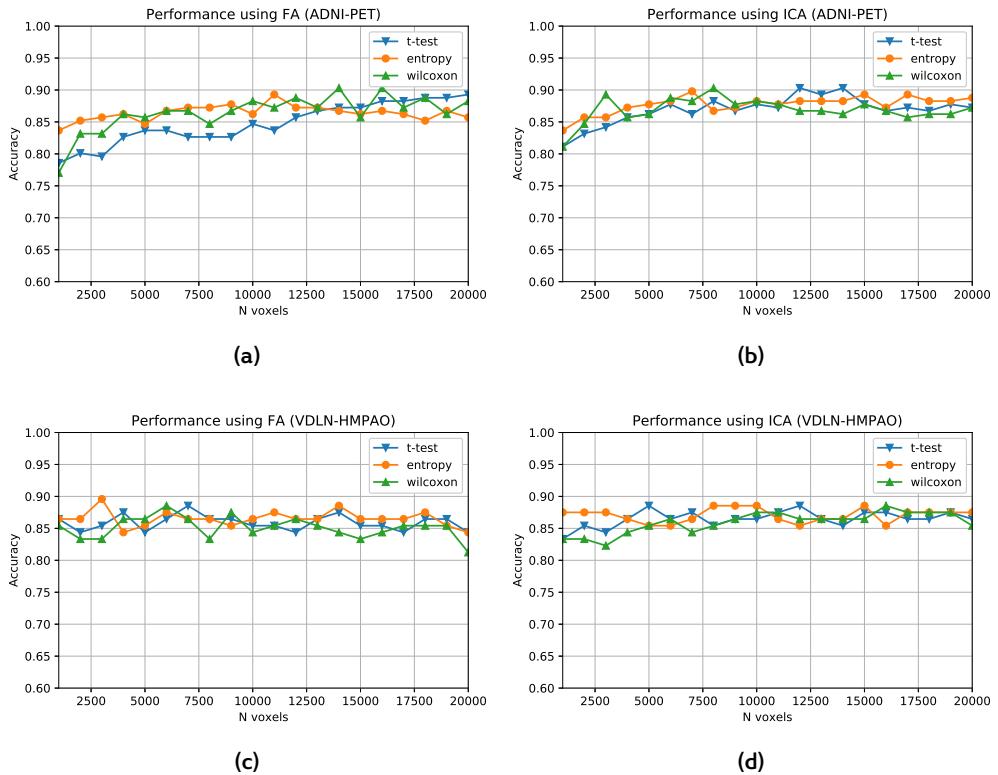


Figure 4.7: Performance of the proposed system using the two AD datasets: ADNI-PET and VDNL-HMPAO at the operation point, and how they vary over the number of selected voxels.

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 there is always a tendency of slightly increase in both datasets and decomposition strategies, as can be seen in 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 VDNL-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 there is

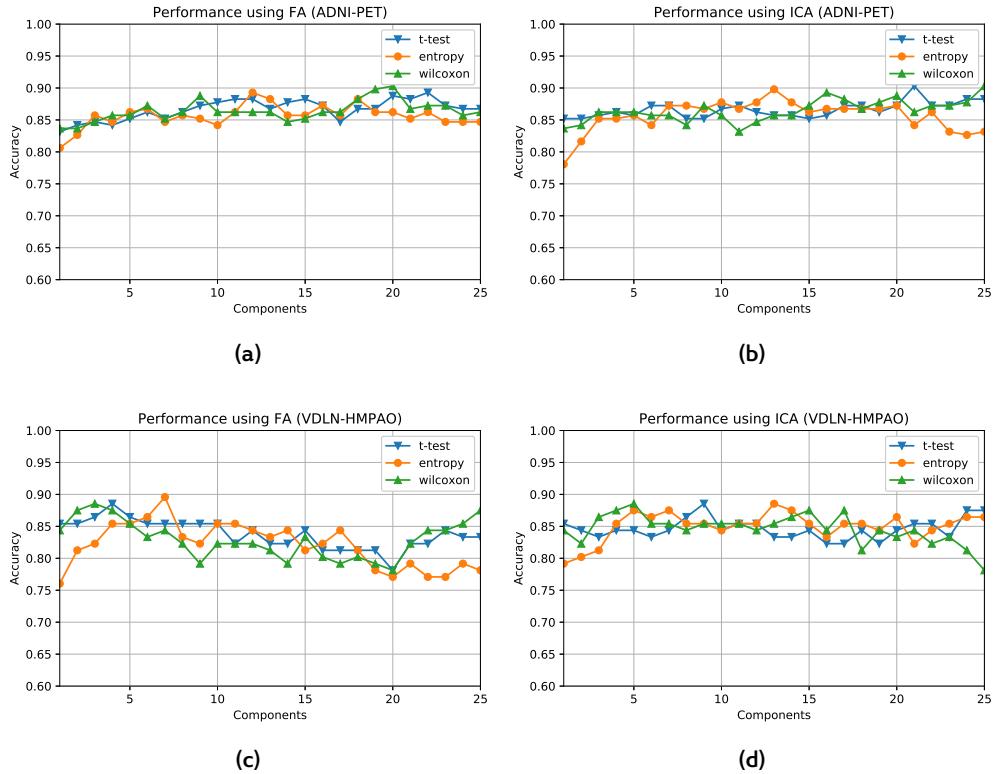


Figure 4.8: Performance of the proposed system using the two [AD](#) datasets: [ADNI-PET](#) and [VDLN-HMPAO](#) at the operation point, and how they vary over the number of components used in the decomposition.

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 between 5 components. But with the [ADNI-PET](#) dataset, the performance keeps 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 test combining decomposition algorithms and selection criteria, for the two datasets analysed in this section. It is obvious that both datasets obtain similar performance in almost every case, with values close to 0.9. These values are compared to the baseline classification performance, quantified using [VAF](#) [119]. From this, we can see that the decomposition systems always perform

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	entropy	0.893 ± 0.074	0.894 ± 0.092	0.891 ± 0.088
		wilcoxon	0.903 ± 0.066	0.917 ± 0.079	0.891 ± 0.082
	ICA	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
		wilcoxon	0.903 ± 0.066	0.906 ± 0.097	0.901 ± 0.094
VDLN-HMPAO	VAF	-	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
	FA	entropy	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
	ICA	t-test	0.885 ± 0.073	0.923 ± 0.130	0.825 ± 0.154
		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.

better than the baseline, but this difference is 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 components, 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 varying 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. In the VDLN-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 VDLN-DAT in almost all cases.

Their behaviours are consistent. Usually, there is no variation with the number of voxel selected (except in the obvious case of the VDLN-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 achieve a decent number (between 4 and 6), the performance starts to decrease. This could mean that the decomposition in more than 5 or 6 components only introduces noise and leads to a wrong decomposition.

Again, this behaviour is consistent with the VDLN-DAT dataset, except for one case. It does seem that the interaction between the VDLN-DAT dataset and the relative entropy selection criterion leads to a wrong 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 VDLN-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 VDLN-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.

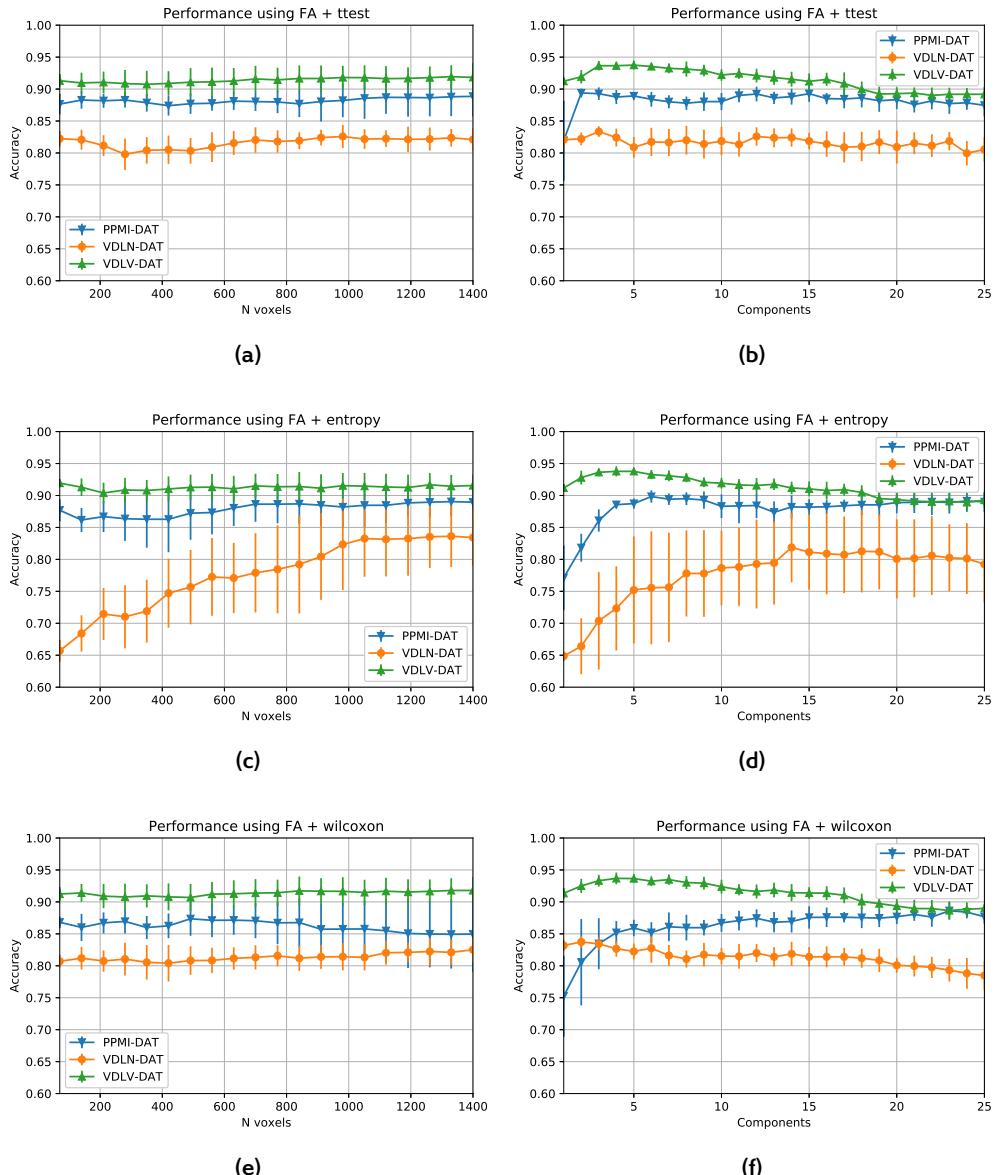


Figure 4.9: Average performance and standard deviation of the proposed system using the three PKS datasets, FA and the three feature selection criteria: t-test ((a) and (b)), relative entropy ((c) and (d)) and wilcoxon ((e) and (f)).

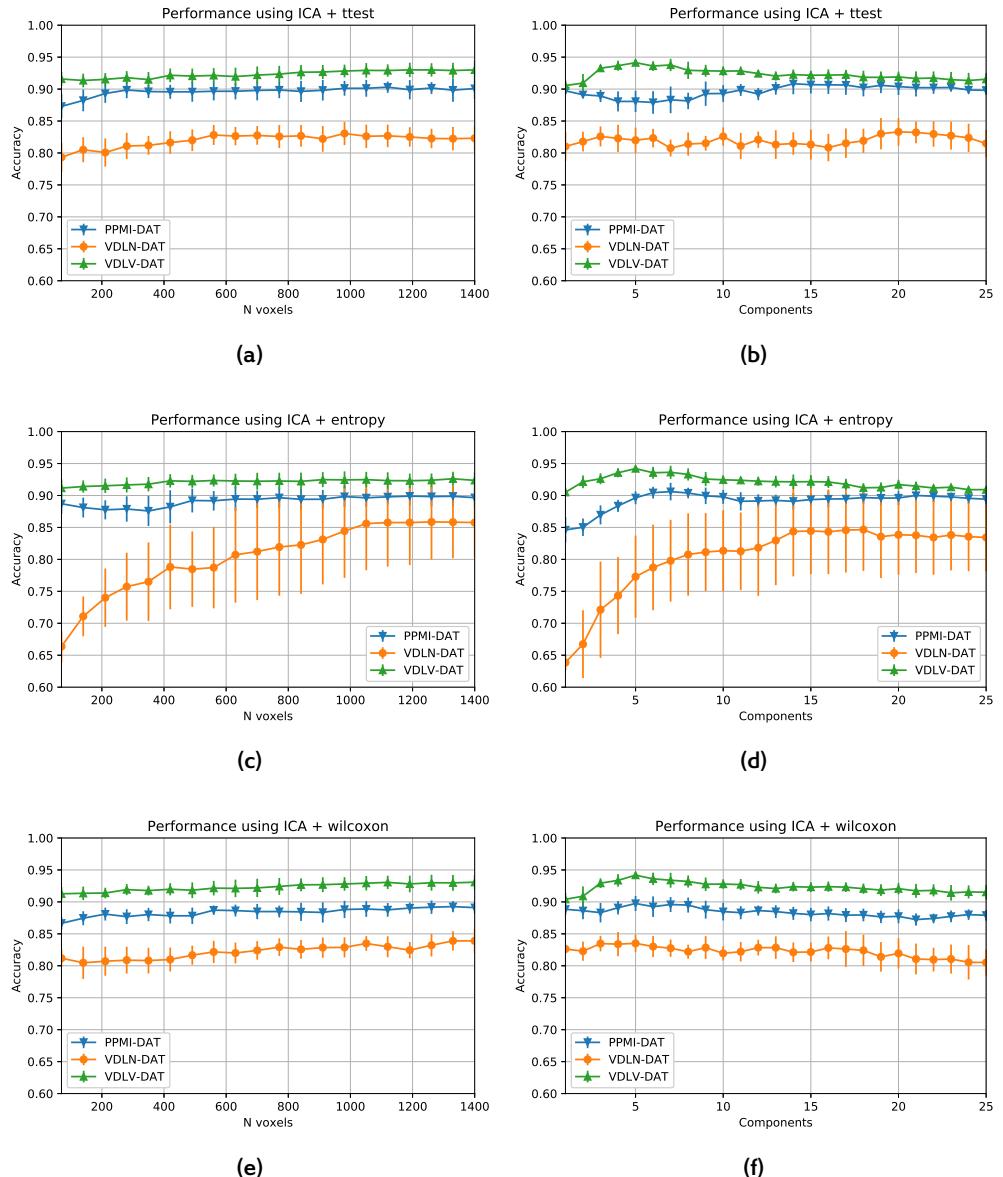


Figure 4.10: Average performance and standard deviation of the proposed system using the three PKS datasets, ICA and the three feature selection criteria: t-test ((a) and (b)), relative entropy ((c) and (d)) and wilcoxon ((e) and (f)).

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 when tested on both the PPMI-DAT and the VDLV-DAT datasets. In contrast, the performance increases when increasing the number of selected voxels when testing the VDLN-DAT dataset, especially when using the relative entropy criterion. This latter dataset also obtains less performance, whereas the VDLV-DAT achieves the best.

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, VDLN-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 will discuss this issue in more detail later.

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 VDLN-DAT achieves striking results. For this dataset, the higher performance is obtained with more than 10 components ($c = 14$), and 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 VDLN-DAT (in VDLV-DAT all combinations perform equally well).

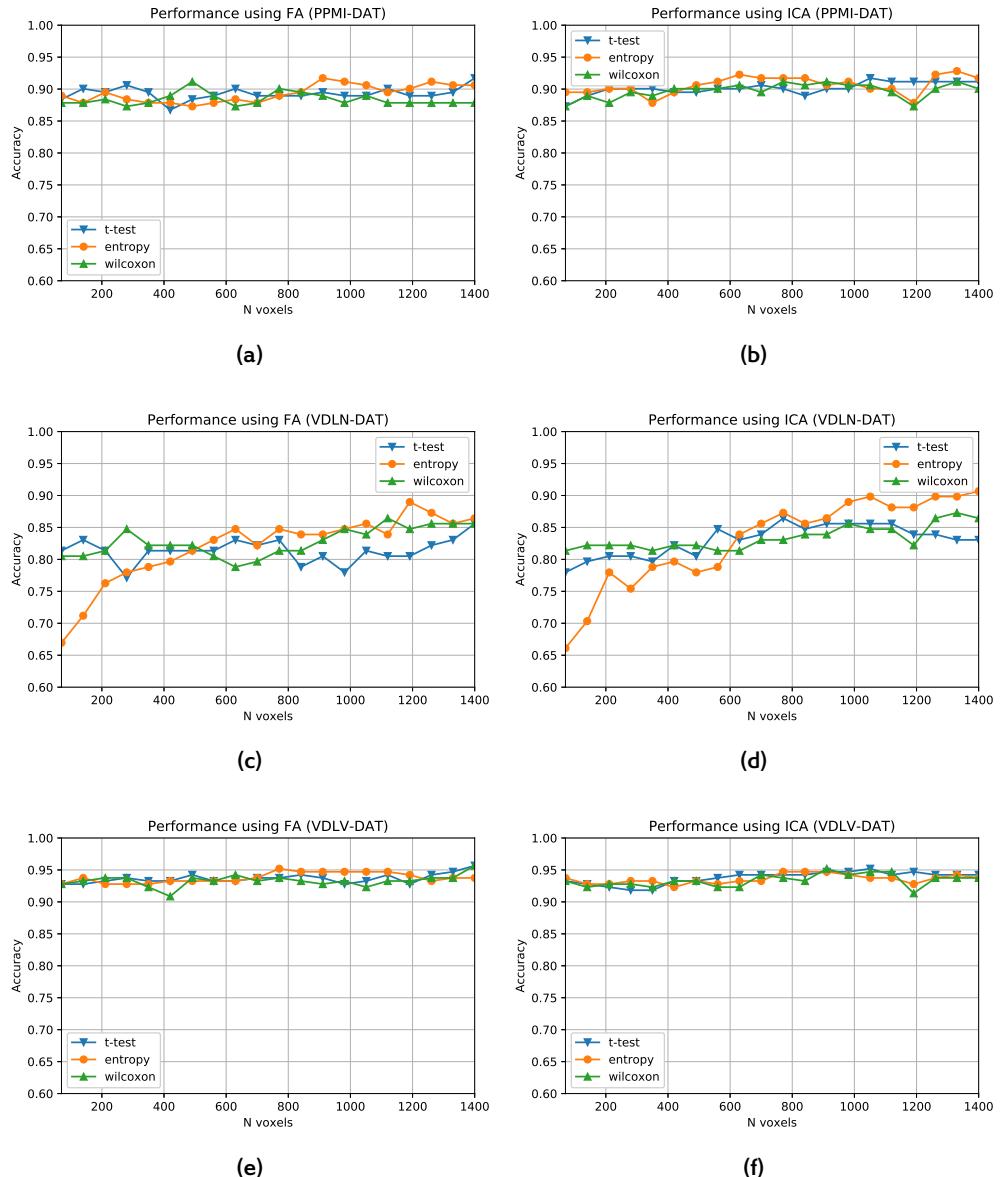


Figure 4.11: Performance of the proposed system using the two PKS datasets: PPMI-DAT, VDLN-DAT and VDLV-DAT at the operation point, and how they vary over the number of selected voxels.

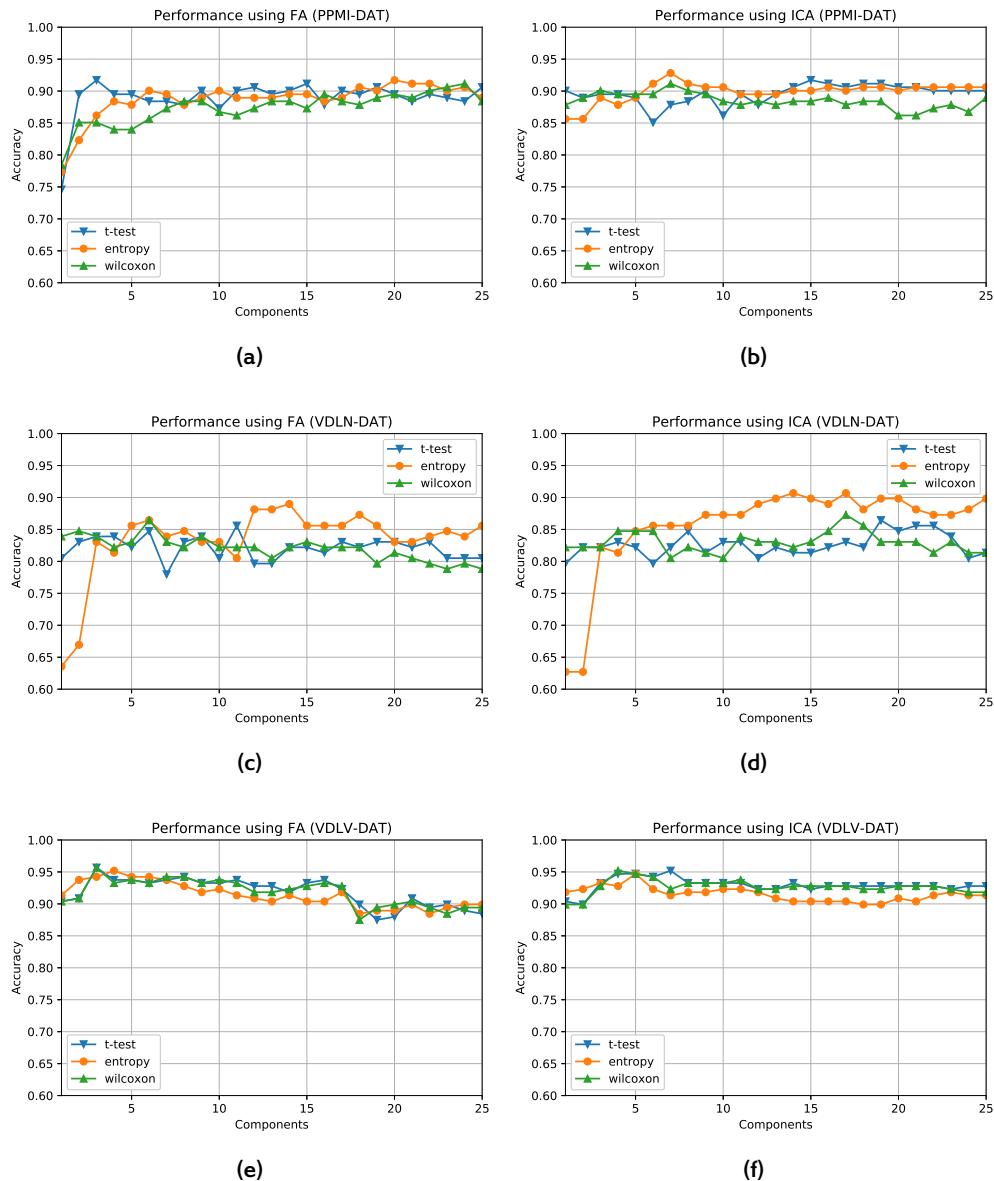


Figure 4.12: Performance of the proposed system using the two PKS datasets: PPMI-DAT, VDNL-DAT and VDLV-DAT at the operation point, and how they vary over the number of components used in the decomposition.

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
VDLN-DAT	VAF	-	0.796 ± 0.129	0.860 ± 0.143	0.675 ± 0.208
		t-test	0.856 ± 0.111	0.887 ± 0.178	0.795 ± 0.164
	FA	entropy	0.890 ± 0.098	0.875 ± 0.118	0.910 ± 0.116
		wilcoxon	0.864 ± 0.070	0.916 ± 0.114	0.780 ± 0.183
	ICA	t-test	0.864 ± 0.101	0.873 ± 0.174	0.840 ± 0.166
		entropy	0.907 ± 0.075	0.889 ± 0.124	0.935 ± 0.131
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 by means of **VAF** [119] in both cases. In [58] and [46], 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, linear **SVC** has been proven to be better generalizable than other systems, even in environments where the small sample size is the norm [152].

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

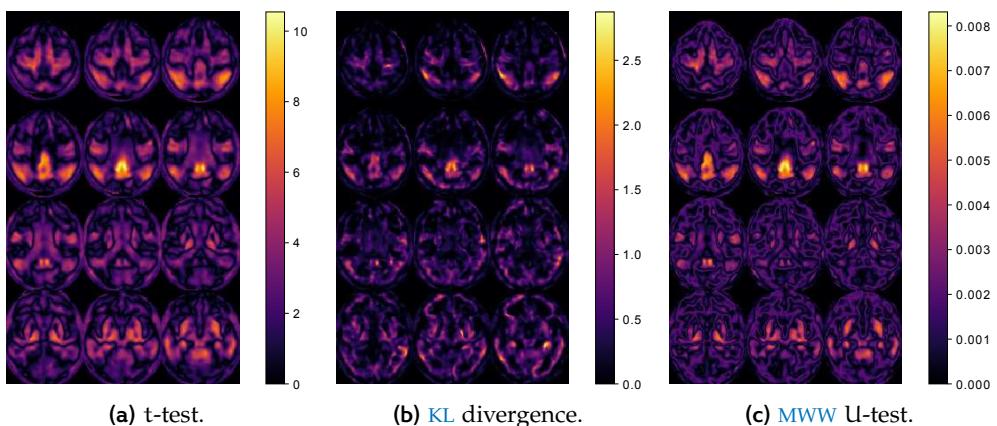


Figure 4.13: Comparison between the different filtering methods, and the regions selected by them, in the ADNI-PET dataset.

In the VDNL-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.

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.

Region	ADNI-PET			VDLN-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.

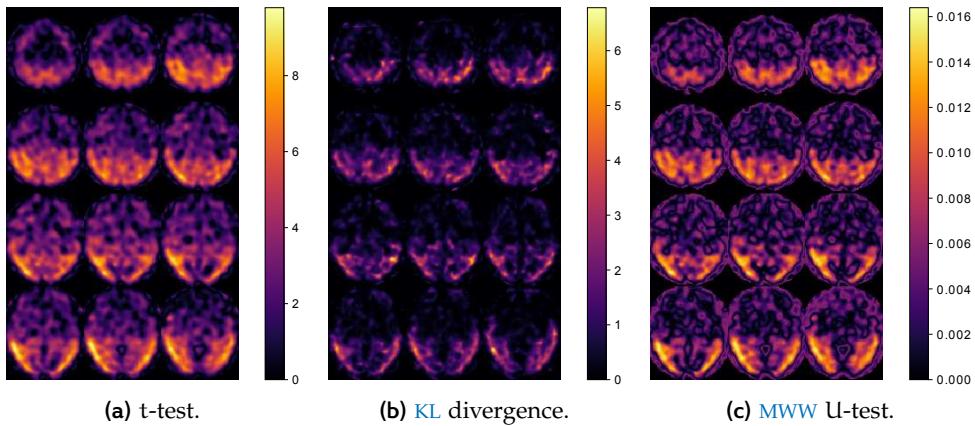


Figure 4.14: Comparison between the different filtering methods, and the regions selected by them, in the ADNI-PET dataset.

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, and other regions with a widely documented relation to **AD** are highlighted [95, 159], such as the cingulum, hippocampus and parahippocampal gyrus.

It is clearly noticeable that the relative entropy selection criterion focuses on many different 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 [65], and leads to a higher performance in the systems 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 VDLP-HMPAO, we can observe differences 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 both datasets, regardless of the noise contained in the images, although the differences with the baseline

system are much higher in the case of the VDLN-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 VDLN-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 VDLN-DAT to the VDLV-DAT. These databases differ in the number of subjects that 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 found throughout all figures and tables, although in general, the VDLN-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 VDLN-DAT dataset.

For its part, there are more significant 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 VDLN-DAT. The optimal c is usually located at $c \in [3, 5]$ in most cases, as can be seen in Figure 4.9. A very similar behaviour is achieved when using the [ICA](#) decomposition, as Figure 4.10 shows.

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 in 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.

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. How-

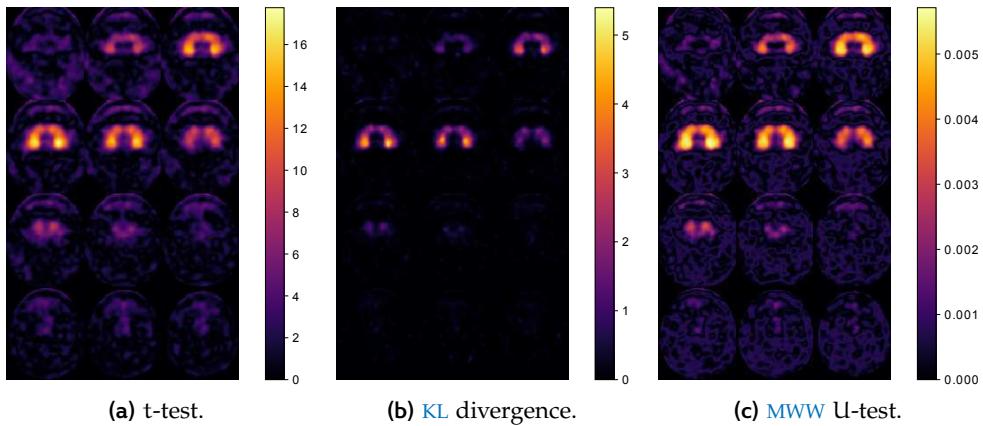


Figure 4.15: Comparison between the different filtering methods, and the regions selected by them, in the PPMI-DAT dataset.

ever, 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 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 defined as any procedure by which we can quantify and classify the spatial variation of intensity throughout an image. In neuroimaging, texture has been widely used in segmentation (tissue classification) of [MRI](#) images [79, 120, 129], although there exist a number of works using it for feature extraction in [CAD](#)-like systems, like the works in [16, 134], or our work on [PKS](#) feature extraction [25, 33].

Texture features can be classified in first, second and higher order analysis, 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 [90].

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 most popular the Haralick texture analysis [171].

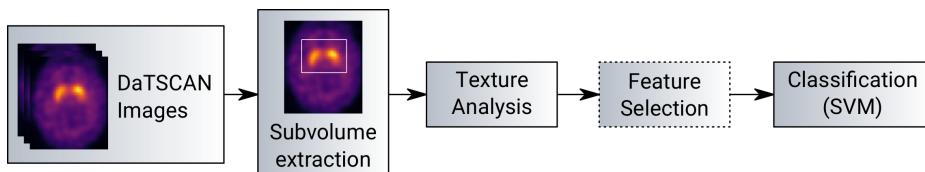


Figure 5.1: Schema of the proposed Texture-based [CAD](#) system, including an optional feature selection block.

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, 33]. First we will provide an 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 computation of the Grey Level Co-occurrence Matrix (**GLCM**) to ignore background voxels.

In this work, 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 I is higher than the threshold:

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

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

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

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

The resulting subvolume I_{sub} is the minimum box-shaped volume containing all the values for which $I > I_{th}$, 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 $I_{th} = [0.25, 0.30, 0.35]$.

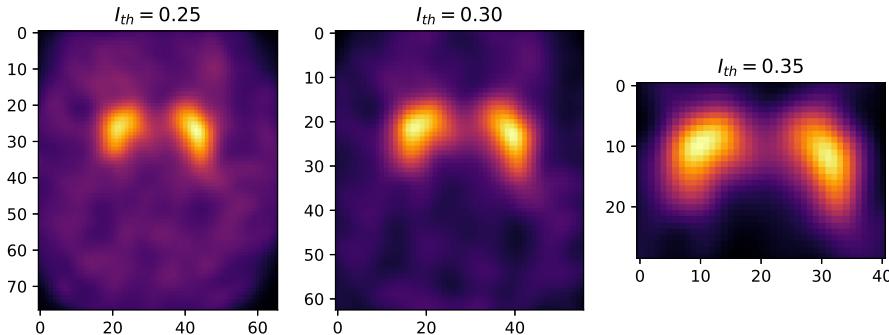


Figure 5.2: Comparison of the different I_{th} values for a random subject extracted from the PPMI database.

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](#)), which is a form of evaluating second-order texture statistics. 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 [89], that uses thirteen spatial directions to generalize the standard 2D [GLCM](#) to 3D. These offset define different angles and are used to get some degree of rotational invariance [89].

Medical images have different number precision, 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 \mathbf{C} over a $n \times m \times k$ three-dimensional image \mathbf{I} is defined as:

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

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 [164, 171], 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 feature 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, MWW U-test and the relative entropy (KL divergence) criteria (see Section 4.1).

5.2.3 Experiments

For evaluating 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 -and 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 some of these criteria.

All images used 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).

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 [\[33\]](#). We have tested the computation of the [GLCMs](#) over the image subvolumes 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 obvious differences can be found between normalization procedures. As a general trend, the integral normalization barely has an 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 no 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 as happens to the other dataset, 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 that eliminating the background is beneficial for the texture analysis and a posterior classification of the images.

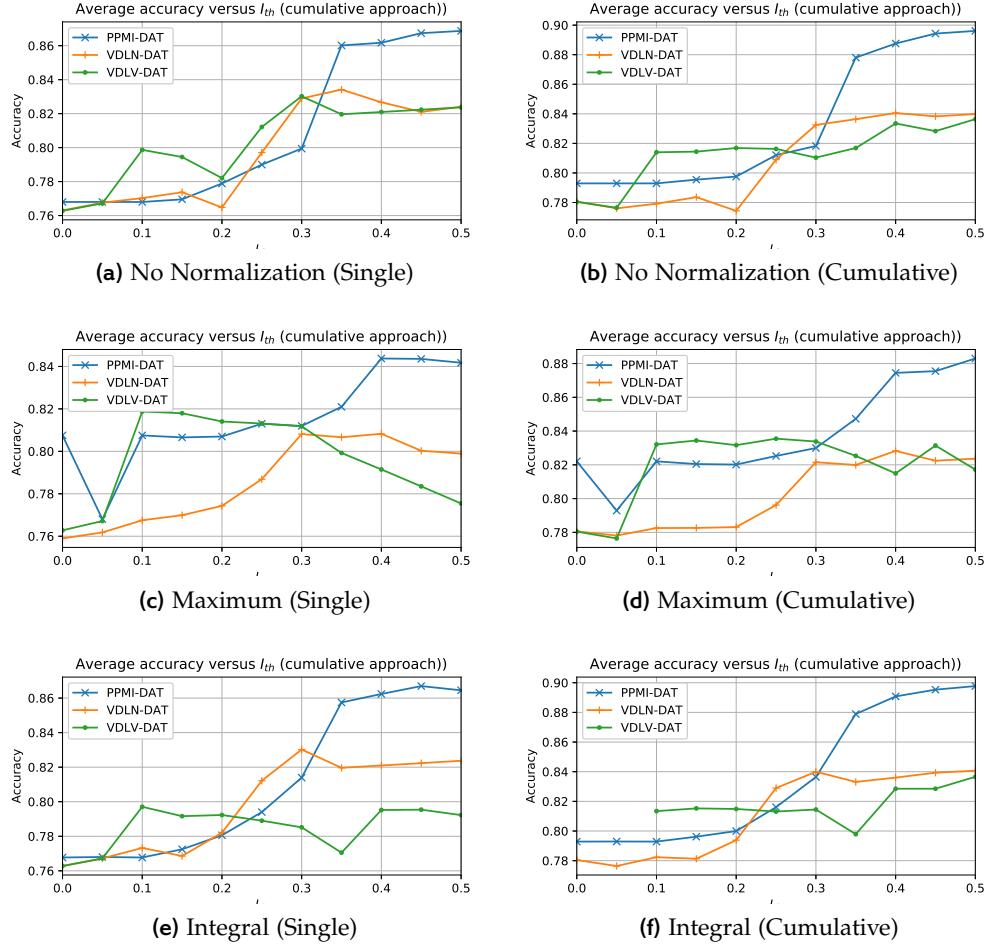


Figure 5.3: Evolution of the average accuracy values obtained for the single approach and the cumulative approach over the intensity threshold, using no normalization, normalization to the maximum and integral normalization.

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 show 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 performance data for the databases normalized 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. As commented before, that will be discussed later. Secondly, as anticipated, the best scoring feature is Cluster Tendency in most cases, and the systems perform better when the intensity threshold is more restrictive.

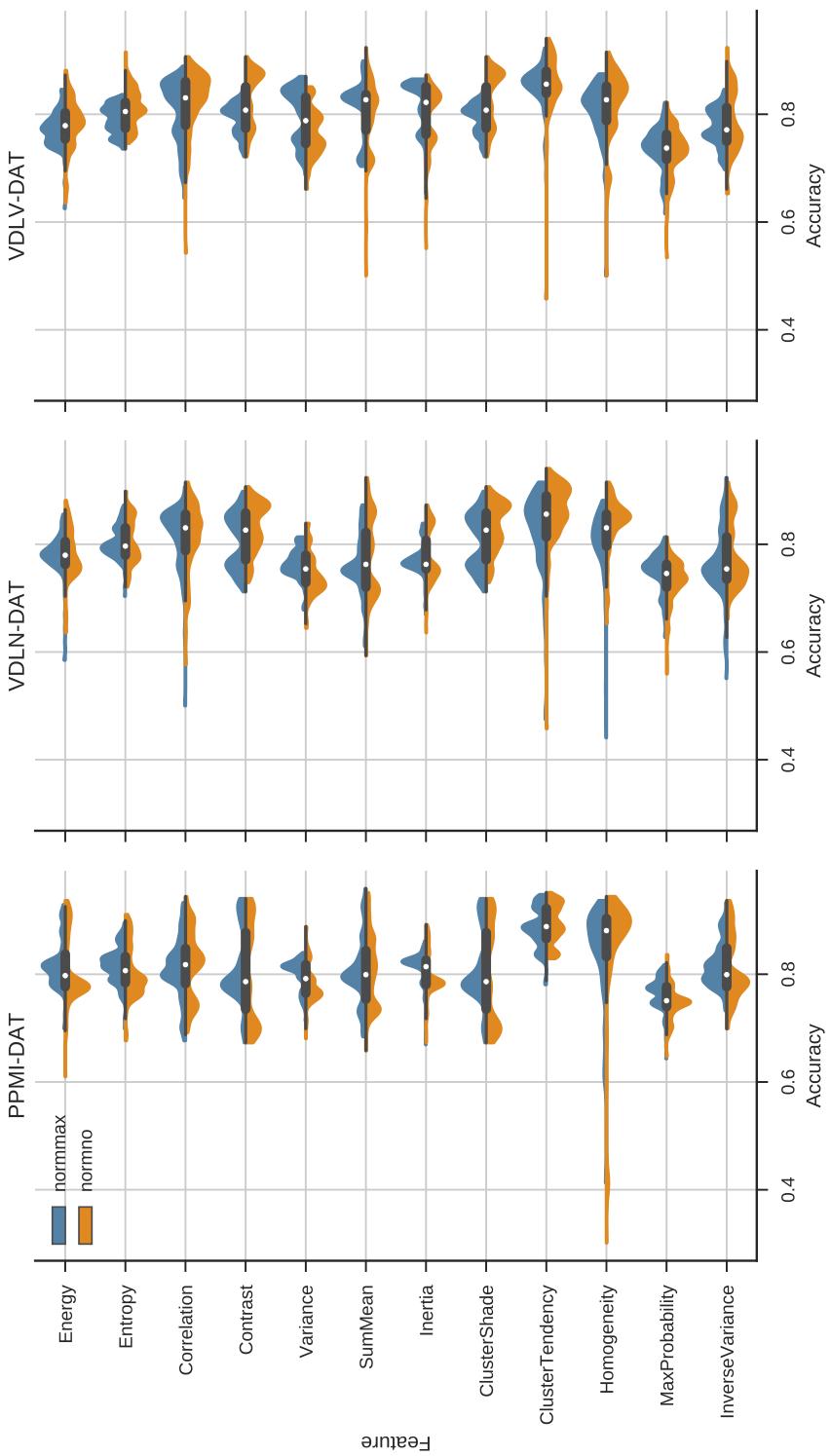


Figure 5.4: Violin plot of all accuracy values, grouped by database and showing the differences between normalization to the maximum and the original images.

	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.

When comparing the single approach with the cumulative one, it is obvious that 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 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 previous case, the integral normalization performs similarly to the original, non-normalized images. In this case, the normalization to the maximum strategy

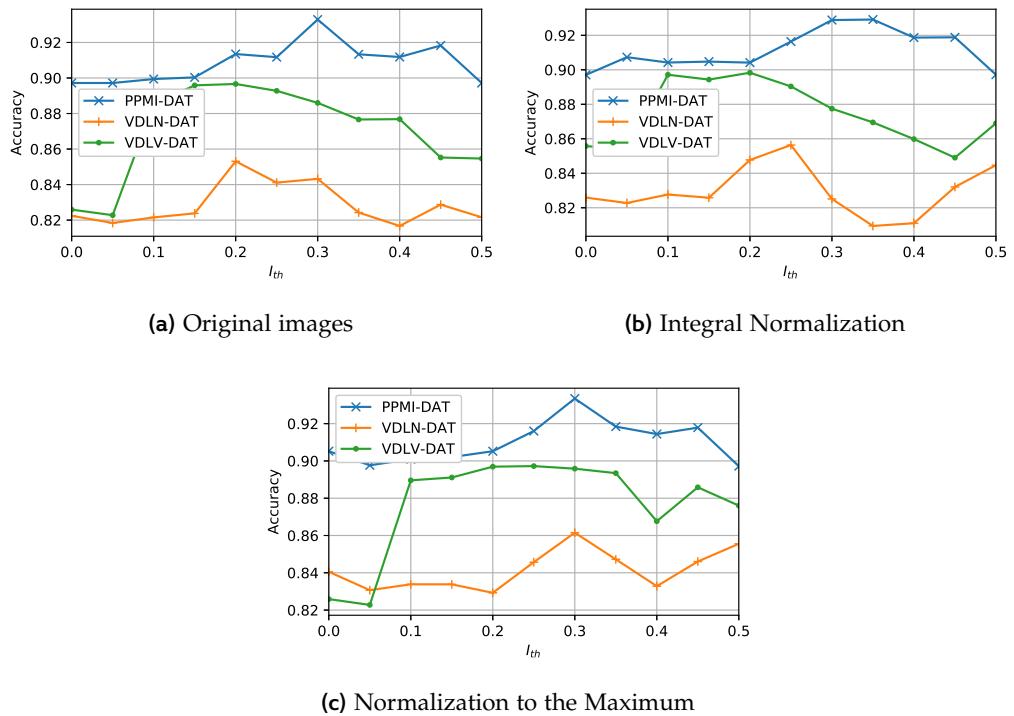


Figure 5.5: Accuracy obtained by averaging all accuracy values using a given volume selection threshold I_{th} .

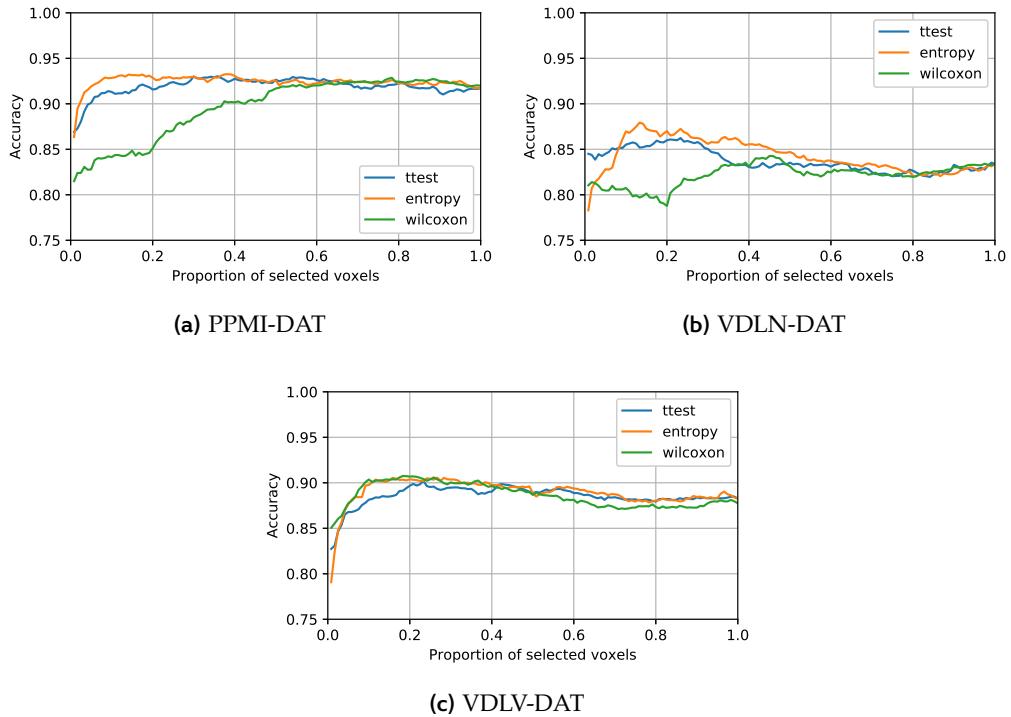


Figure 5.6: Average accuracy computed for each selection criteria, using all accuracy values for intensity thresholds of 0.10 to 0.45, compared to the proportion of selected features for the PPMI-DAT, VDLV-DAT and VDLN-DAT databases.

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 VDLN-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 methods. 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.

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 VDLN-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 obtained when using the volume extraction method proposed in Sec. 5.2.1, with values of I_{th} between 0.25 and 0.30. As for the datasets, we see that PPMI-DAT achieves 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. This was pointed in other analysis, and can be checked again at this table. 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 VDLN-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 VDLV-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 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).

	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
	Max.	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
VDLN-DAT	Orig.	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
	Int.	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
	Max.	entropy	20	0.937	0.953	0.920	0.608
		t-test	35	0.932	0.935	0.930	0.341
		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.

5.4 Discussion

The system proposed in this chapter was published in [25, 33], 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 [44], Singular Value Decomposition (SVD) [48] or PLS [37]. 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 pixel relations. When analysing the behaviour of our system in both experiments, we can infer that normalizing the images does not pose any further improvement over using the original images themselves.

The different experiments still find improvement in using normalization, but these are very small. 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 we can consider an advantage.

Now, regarding the volume selection algorithm, it has proved beneficial in almost any case. This can be due to a series 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 $2 \times 2 \times 2$ 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 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, the texture changes only correspond to real changes represented by the dopamine distribution, and no to the contrast between the brain and the background. And thirdly, computing the texture fea-

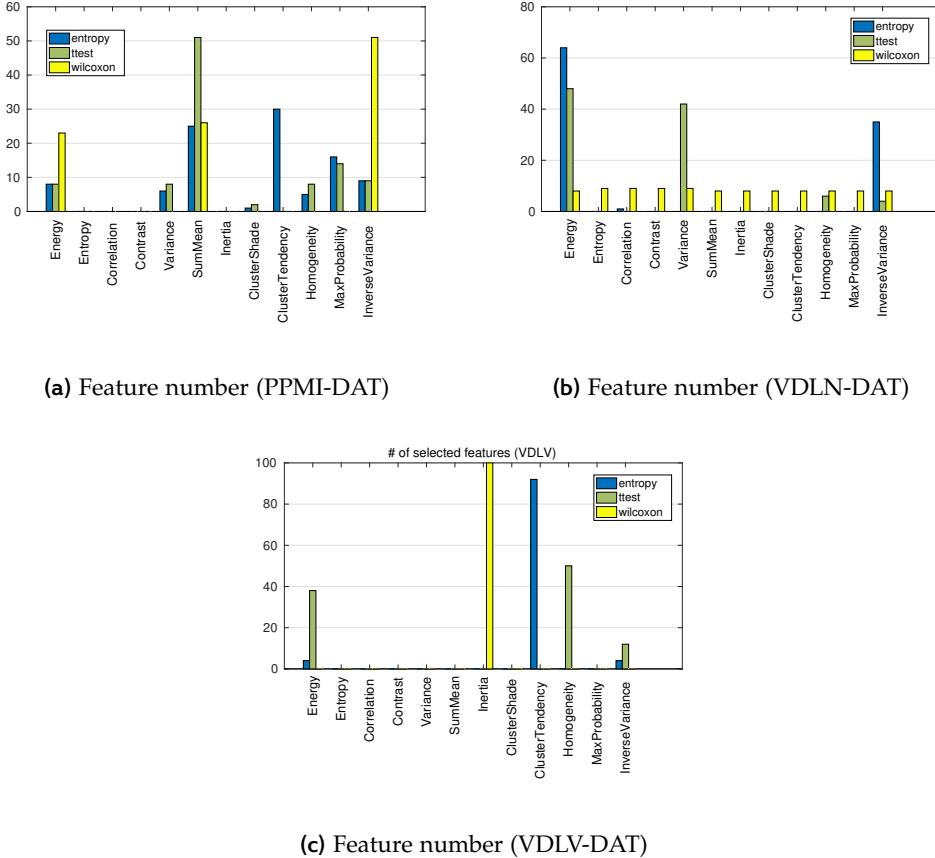


Figure 5.7: Distribution of the 100 first selected features by means of the different selection methods (using the $I_{th} = 0.35I_{max}$) for PPMI-DAT, VDLN-DAT, VDLV-DAT database.

tures 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 fundamentally 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 [44], and two additional methods. These are an asymmetrical Single Value Decomposition (SVD) [48] that applied SVD on both sides of the brain (since PD often appears only in one hemisphere), and a Empirical Mode Decomposition (EMD) [37] 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.

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 [37, 48], 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.

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	0.951	0.945	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) [44], a SVD-based approach [48] and EMD using the third independent mode function (IMF3) [37].

6

SPHERICAL BRAIN MAPPING

6.1 Introduction

In this section we will present a 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 even project the information to a bidimensional maps, 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].

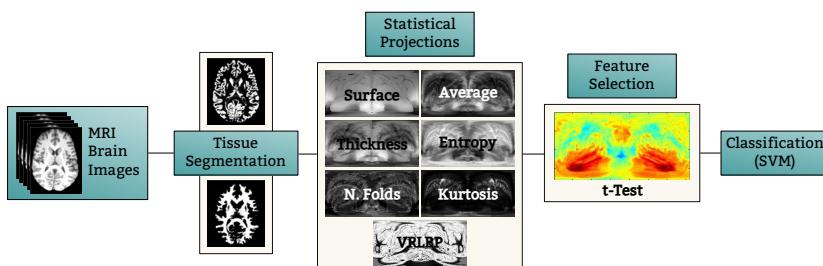


Figure 6.1: Flow diagram of the procedure used in the textural analysis of projected MR brain images.

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 creation

algorithm that follows a minimum-intensity change path 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 along the path. This was extended to make use of the GLCM and Haralick texture analysis (see Chapter 5) to characterize the brain texture along 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 P voxels crossed by the mapping vector $\mathbf{v}_{\theta,\varphi}$.

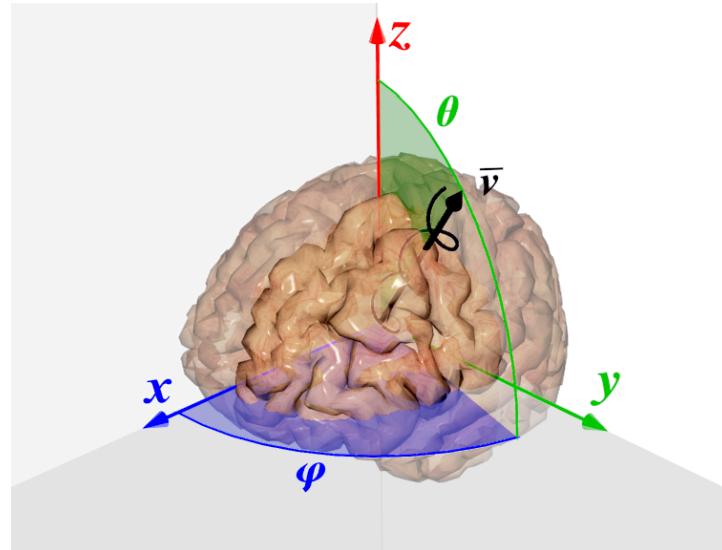


Figure 6.2: Illustration of the computation of the mapping vector $\mathbf{v}_{\theta,\varphi}$, the angles θ and φ and the r -neighbourhood of \mathbf{v} (see Section 6.2.2).

The basic form in which SBM works is by computing 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 is intended to characterize 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 [141], 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 $V_{\theta,\varphi}$:

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

- 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**, which 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 whole algorithm that produces these maps can be downloaded at <http://pakitochus.github.io/mapBrain/>.

This methodology defines the sampling set as the voxels that are crossed by the sampling vector $\mathbf{v}_{\theta,\varphi}$. This implies a loss of information on the neighbourhood of $\mathbf{v}_{\theta,\varphi}$ that increases with the distance to the origin. To overcome this problem, two different approaches 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 $\mathbf{v}_{\theta,\varphi}$ 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 addition that can be made to the original **SBM** is the inclusion of the r -neighbourhood of the mapping vector $\mathbf{v}_{\theta,\varphi}$ in the computation of v . We do so

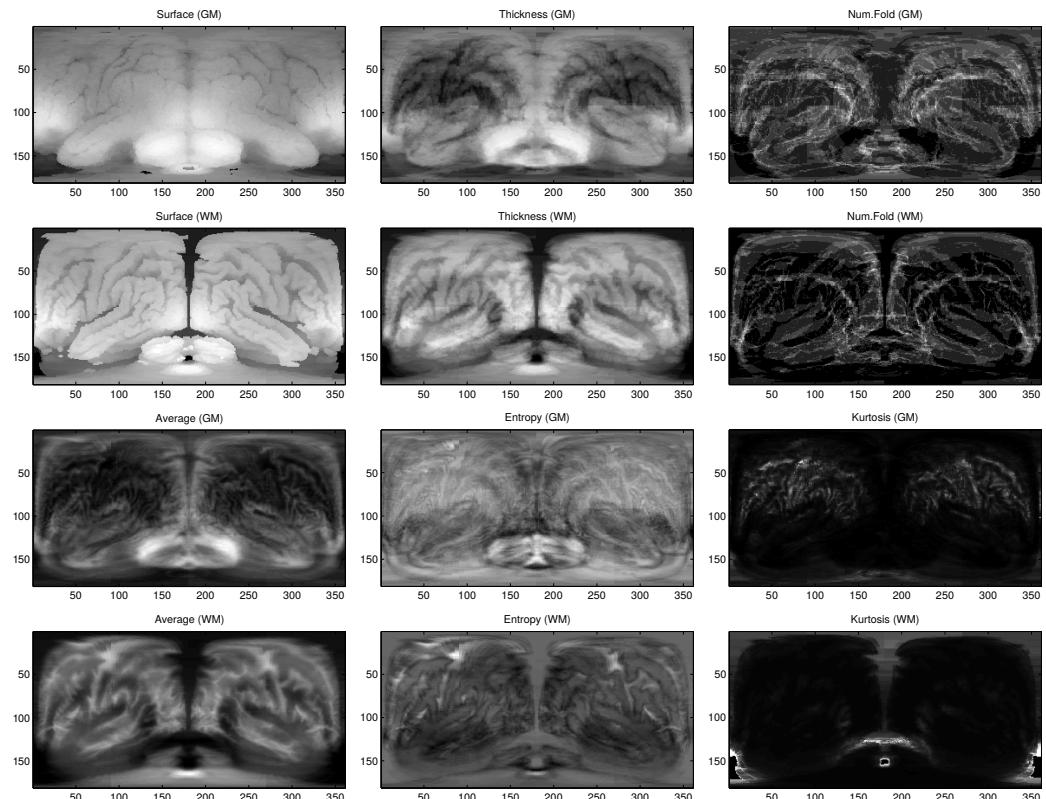


Figure 6.3: Resulting GM and WM maps of the same control subject using the six proposed measures: Surface, Thickness, Number of Folds, Average, Entropy and Kurtosis.

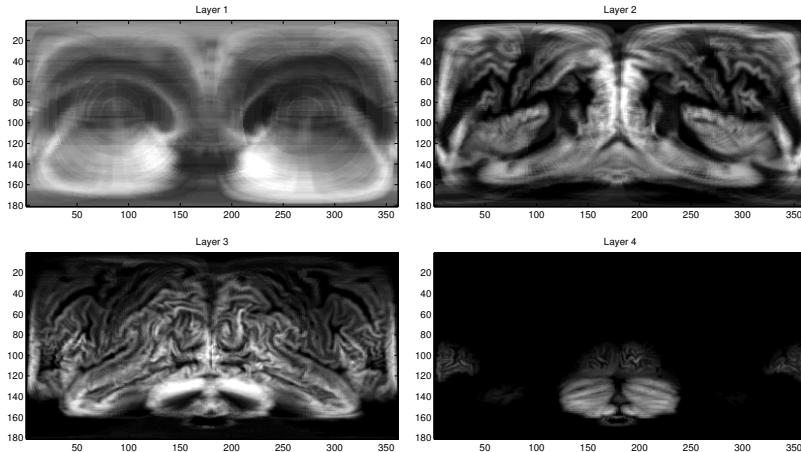


Figure 6.4: Example of the layered approach using the average measure on GM maps. Some internal GM structures such as the Putamen or Globus Pallidus can be identified at Layer 2 (see anatomical reference at Figures 6.7 and 6.8).

by computing the Volumetric Radial LBP (VRLBP), based on the LBP descriptors proposed in [154].

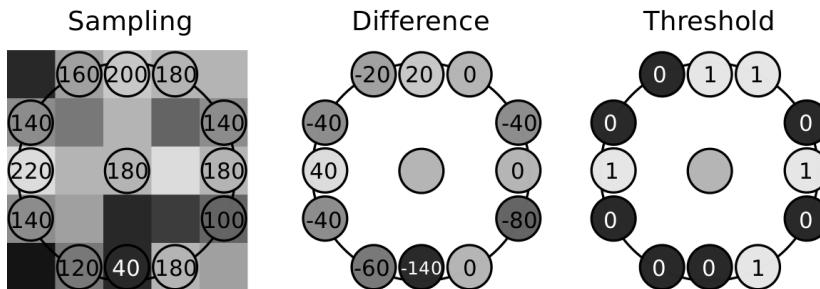


Figure 6.5: Example of how the basic LBP is computed.

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 pth voxel and the central voxel for which 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 [102], 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 will update the sampling procedure proposed in [102] 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:

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

being n 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 [102].

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 similar to the definition of SBM found in Section 6.2. The value v of the VRLBP approach is therefore defined as:

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

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 [130] 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.

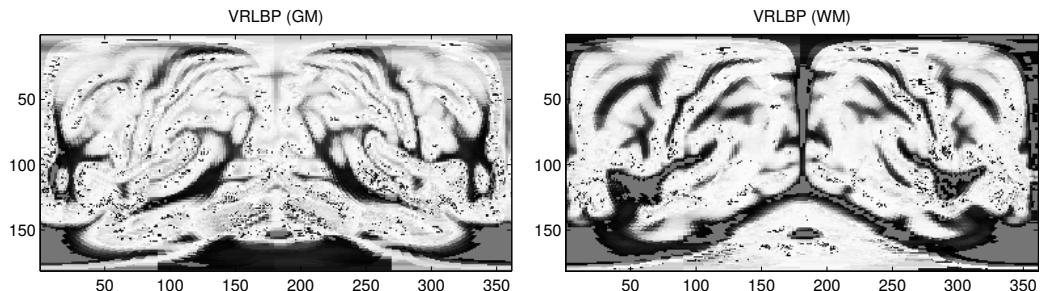


Figure 6.6: An example of the VRLBP projection for GM and WM Tissues.

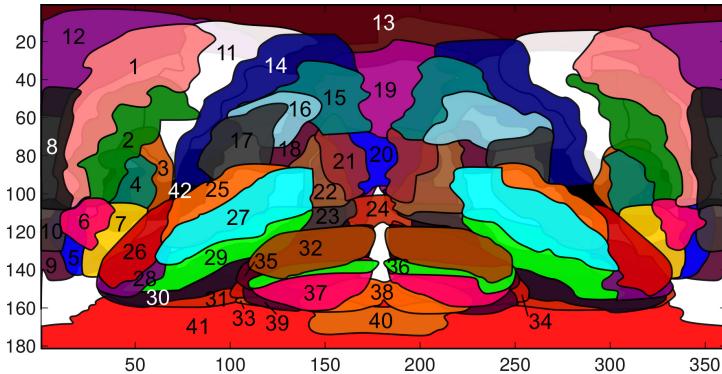


Figure 6.7: SBM mapping of different cortical regions. In the Frontal region, we can find: 1) Frontal Sup., 2) Frontal Mid., 3) Frontal Inf. Oper., 4) Frontal Inf. Tri., 5) Frontal Sup. Orb, 6) Frontal Mid. Orb, 7) Frontal Inf. Orb, 8) Frontal Sup. Medial, 9) Rectus, 10) Frontal Med. Orb., 11) Precentral, 12) Supp. Motor Area. In the Parietal region: 13) Paracentral Lobe, 14) Postcentral, 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.

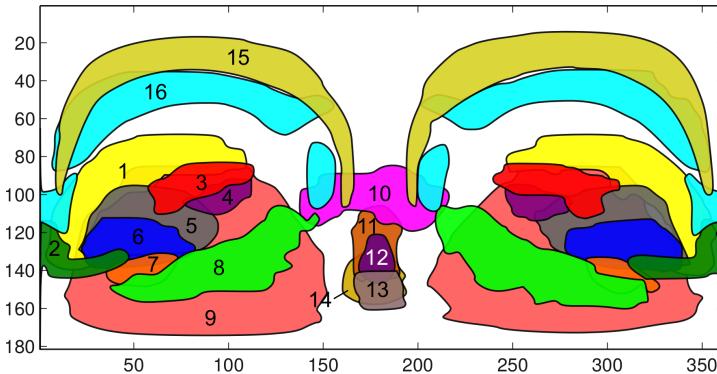


Figure 6.8: SBM mapping of some important subcortical regions and organs. We observe the following subcortical structures: 1) Caudate Nucleus, 2) Olfactory Bulb, 3) Rolandic Operculum, 4) Heschl's gyri, 5) Putamen, 6) Globus Pallidus, 7) Amygdala, 8) Hippocampus, 9) Thalamus, 10) Lingual, 11) Vermis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

6.3 Sampling Paths via Hidden Markov Model (HMM)

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 the technique used to define minimum intensity change sampling paths via Hidden Markov Models, that was 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 that contains 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 VBM 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 [81]:

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

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:

$$P(\mathbb{P}_{\theta,\varphi}) = P(p_0, p_1, p_2, \dots, p_N) \quad (6.15)$$

with p_0 being the origin of the spherical coordinates, and p_N is the last possible coordinate within \mathbb{I} in the current direction (θ, φ) . In this work, we have placed 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 [162], 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:

$$P(p_i | p_{i-1}, p_{i-2}, \dots, p_0) \approx P(p_i | p_{i-1}) \quad (6.16)$$

Using this assumption, Eq. 6.15 becomes:

$$P(\mathbb{P}_{\theta,\varphi}) = P(p_0, p_1, \dots, p_N) = \prod_{i=1}^N P(p_i | p_{i-1}) \quad (6.17)$$

Using this HMM definition, the hidden state of each node will be its intensity $I(p_i)$. Similarly to the original SBM, let $V_{\theta,\varphi} = \{I(p_0), I(p_1), \dots, I(p_N)\}$ 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:

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

$$P(\mathbb{P}_{\theta,\varphi} | I) = P(p_0, \dots, p_N | I(p_0), \dots, I(p_N)) \quad (6.19)$$

$$= \frac{P(I(p_0), \dots, I(p_N) | p_0, \dots, p_N) \cdot P(p_0, \dots, p_N)}{P(I(p_0), \dots, I(p_N))} \quad (6.20)$$

where:

$$P(I(p_0), \dots, I(p_N) | p_0, \dots, p_N) = \prod_{i=1}^N P(I(p_i) | p_i) \quad (6.21)$$

and $P(I(p_0), \dots, I(p_N))$ 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 computation 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:

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

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 defining the last term $P(\mathbf{p}_0, \dots, \mathbf{p}_N)$ in Eq. 6.20, setting an attractor located in the position \mathbf{p}_N , the last possible coordinate within \mathbb{I} in the current direction (θ, φ) . It should affect the transition probability between states 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 gaussian 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} not only contains 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 point. However, what we are really interested on 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 [158]. If we note $G_m \equiv \{I_0, I_1, \dots, I_m\}$ the set containing all intensity levels in the image \mathbf{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 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° .

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 $I_{th} = I_s$ where s maximizes $TE(s)$;

Set p_0 to the AC;

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

$p_i \leftarrow p_0$;

while ($i < IterLimit$) $\&$ ($I(p_i) > I_{th}$) $\&$ ($p_i \in \mathbb{I}$) **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 $P(I(p_c)|p_i)$;

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

$i = i + 1$;

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

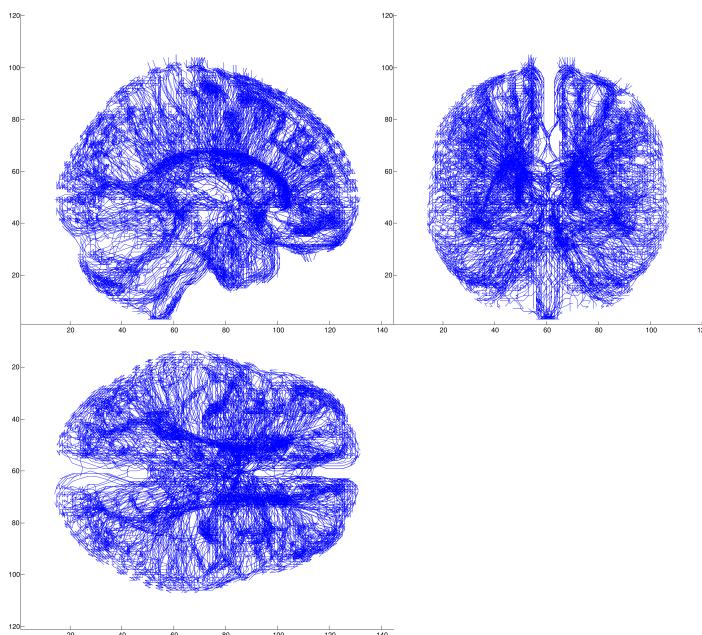


Figure 6.9: Set of HMM based paths over the MRI DARTEL template.

6.3.1 Radial Texture Features

The paths $\mathbb{P}_{\theta,\varphi}$ proposed above could theoretically 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 $\mathbb{P}_{\theta,\varphi}$.

In [14] a helical sampling was proposed to define the neighbourhood of the mapping vector $v_{\theta,\varphi}$. 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 p_i and p_{i+1} , is stored along the whole path $\mathbb{P}_{\theta,\varphi} = \{p_0, p_1 \dots p_N\}$. Mathematically, the computation of the GLCM in each point in the path will be:

$$C_{\Delta_n}(i, j) = \sum_{n=0}^{N-1} \begin{cases} 1 & I(p_n) = i, I(p_{n+1}) = j \\ 0 & \text{otherwise} \end{cases} \quad (6.30)$$

where the offset is different for each pair of nodes $\Delta_i = p_{i+1} - p_i$.

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 p_i in the [HMM](#) path, which we have noted as the set X_i . Under this generalization, equation 6.30 becomes:

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

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 [171] as well as in [144] and [126]. These are Dissimilarity[144] (eq. 6.32), Difference Variance[171] (D. Variance, eq 6.33), Difference Entropy[171] (D. Entropy, eq 6.34), Inverse Difference Normalized[126] (IDN, eq 6.35) and Inverse Difference Moment Normalized[126] (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 six 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 use 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 structures that are mapped to these directions, including mid temporal lobe, amygdala,

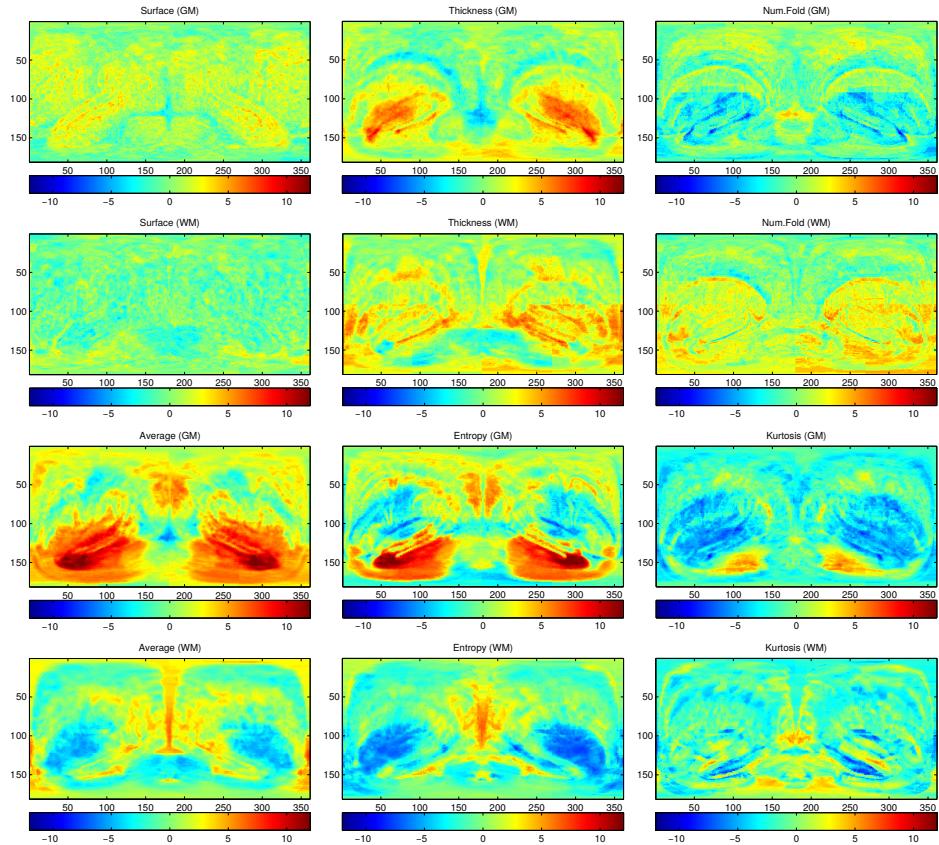


Figure 6.10: t-maps to assess the regions with a higher statistical relevance in the AD vs CTL paradigm, for each SBM measure and using GM and WM maps.

dala, hippocampus or parahippocampal lobe, considered a strong indicator in the NINCDS-ADRDA criteria [95]. 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) [36].

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 [132].

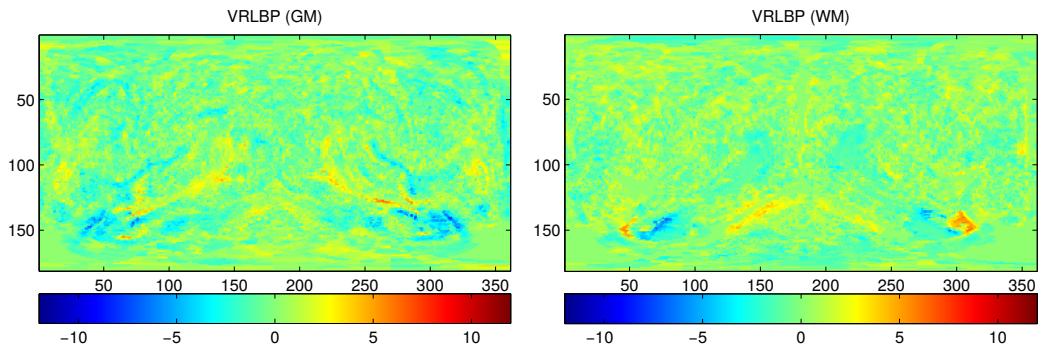


Figure 6.11: Maps that present the level of the t statistic in the AD vs. CTL paradigm, for the VRLBP projections mapping over a) GM and b) WM.

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 statistical measures (average, entropy and kurtosis) clearly outperform the morphological ones (surface, thickness and number of folds), although the tissue thickness is the best performing 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),

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.

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 these 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.

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 could 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.

Performing a significance analysis on the 4 layer extension of the 7 proposed **SBM** maps over the **GM** and **WM** images could be infeasible for this thesis, 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.

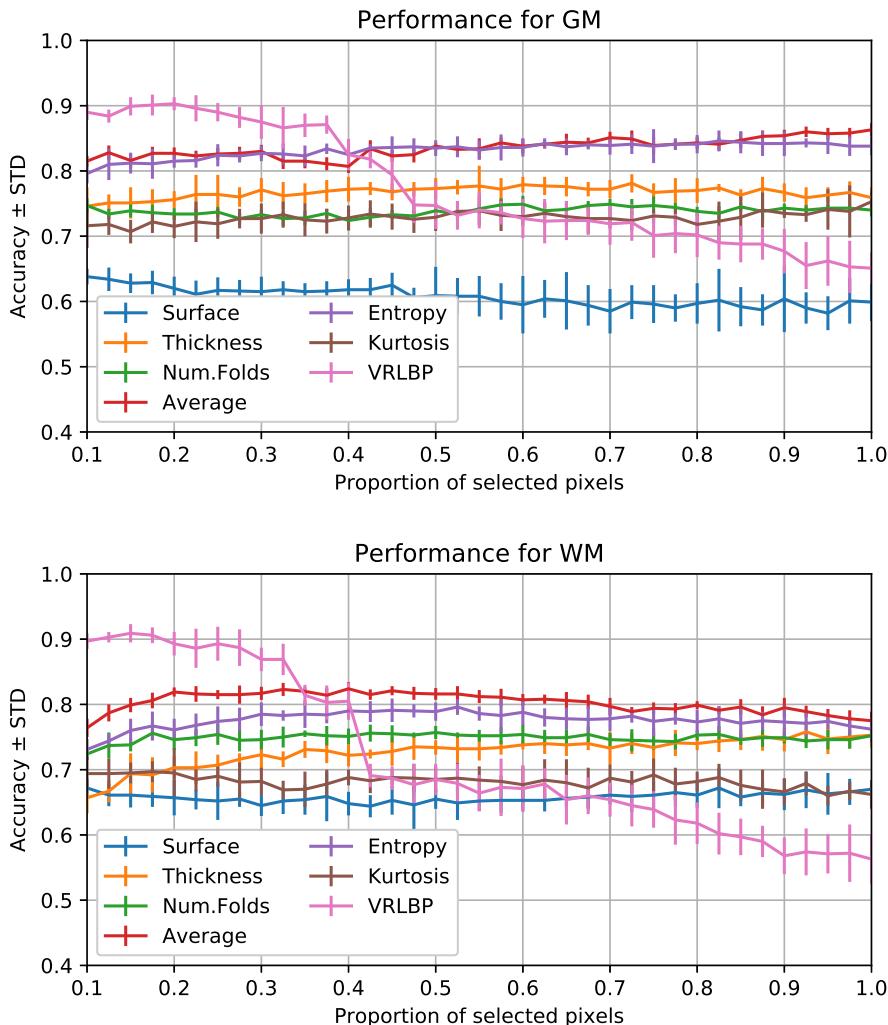


Figure 6.12: Performance for the different SBM approaches over the: a) Grey Matter and b) White Matter.

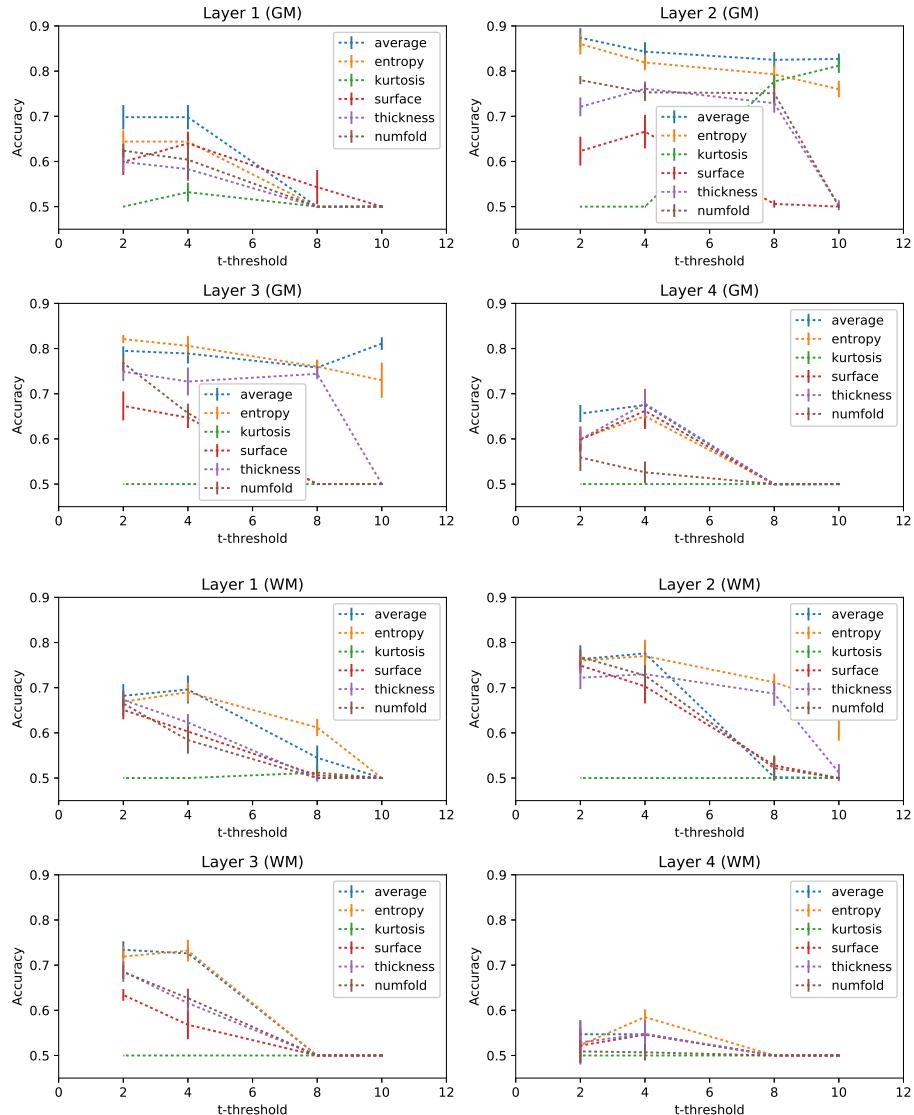


Figure 6.13: Performance for the different four-layered mappings over the: a) Grey Matter and b) White Matter at different levels of statistical significance.

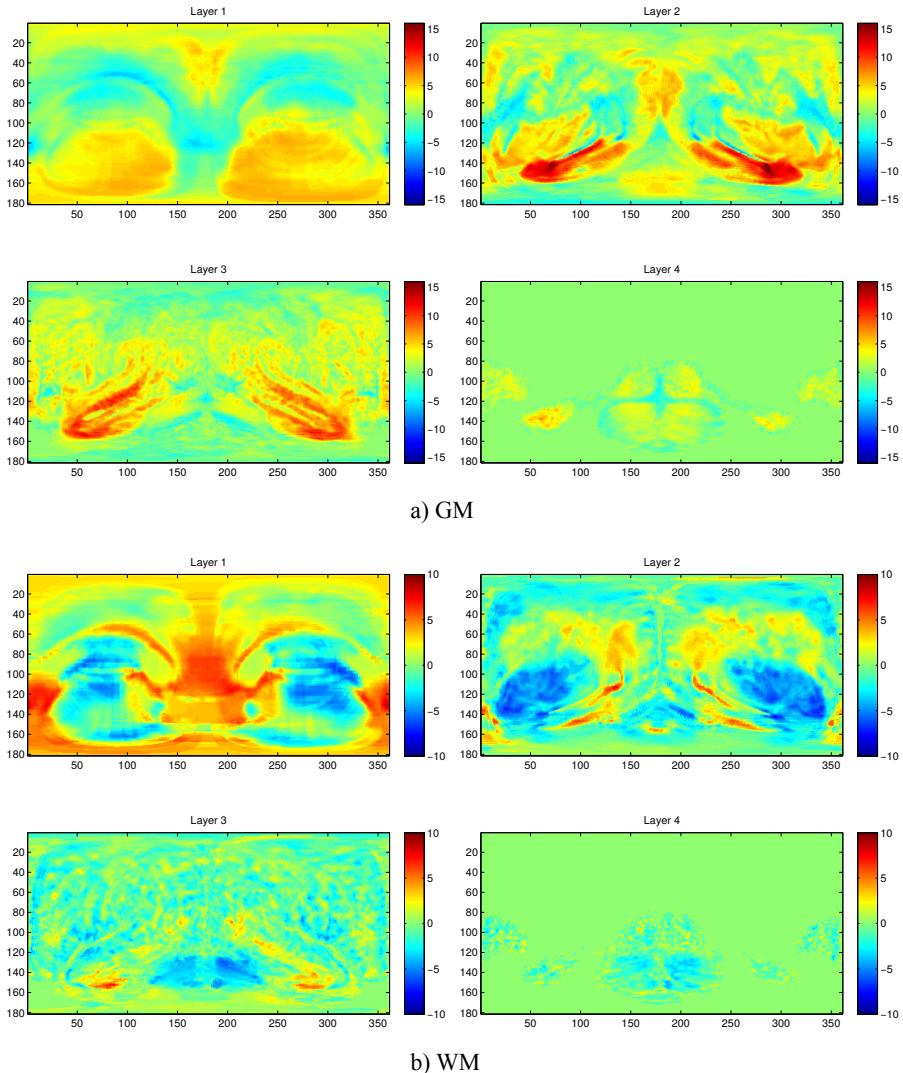


Figure 6.14: t-maps that present the level of statistical relevance in the [AD](#) vs [CTL](#) paradigm, for a four-layered average mapping over a) [GM](#) and b) [WM](#).

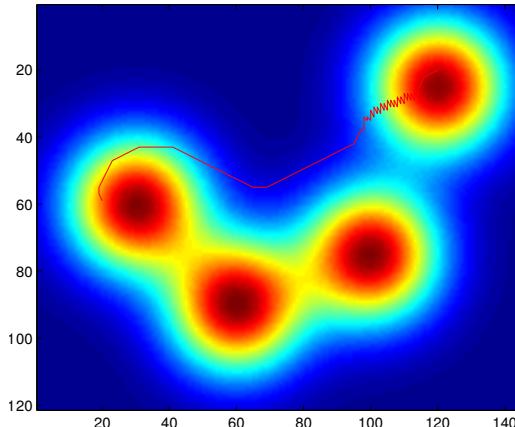


Figure 6.15: Path traced over a gaussian mixture distribution of 4 isotropic gaussian kernels.

In [GM](#), the most obvious changes are located in layers 2 and 3, specifically at the hippocampus, parahippocampal lobe 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 [36, 95] 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 [132].

6.5.3 Experiment 3: [HMM](#) on Synthetic Datasets

Now, we will analyse the behaviour of the [HMM](#) paths on some examples. In Figure 6.15, we used a Gaussian mixture probability density function, with four isotropic Gaussian kernel, 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

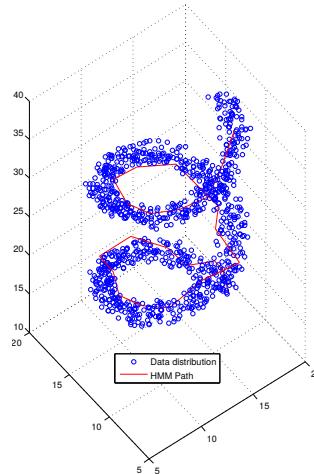


Figure 6.16: HMM path computed inside a density distribution defined by an helix.

approximation to \mathbf{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 \mathbf{p}_0 to the point on the helix distribution with smallest z coordinate, and \mathbf{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 was acquired by the LANDSAT SRTM30+ mission, and its intensity represents height. We tested multiple paths by establishing sequentially \mathbf{p}_0 and \mathbf{p}_N in ten different cities. The paths shown at Figure 6.17 optimize both the distance and height variation as well as resembling -in most cases- the 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 $r = 30$.

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

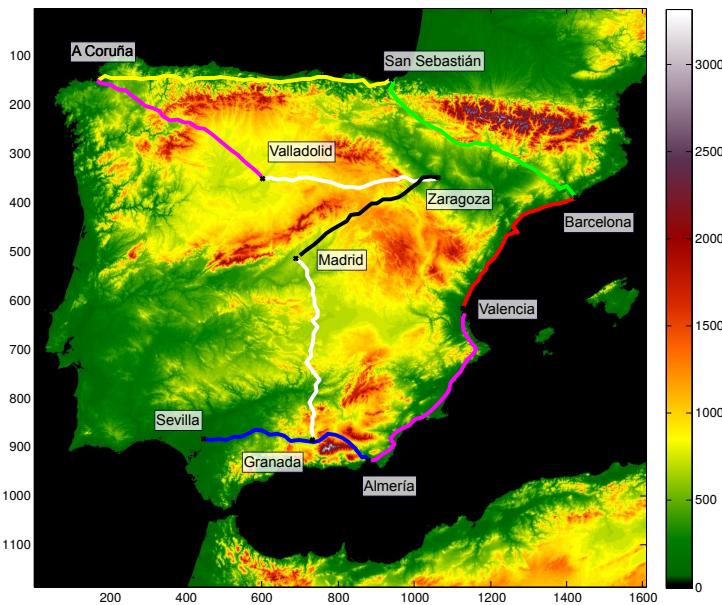


Figure 6.17: Simulation of the [HMM](#)-based path tracing over an Iberian Peninsula height map, interconecting different cities.

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 will be 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.

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, under the AD vs NOR scenario (76.8% for [GM](#) and 0.64% for [WM](#)) XXXXMODI-

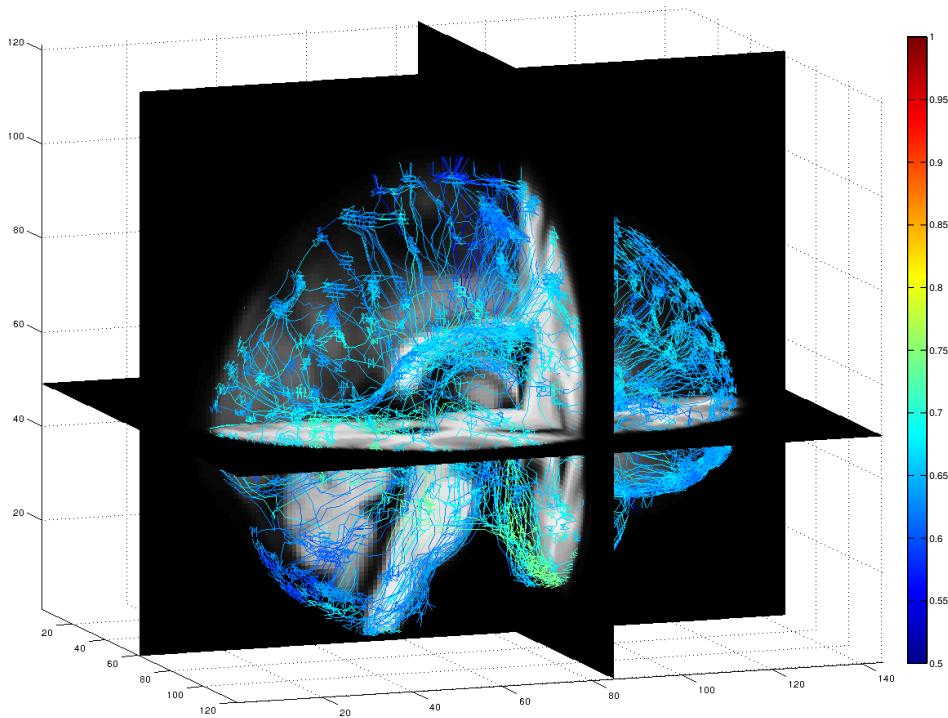


Figure 6.18: canonical paths computed in each direction (θ, φ) . Each path's colour represent the accuracy in a differential diagnosis. Only one in every five paths are shown for clarity purposes.

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).

FICAR, 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 term, which proved to be a good approach. 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

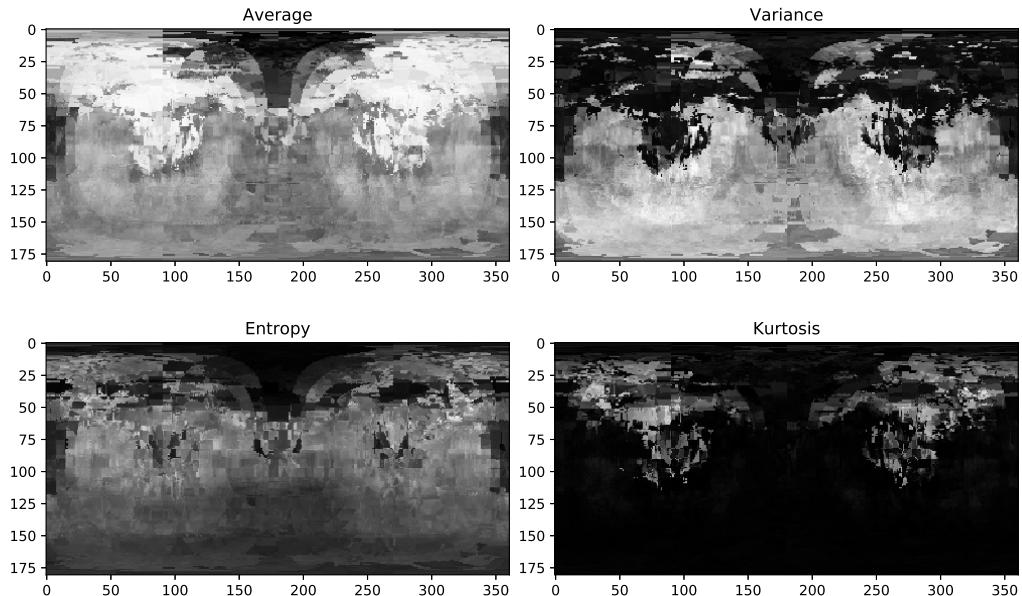


Figure 6.19

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.

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.

per texture feature and direction, which allow us to construct texture **SBM** maps. In Table ?? we show the performance values for the maps built using each of the texture features proposed in Section 6.3.1

Difference variance achieves here the higher accuracy, 0.736 ± 0.070 , a value similar to 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 Spherical Brain Mapping 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 Alzheimer’s Disease, these occur mainly in **GM** [36, 73, 107, 115, 119, 132], 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 neg-

ative in the areas where we found positive values in [GM](#) maps. This can be also appreciated in the layered approach, at Figure [6.14](#).

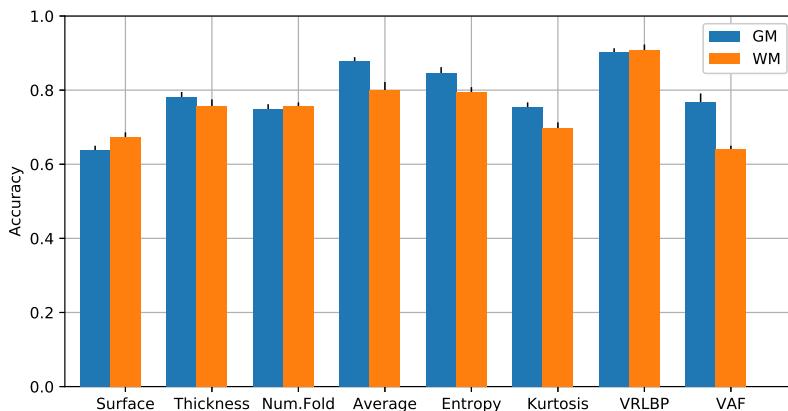


Figure 6.20: Performance at the operation point for the different mappings over the Grey Matter and White Matter, compared with the performance of [VAF](#).

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 $v_{\theta,\varphi}$. We classified all the [SBM](#) measures in three groups: 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 [\[115, 141\]](#), 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 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 lobe or some structures of the basal ganglia, such as caudate nucleus, globus pallidus or putamen [36, 95].

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 minimum 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** [54, 88]. **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

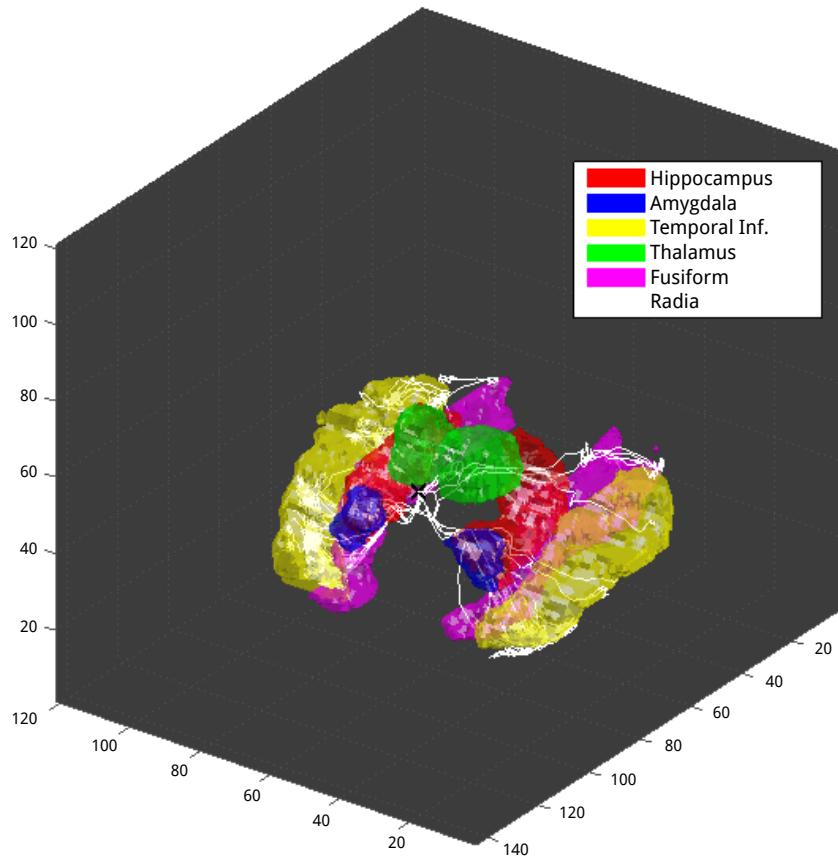


Figure 6.21: Paths that obtain more than 75% accuracy, and a three-dimensional representation of the structures crossed by them.

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 [130].

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 [84, 95, 132], as well as its surrounding structures (Amygdala, Parahippocampal lobe and Fusiform) [132]. Some studies have also reported neurodegeneration at the Thalamus and Putamen, also related to early AD [84], and in advanced AD, atrophy in most of the neocortex [84, 95, 132]. 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 intensity changes are small and progressive, and therefore, higher order statistics that focus on this (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 General Discussion of SBM

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 [119] 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

App.	Feature	Tis.	Accuracy	Sensitivity	Specificity
VAF	Intensity	T ₁	0.768 ± 0.011	0.752 ± 0.016	0.785 ± 0.016
	Intensity	GM	0.768 ± 0.011	0.752 ± 0.016	0.785 ± 0.016
	Intensity	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
	average	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
	VRLBP	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
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*
SCA	-	WM	0.808 ± 0.0*	0.817 ± 0.0*	0.800 ± 0.0*

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

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

paths and selected features from the paths), the hybrid **HMM-SBM** tool (variance and kurtosis) and the best scoring texture feature (difference variance).

Then, 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) [35] 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 segmentation 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 $V_{\varphi,\theta}$, 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 samples 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 [91, 108], 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 [41, 105], 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 [78], 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 [78]. Furthermore, results in [74] 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 [91, 108]. Furthermore, there exist a number of diffeomorphic registration algorithms, such as DARTEL [93] or ANTS [60], 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 [82] that 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, in this case, undesired intensity variance introduced by multi-site image acquisition. **SWPCA** takes any dataset of pre-processed images, spatially normalized, and decomposes them into their variance components to then provide a corrected dataset where these undesired variance components have been reduced. To do so, **PCA** was 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** estimated 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 was then identified by its corresponding *p*-value. Finally, these *p*-values were transformed into a weighting matrix Λ that weighted the influence of each variance component in a final **PCA** reconstruction of the corrected maps. The procedure is summarized in Figure 7.1.

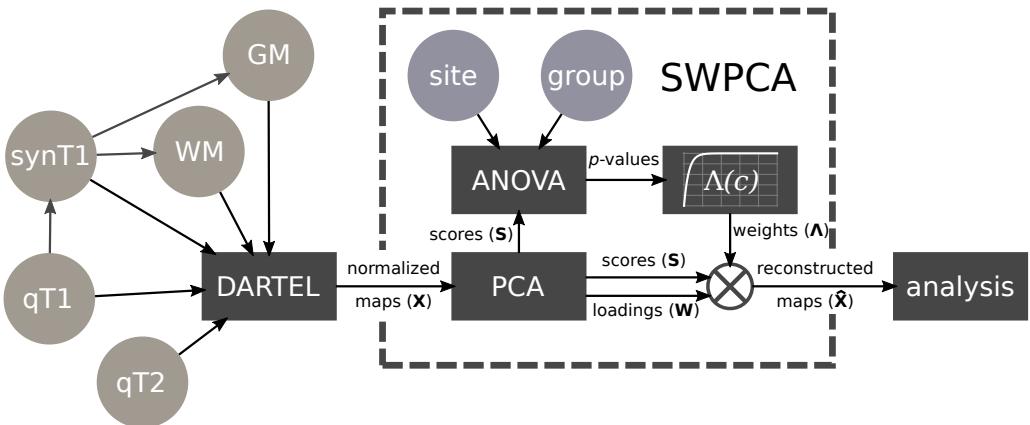


Figure 7.1: Summary of the [SWPCA](#) algorithm, along with its context in the pipeline used in this article. Circles represent the input data, both images (green shading) and class (group and acquisition site, purple shading). Rectangles represent the different procedures applied, comprising the DARTEL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

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 $K \times N$ matrix, with K participants (in this case, with one image per participant) and N the number of voxels, into a set of N linearly uncorrelated variables called Principal Components (PC, also known as component loadings or the mixing matrix) \mathbf{W} of size $N \times N$ whose linear combination using a vector of component scores \mathbf{s}_K 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 $K \times N$) was estimated as:

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

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 eigenvalue decomposition 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 eigenvalue decomposition 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 Eigenvalue Decomposition of $\mathbf{X}^T \mathbf{X}$. Since the eigenvalue decomposition 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 $K \times K$ orthogonal matrix, Σ is a $K \times N$ diagonal matrix with non-negative real numbers on the diagonal, and the $N \times N$ unitary matrix \mathbf{V}^* denotes the conjugate transpose of the $M \times M$ 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 $\mathbf{S} = \mathbf{X} \mathbf{W} = \mathbf{U} \Sigma$.

We obtained both the component scores and estimates of the ‘eigenbrains’ \mathbf{W} using this decomposition. 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:

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

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

images, $C = K$. 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,

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

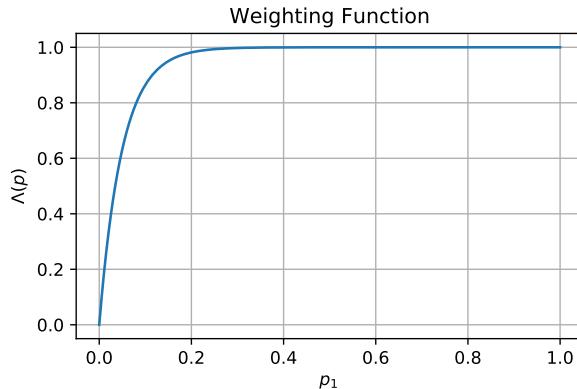


Figure 7.2: Weighting function $\Lambda_c(p_c, p_{th})$ used in SWPCA.

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) have been parsed out while keeping all other sources of variance.

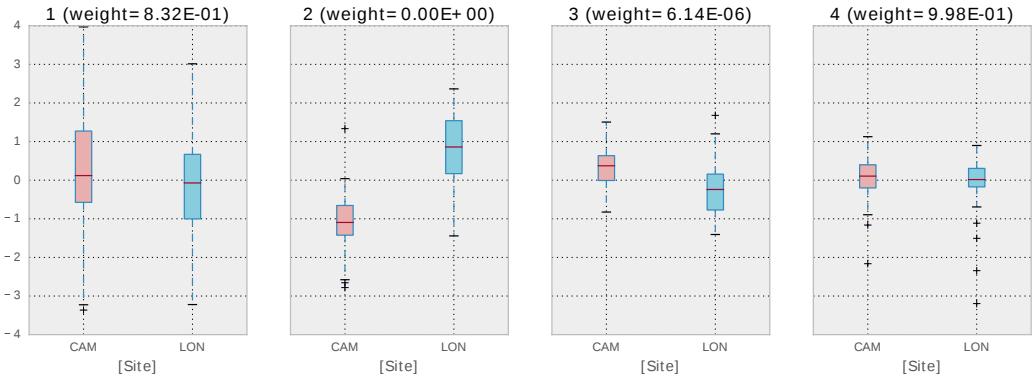


Figure 7.3: Box-plot of the distribution of the component scores at each site of the AIMS-MRI dataset (see Sections 7.3 and A.1.2) in the four first components. We assume that bigger differences between distributions imply a bigger influence of the acquisition site on the portion of variance modelled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

7.2 Evaluation of SWPCA

To validate the effects of the SWPCA algorithm on the inter-site variance, experiments were undertaken to assess the reduction of the undesired site variance in the original datasets, and its impact on the between-group signal. Two kind of analysis were performed: a characterization of voxel-wise differences, and a classification analysis.

Voxel-wise differences between groups were characterized using Voxel Based Morphometry (VBM) [136], 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 computation of the test. 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 [168]. Additionally, a Component Based Morphometry (CBM), based on Source Based Morphometry (SBM) [80] 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 [156].

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**, were 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 (**qT₁**, **qT₂**, simulated T₁ - weighted Inversion Recovery (**synT₁**), **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 **qT₁** and **synT₁** maps, where high variability between acquisition sites, especially in the right hemisphere, was substantially reduced after the application of **SWPCA**. The reduction in the **qT₂**, **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 low. 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. This is an estimation of the discrimination ability of the different modalities without the influence of the site effects; Table 7.2. 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 in 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

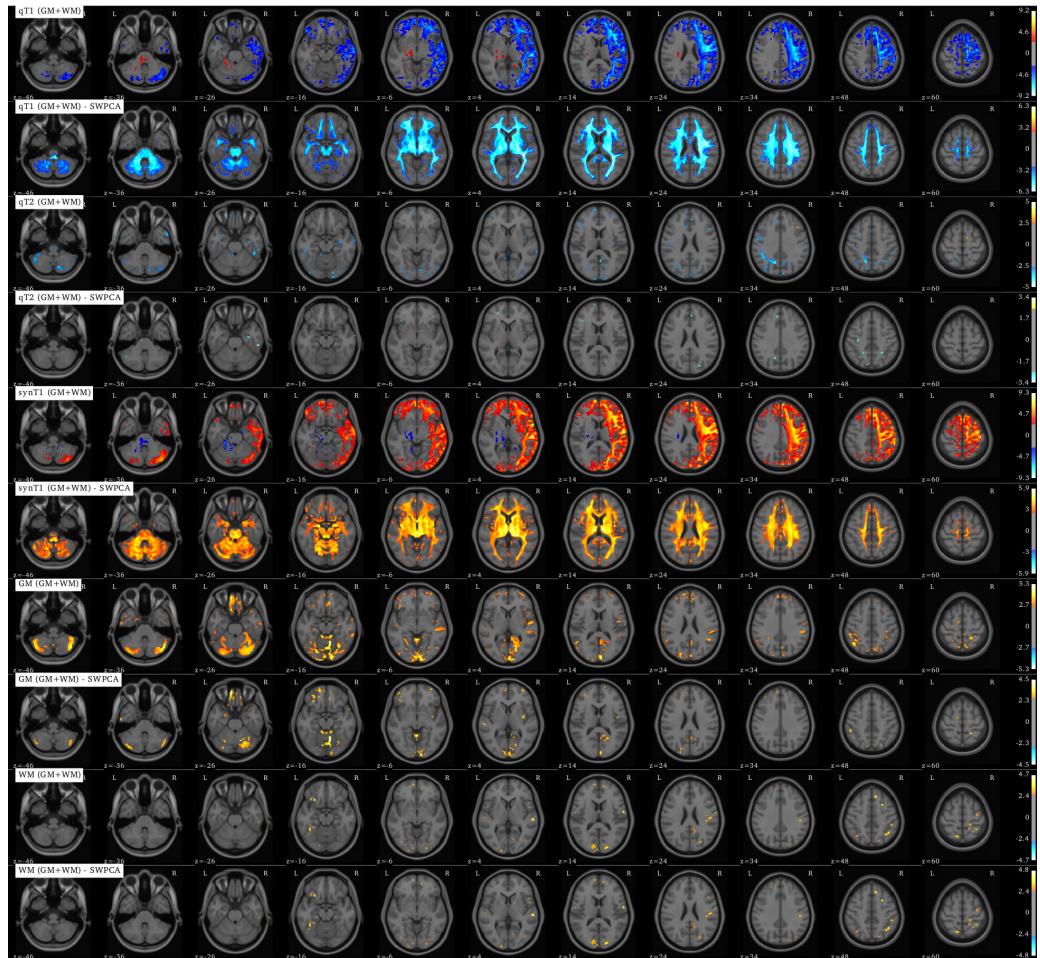


Figure 7.4: Brain t-map (VBM) of significant ($p < 0.01, |t| > 2.57$) GM and WM between-group differences using qT₁, qT₂, synT₁, GM and WM modalities after applying SWPCA to remove site effects.

Modality	Mask	ALL			CTL			ASD		
		no- <i>SWPCA</i>	<i>SWPCA</i>	no- <i>SWPCA</i>	<i>SWPCA</i>	no- <i>SWPCA</i>	<i>SWPCA</i>	no- <i>SWPCA</i>	<i>SWPCA</i>	
<i>qT₁</i>	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			
<i>qT₂</i>	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			
<i>synTr₁</i>	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 \pm 0.175	0.512 \pm 0.260	0.692 \pm 0.237	0.504 \pm 0.193	0.492 \pm 0.276	0.515 \pm 0.307
	GM	0.501 \pm 0.157	0.440 \pm 0.244	0.565 \pm 0.245	0.484 \pm 0.201	0.488 \pm 0.300	0.480 \pm 0.327
	WM	0.505 \pm 0.174	0.485 \pm 0.248	0.526 \pm 0.242	0.451 \pm 0.197	0.465 \pm 0.297	0.435 \pm 0.296
qT_2	GM+WM	0.628 \pm 0.168	0.535 \pm 0.246	0.719 \pm 0.237	0.467 \pm 0.181	0.527 \pm 0.307	0.417 \pm 0.314
	GM	0.539 \pm 0.149	0.425 \pm 0.220	0.654 \pm 0.222	0.491 \pm 0.196	0.548 \pm 0.316	0.430 \pm 0.298
	WM	0.619 \pm 0.194	0.585 \pm 0.262	0.655 \pm 0.250	0.472 \pm 0.195	0.448 \pm 0.283	0.492 \pm 0.290
synt $_1$	GM+WM	0.665 \pm 0.158	0.578 \pm 0.224	0.755 \pm 0.238	0.479 \pm 0.201	0.478 \pm 0.318	0.475 \pm 0.316
	GM	0.547 \pm 0.159	0.475 \pm 0.237	0.622 \pm 0.252	0.514 \pm 0.218	0.477 \pm 0.322	0.555 \pm 0.342
	WM	0.515 \pm 0.185	0.520 \pm 0.288	0.506 \pm 0.254	0.509 \pm 0.209	0.472 \pm 0.317	0.542 \pm 0.316
GM	GM+WM	0.513 \pm 0.171	0.507 \pm 0.252	0.518 \pm 0.245	0.488 \pm 0.202	0.445 \pm 0.318	0.528 \pm 0.285
	GM	0.586 \pm 0.174	0.610 \pm 0.247	0.564 \pm 0.270	0.521 \pm 0.187	0.522 \pm 0.303	0.535 \pm 0.289
	WM	0.471 \pm 0.181	0.455 \pm 0.245	0.488 \pm 0.278	0.489 \pm 0.206	0.502 \pm 0.319	0.483 \pm 0.314
WM	WM	0.465 \pm 0.174	0.445 \pm 0.243	0.484 \pm 0.268	0.468 \pm 0.210	0.488 \pm 0.292	0.448 \pm 0.305

Table 7.2: Classification accuracy (Acc), sensitivity (Sen) and specificity (Spec) \pm standard deviation for each modality and mask using the participants acquired at the LON and CAM sites.

`synT1` modalities related to some abnormal participants that will be discussed 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 \[130\]](#) areas of: A) right superior frontal gyrus, Brodmann areas 6 ($z=60$); B) the pars opercularis of the left inferior frontal gyrus, Brodmann areas 44; C) the pars triangularis of the left inferior frontal gyrus, Brodmann areas 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.

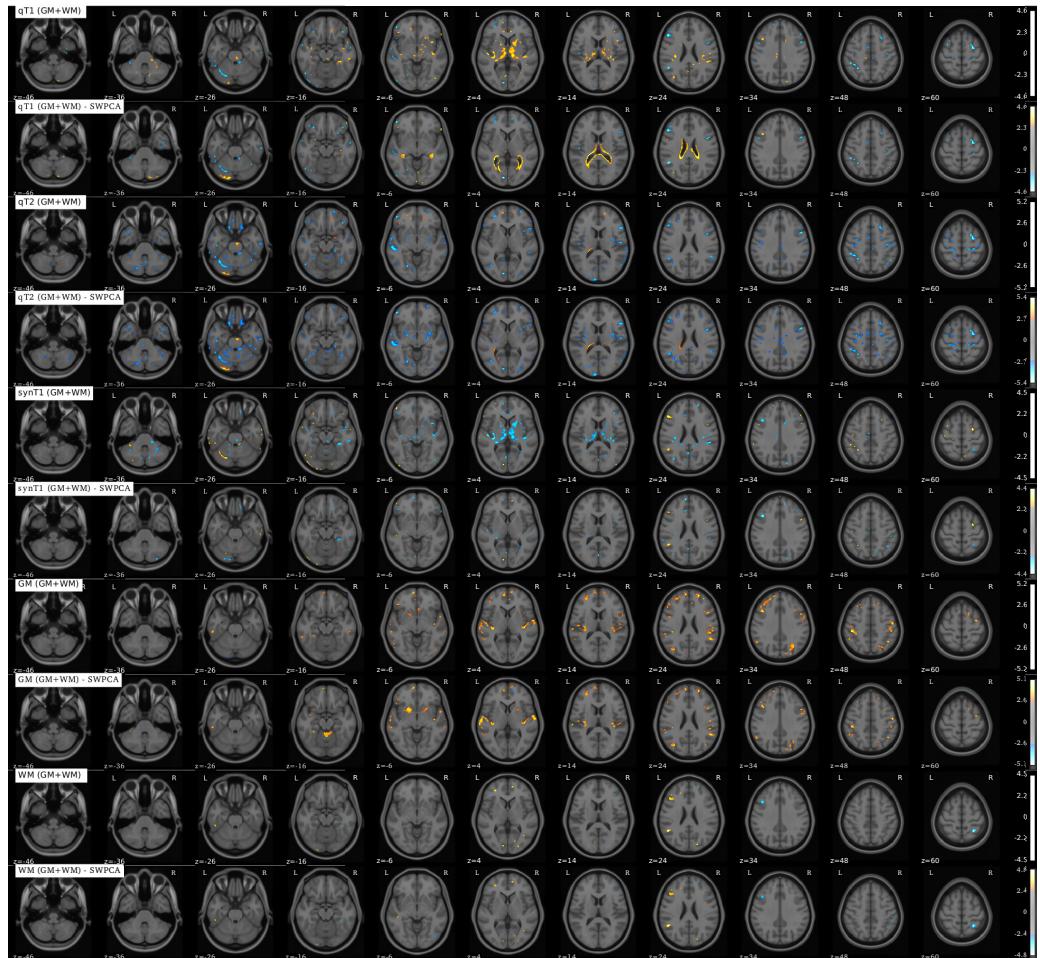


Figure 7.5: Brain t-map (VBM) of significant ($p < 0.01, |t| > 2.57$) GM and WM differences in ASD using **qT₁**, **qT₂**, **synT₁**, **GM** and **WM** maps before and after applying **SWPCA** to remove site effects.

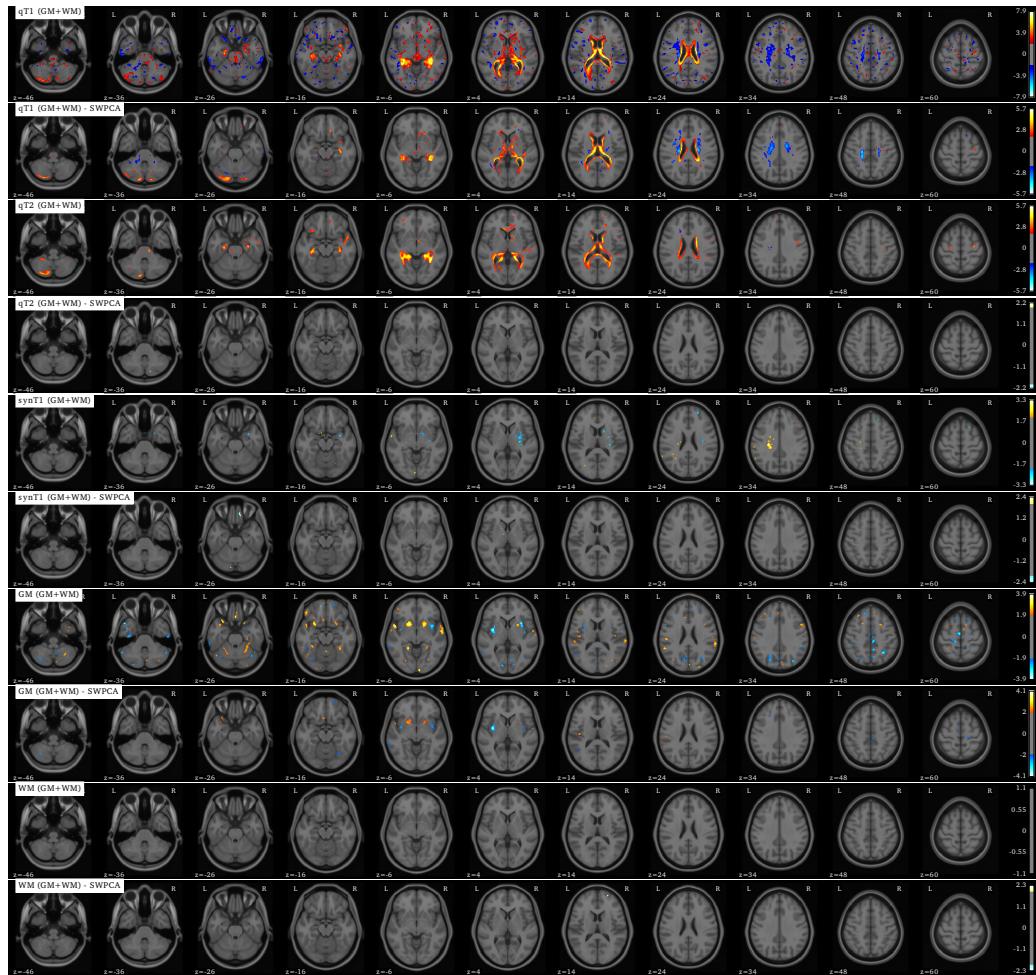


Figure 7.6: Brain Z-map (CBM) of significant ($p < 0.01, |t| > 2.57$) GM and WM differences using qT₁, qT₂, synT₁, GM and WM maps before and after applying SWPCA to remove site effects.

Modality	Mask	no-SWP _{CA}			SWP _{CA}		
		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 (Sen), 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](#), [32](#)]. 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 [[74](#)] and technical effects that range from the usage of different equipment or acquisition parameters [[78](#)] to physical changes that affect the performance of [MRI](#) scanners across time [[74](#)]. 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 [[105](#)].

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 the impact of the site-related geometric differences.

Regarding intensity correction, in the [MRC-AIMS](#) database used in this study [[31](#), [43](#)], a standardization procedure based on quantitative imaging [[82](#)] 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 vari-

ance that are due to multi-site acquisition and reduce them. Decomposition 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 [50]. 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 [40, 85]. 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](#) [105] 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 Partial Least Squares ([PLS](#)) algorithm [68] or Surrogate Variable Analysis ([SVA](#)) [98]. 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 [98], 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



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

complexity model that starts by eliminating the contribution 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 [111]), is a region where mirror neuron dysfunction has been consistently reported in **ASD**-affected children [104] and adults [28, 87, 106]. Wernicke's area, contained in the left TPJ, is also linked to language, and has been associated with **ASD** in several works [28, 106]. Additionally, the right TPJ has been proposed as related to mentalizing and has been repeatedly implicated in autism [113], including a **fMRI** study of a subsample of this same AIMS dataset [56]. The right superior frontal gyrus (region A) is more equivocal, with some studies [43, 64] reporting abnormalities in this area, while others [26, 106] report no significant differences. Our analyses reveal no differences in the insula and amygdala, brain structures frequently linked to autism.

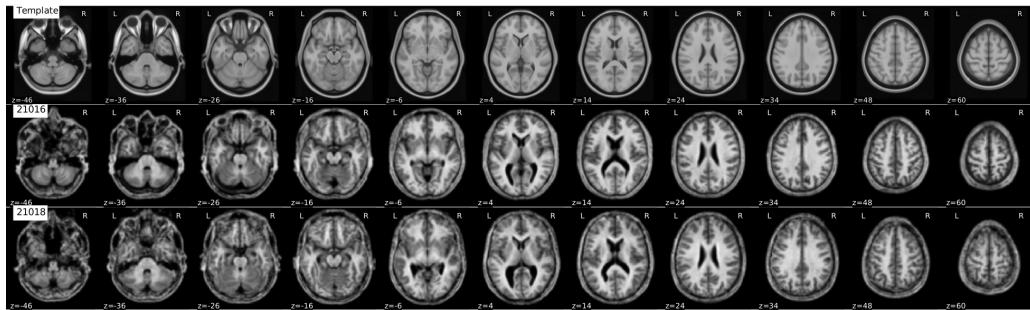


Figure 7.8: The template used in this work compared to two of the participants with abnormal ventricle size (21016 and 21018). Atrophy of the cerebellum in participant 21016 can also be appreciated, responsible for some of the ‘highlighted’ areas in [qT₁](#), [qT₂](#) and [synT₁](#) t-maps (see Fig.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

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 mea-

sures, 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 is focused 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 then projected to the brain space.

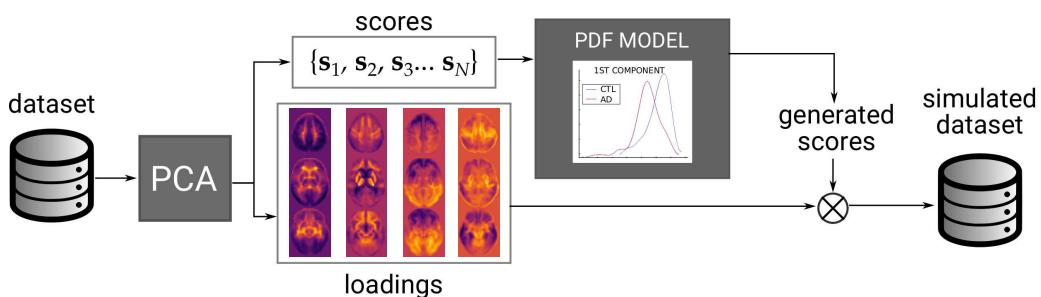


Figure 8.1: Schema of the brain image synthesis algorithm.

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 maximum strat-

egy, which proved to be very useful in most applications [25, 46] (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, 55], and that was already defined 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 the new samples, 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. To do so, 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 model of this distribution using a Multivariate Normal Distribution, and the empirical Kernel Density Estimation.

8.1.2.1 Multivariate Normal distribution ([MVN](#))

The Multivariate Normal distribution ([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:

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

8.1.2.2 Kernel Density Estimation (KDE)

Kernel Density Estimation (KDE) is an increasingly used method to estimate the PDF of a set of data [49, 63]. The KDE estimates the probability density function of the i^{th} coordinate of class c as:

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

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:

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

It also must be centred:

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

and its covariance matrix must be close to identity:

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

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 [63].

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 Cumulative Density Function (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 $\hat{\mathbf{S}}^c$.

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

8.2 Experimental Setup

Validation of a simulated dataset is still a matter of discussion. Most works exploring that possibility [143, 163] 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](#) [119] 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](#) [97] 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 training set is used to generate a new training set of completely new images, and after the classifier is trained 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 us-

ing 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 [97] has been performed on both the original and the simulated datasets. Mass-univariate two-sample t-maps are provided in Figure 8.7, 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 are 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 class, which in the best case ([AD](#) vs [CTL](#)), only achieves up to 80% accuracy. However, the scenarios including [MCI](#) perform similarly to the original dataset.

On the other hand, we take a look at the PPMI dataset. In this case, we have highly specific functional imaging, in which the dopaminergic level (cause of Parkinson's Disease) is assessed. Therefore, the application of [VAF](#) to the

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
		PD vs CTL	0.997 ± 0.007	0.995 ± 0.015	1.000 ± 0.011
		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.

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. We will discuss this later.

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 confused 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 achieves 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 an 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, 55]. In [AD](#), the computed eigenbrains model different 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.

Database	Est	Scenario	Test set	acc ($\pm SD$)	sens ($\pm SD$)	spec ($\pm SD$)
ADNI-PET	MVN	AD vs CTL	Original	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	
	KDE	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	
PPMI-DAT	MVN	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	
	KDE	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	
PPMI-DAT	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	

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

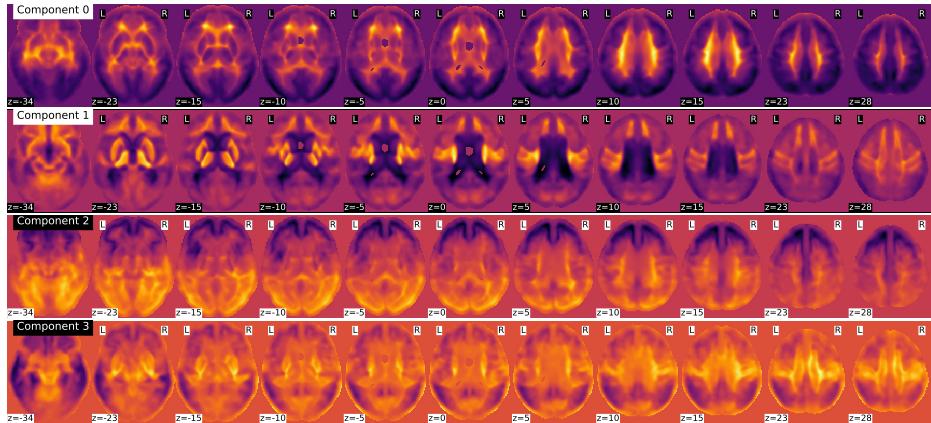


Figure 8.2: Illustration of the three first eigenbrains in PCA.

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 [55, 119]. 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.

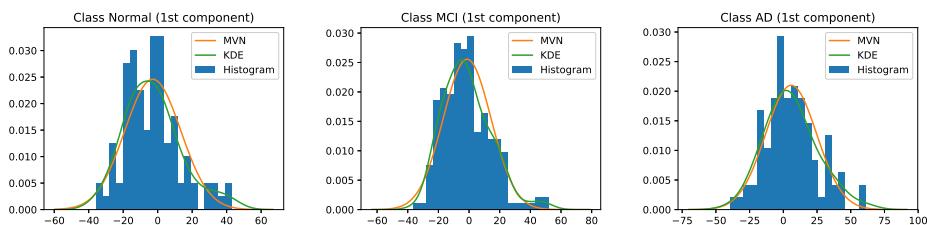


Figure 8.3: Comparison between the MVN and KDE PDF estimation methods for the first component and the three groups, setting the histogram as reference.

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.

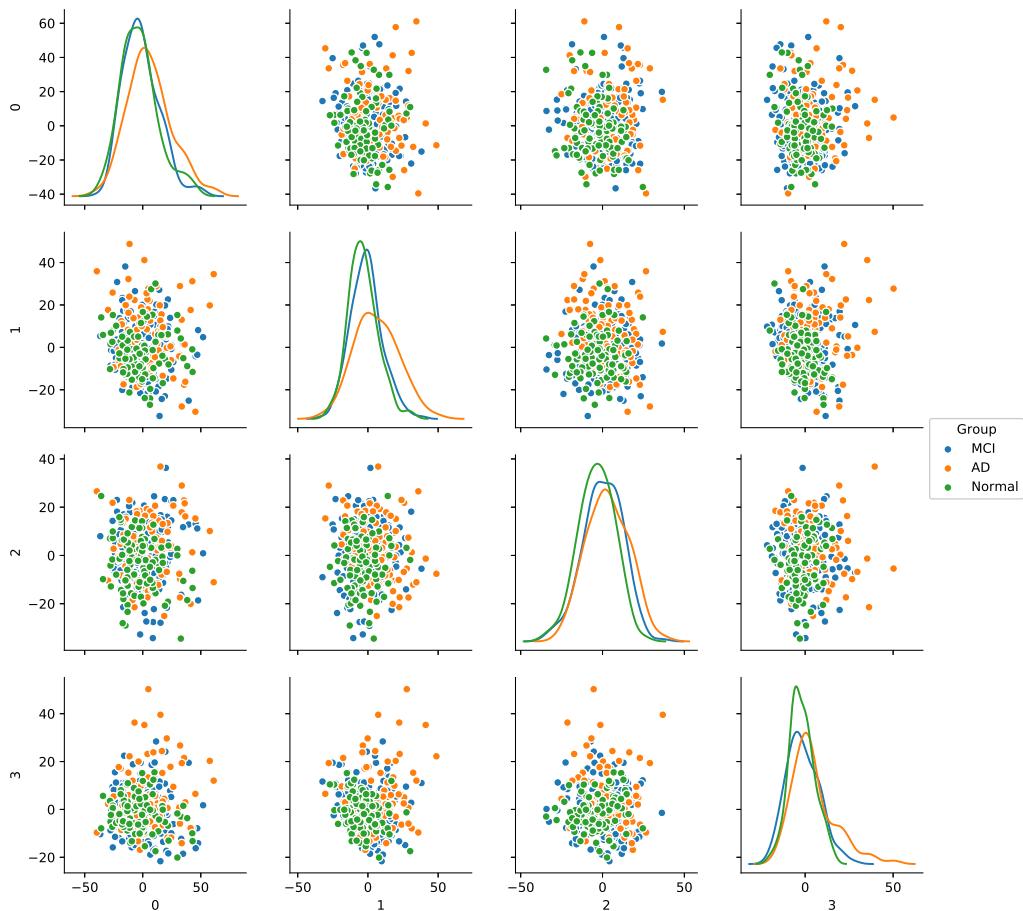


Figure 8.4: Pairplot plotting the coordinates of the subjects in the ADNI-PET database over the four first eigenbrains. In the diagonal, distribution of the classes estimated using **KDE** for each eigenbrain.

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 [55, 119]. In the simulated DaTSCAN images (Fig. 8.6) differences

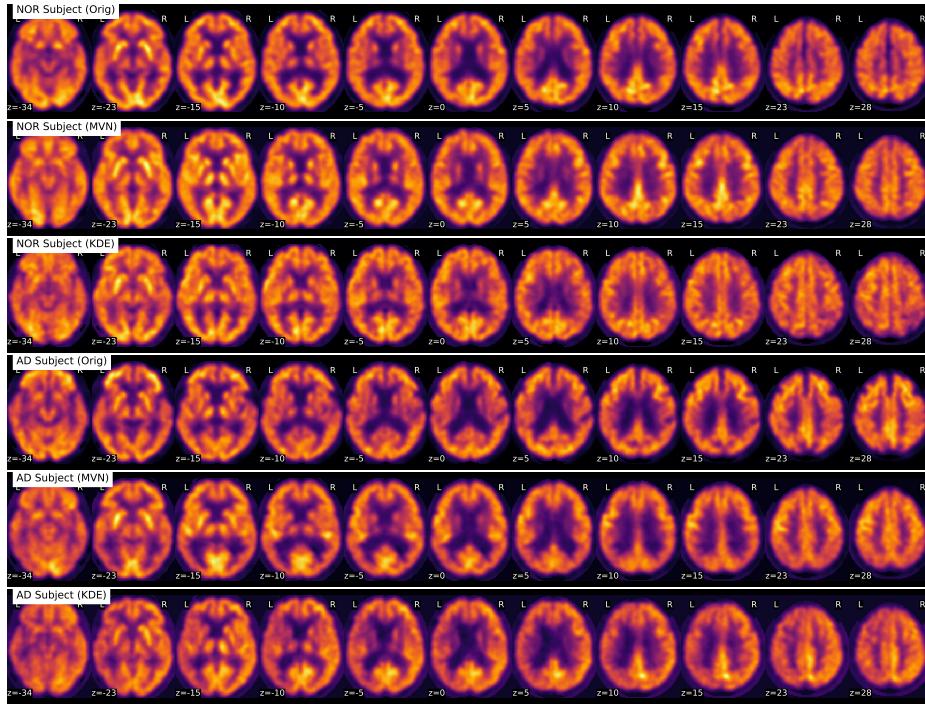


Figure 8.5: Examples of some original and simulated subjects from the ADNI-PET dataset.

in shape and intensity of the striatum, and bilateral differences [25, 44, 59] can also be noticed, using both **MVN** and **KDE** modelling.

To deepen the analysis of the synthesized figures, we will perform a **SPM** analysis [97], 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 intense, represented in the **SPM** analysis of the original dataset. The aforementioned regions such as the precuneus, angular, mid-temporal lobe, hippocampus, amygdala, among others, are represented in both the original and simulated 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 signifi-

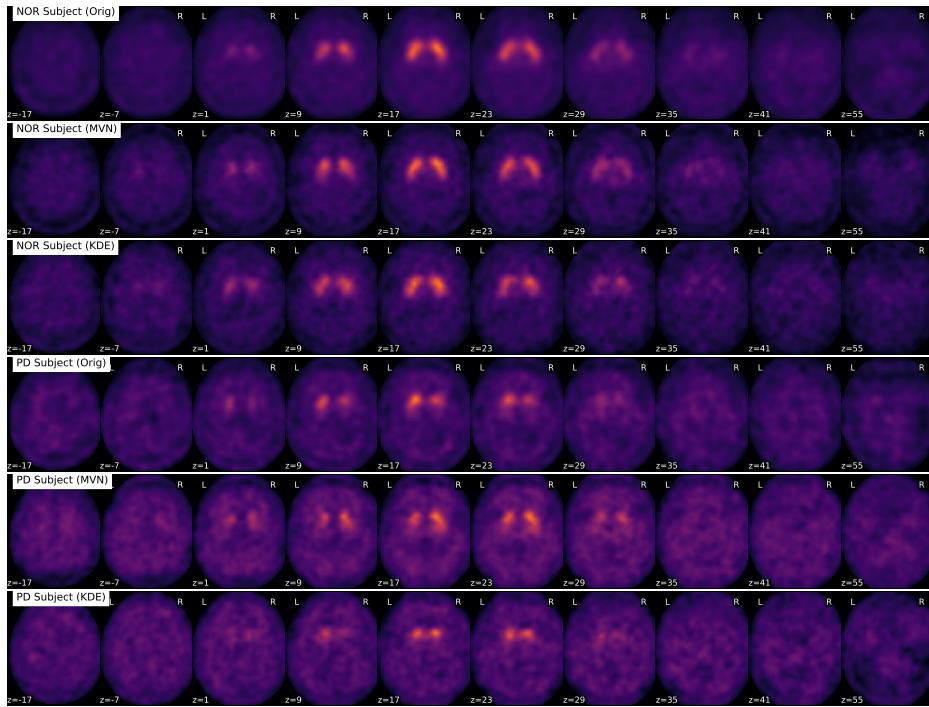


Figure 8.6: Examples of some original and simulated subjects from the PPMI dataset.

cance in the case of the **MVN** model, and a more homogeneous distribution of the negative differences under the **KDE** model.

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, 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 encom-

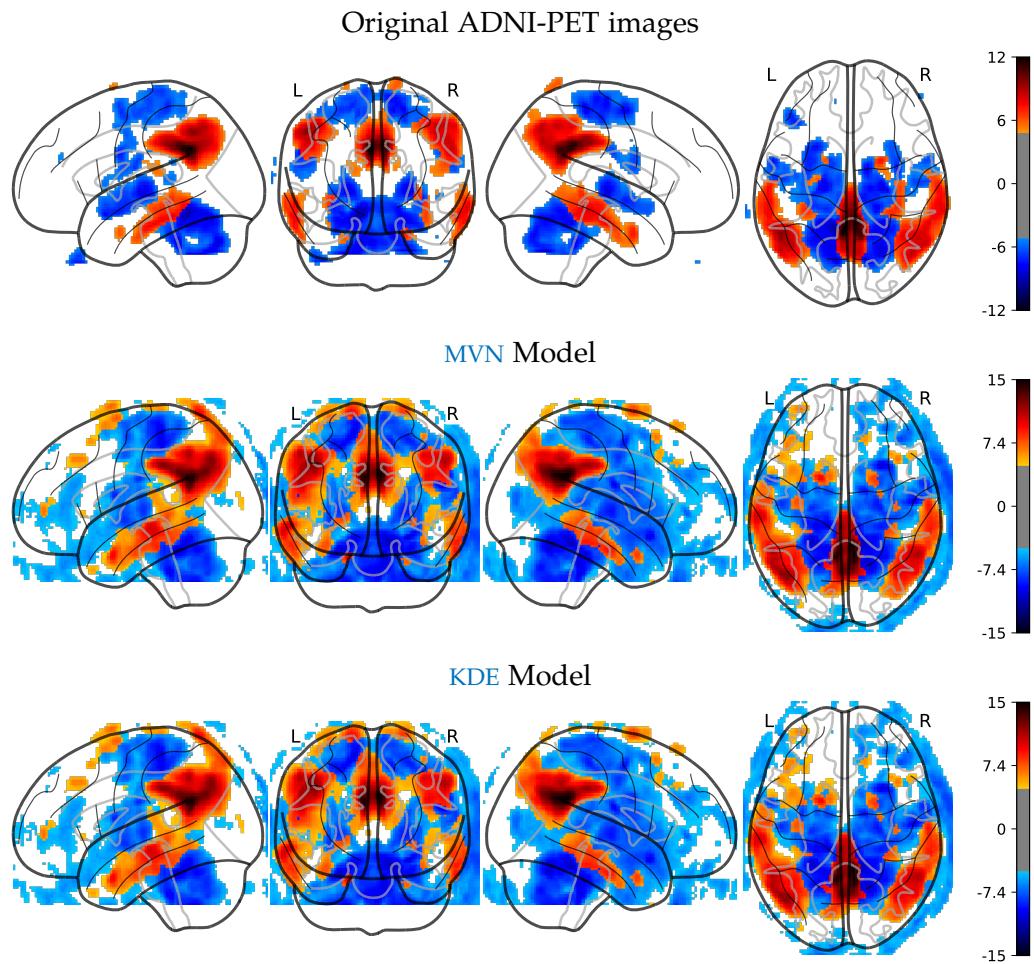


Figure 8.7: SPM Analysis of the ADNI dataset (AD vs CTL), FWE corrected, with $p = 0.05$, for the original and the simulated images.

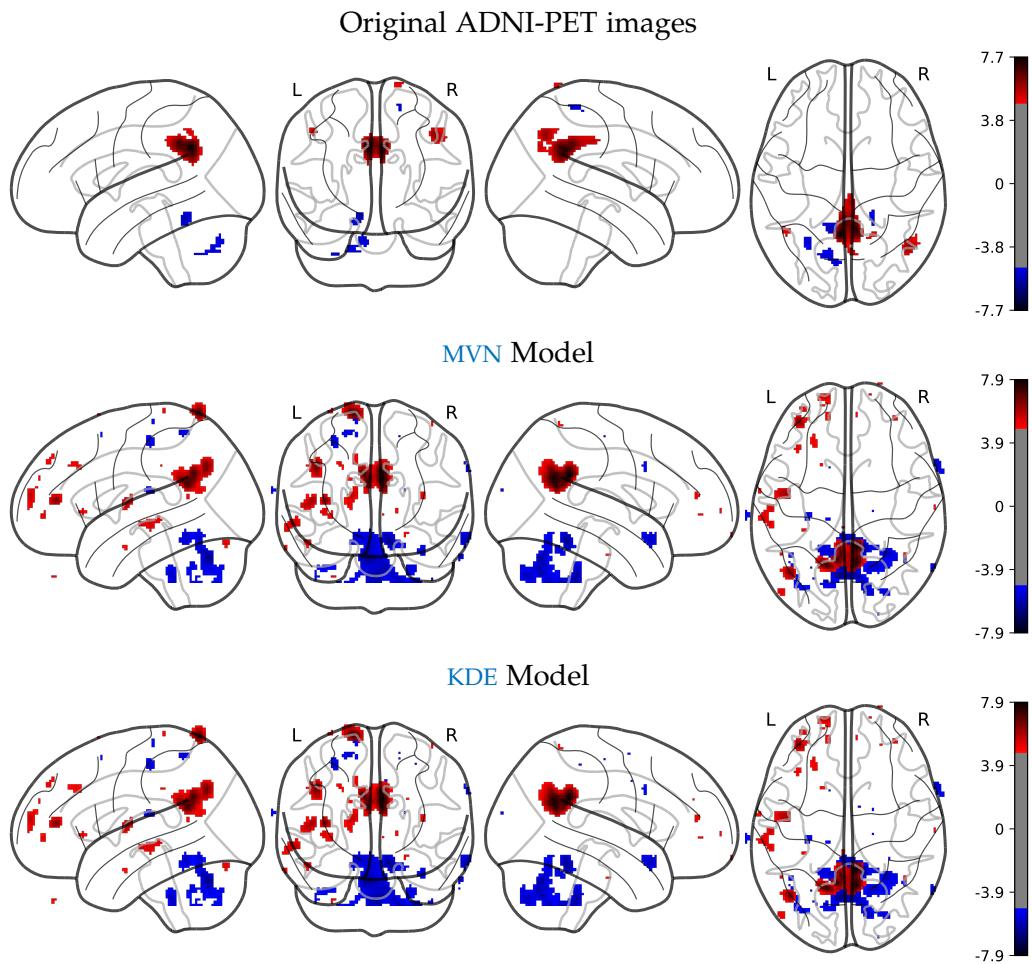


Figure 8.8: SPM Analysis of the ADNI dataset (CTL vs MCI), FWE corrected, with $p = 0.05$, for the original and the simulated images.

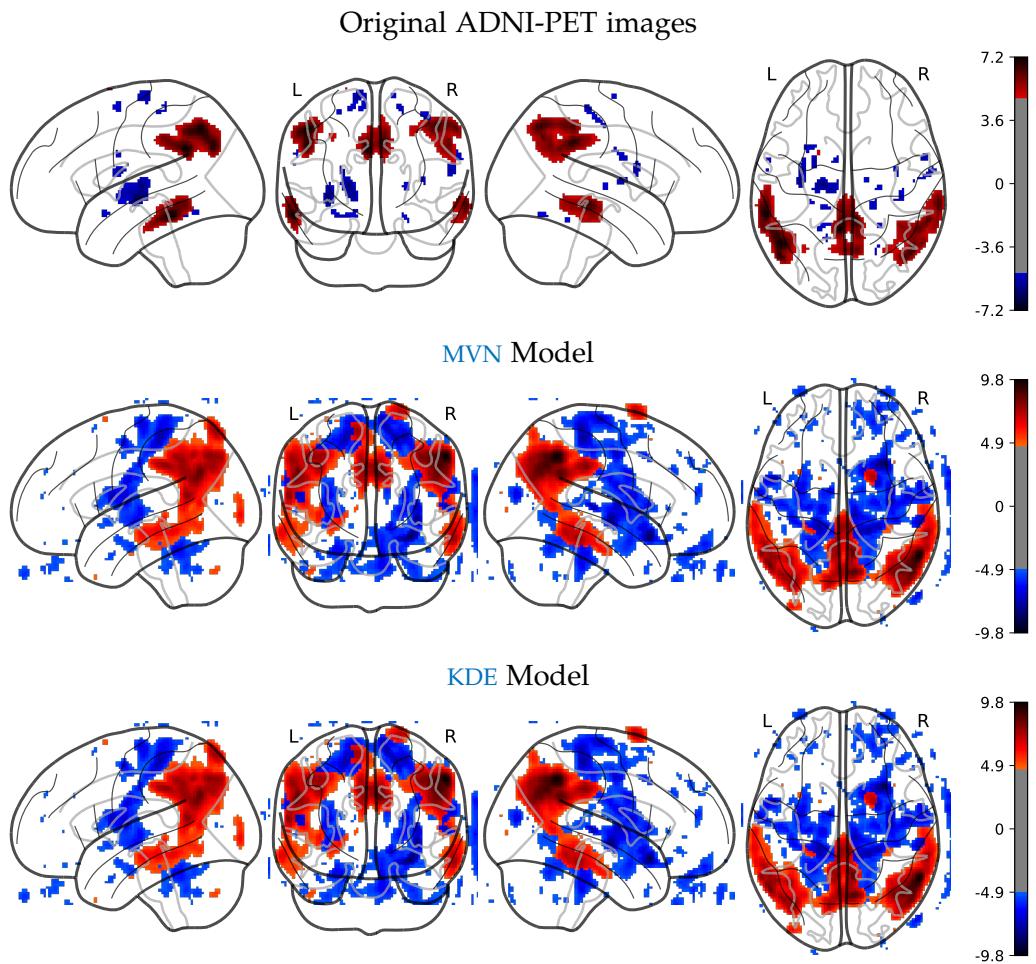


Figure 8.9: SPM Analysis of the ADNI dataset (MCI vs AD), FWE corrected, with $p = 0.05$, for the original and the simulated images.

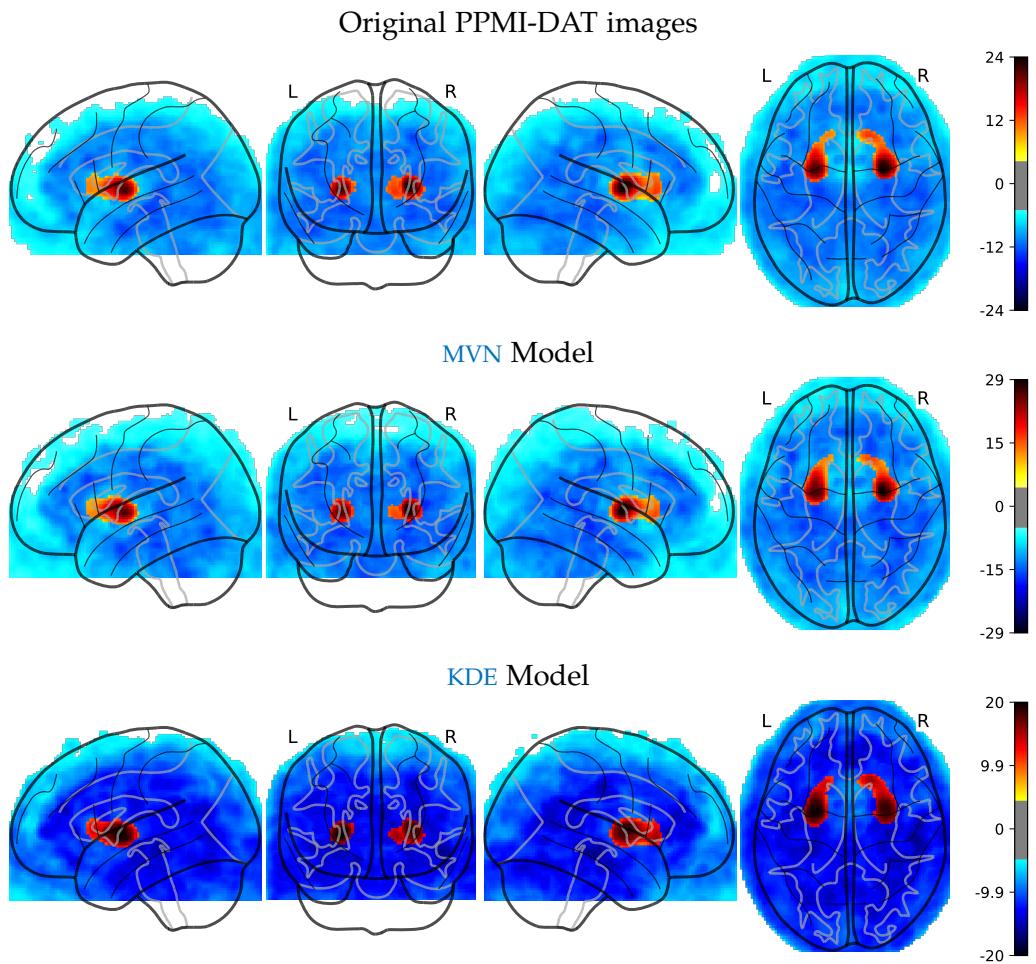


Figure 8.10: SPM Analysis of the PPMI dataset (PD vs CTL), FWE corrected, with $p = 0.05$, for the original and the simulated images.

pass 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 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.

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** 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 softness 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) -COMPROBAR-, 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 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 the nondescript amount of 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 have been 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 correct this behaviour by decomposing the images using PCA and parsing out the components whose contribution was more related to the acquisition site. After ap-

plying [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 merging structural dataset 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 diseases

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, and other regions with a widely documented relation to [AD](#) are highlighted [95, 159], such as the cingulum, hippocampus and parahippocampal gyrus.

It is clearly noticeable that the relative entropy selection criterion focuses on many different 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 [65], and leads to a higher performance in the systems 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 differences in the angular lobe and also all over the occipital lobe and parts of the temporal lobe.

9.2 Conclusions

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](#). These are only recent contributions to the neuroimage processing

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 that will be used throughout this thesis, 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) [97], or segmented [GM](#) and [WM](#) maps. Segmentation was performed using the VBM8 toolbox for SPM [116].

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) [161]. Autism Diagnostic Observation Schedule (ADOS) [140] 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) [147], 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) [82] 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 [31, 43, 45]. The [synT₁](#) images were then segmented using New Segment into [GM](#) and [WM](#) maps, and normalized to the [MNI](#) space using DARTEL in SPM8 [97], 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 `synT1`, `qT1` and `qT2` 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 T₁ contrast, T₁ relaxation time, and T₂ relaxation times, and smoothed with a 3mm FWHM Gaussian kernel. Therefore, there were five different modalities: `qT1`, `qT2`, `synT1` 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 (`qT1` and `qT2`), simulated (`synT1`) 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-PET database also refers from images acquired at the `ADNI`. However, in this case we will use ¹⁸F-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.

Further preprocessing of these images included spatial normalization, via the SPM8 software and its `PET` template, and intensity normalization. The type of

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-PET dataset.

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 possesses 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 [[117](#), [148](#)]. Given that this is an inherent limitation of 'in vivo' studies, our working-assumption is that the labels are true, considering the subject label positive when belonging to any of the [AD](#) classes, and negative otherwise.

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.

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 **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 striatum, 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**, 73 **PD** 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

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 [[166](#), [170](#)], application of a Hanning filter (frequency 0.7) and images were obtained with transaxial cuts, following the method proposed in [[75](#)].

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 caudate 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 dopaminer-

gic transporters through direct competitive mechanism were excluded. A more detailed description of the database can be found in [99].

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

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 [57, 66]. 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

Support Vector Machines (**SVMs**), introduced in the late 70s [149], 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 the ones presented in this thesis [1, 6, 7, 24, 25, 34] and many others [8, 35, 44, 55, 59, 119].

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 \mathbf{X} , which contains N p -dimensional vectors (samples) \mathbf{x}_i and their corresponding class y_i :

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l) \in \Re^p \times \{\pm 1\} \quad (\text{B.1})$$

The linear **SVC** defines a separation hyperplane in a multidimensional space using the following function:

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \omega_0 = 0, \quad (\text{B.2})$$

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:

$$\mathbf{x}_i \mathbf{w} + \omega_0 \geq +1 \quad \text{for } y_i = +1 \quad (\text{B.3})$$

$$\mathbf{x}_i \mathbf{w} + \omega_0 \leq -1 \quad \text{for } y_i = -1 \quad (\text{B.4})$$

$$(\text{B.5})$$

which can be combined to a simpler equation:

$$y_i (\mathbf{x}_i \mathbf{w}) - 1 \geq 0 \quad \forall i \quad (\text{B.6})$$

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:

$$\text{minimize}_{\mathbf{w}} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i^L \xi_i \quad \text{s.t.} \quad y_i(\mathbf{x}_i \mathbf{w}) - 1 \geq 0 \quad \forall i \quad (\text{B.8})$$

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** [52] 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:

$$\text{minimize}_{\mathbf{w}} \quad \frac{\mathbf{w}^\top \mathbf{w}}{2} + C \sum \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i)^2 \quad (\text{B.9})$$

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 [86].

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. 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).
- [5] 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).
- [6] F.J. Martinez-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. Martinez-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. 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: [10.1111/1.4845115](https://doi.org/10.1111/1.4845115).
- [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] 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).
- [31] 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).
- [32] Rhoshel K Lenroot and Pui Ka Yeung. "Heterogeneity within autism spectrum disorders: what have we learned from neuroimaging studies?" In: (2013).

- [33] 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).
- [34] 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).
- [35] 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).
- [36] 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).
- [37] 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).
- [38] 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).
- [39] 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.

- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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).

- [47] 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).
- [48] 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.
- [49] Jeffrey S Simonoff. *Smoothing methods in statistics*. Springer Science & Business Media, 2012.
- [50] 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.
- [51] 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).
- [52] 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).
- [53] 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).
- [54] 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).
- [55] 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](https://doi.org/10.1016/j.ins.2010.10.027).

- [56] 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.
- [57] 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](https://doi.org/10.1016/j.pneurobio.2011.09.005).
- [58] 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).
- [59] 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](https://doi.org/10.1097/MNM.0b013e328347cd09).
- [60] 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.
- [61] Yoav Benjamini. "Simultaneous and selective inference: current successes and future challenges". In: *Biometrical Journal* 52.6 (2010), pp. 708–721.
- [62] 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.
- [63] ZI Botev, JF Grotowski, DP Kroese, et al. "Kernel density estimation via diffusion". In: *The Annals of Statistics* 38.5 (2010), pp. 2916–2957.
- [64] 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.

- [65] 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.
- [66] The Parkinson Progression Markers Initiative. *PPMI. Imaging Technical Operations Manual*. 2nd ed. June 2010.
- [67] 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](https://doi.org/10.1016/j.neuroimage.2010.07.020).
- [68] V Vinzi, Wynne W Chin, Jörg Henseler, Huiwen Wang, et al. *Handbook of partial least squares*. Springer, 2010.
- [69] 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.
- [70] 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.
- [71] 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](https://doi.org/10.1007/978-1-84628-723-7_8).
- [72] 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.
- [73] 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](https://doi.org/10.1016/j.neuroimage.2008.10.031).
- [74] Godfrey Pearlson. *Multisite collaborations and large databases in psychiatric neuroimaging: advantages, problems, and challenges*. 2009.

- [75] 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](https://doi.org/10.1049/el.2009.1111).
- [76] 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](https://doi.org/10.1016/j.neulet.2009.05.056).
- [77] 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.
- [78] John Darrell Van Horn and Arthur W Toga. "Multi-site neuroimaging trials". In: *Current opinion in neurology* 22.4 (2009), p. 370.
- [79] 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.
- [80] 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.
- [81] 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.
- [82] 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.
- [83] 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](https://doi.org/10.1093/brain/awn016).

- [84] 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](https://doi.org/10.1093/brain/awn278).
- [85] 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.
- [86] 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.
- [87] 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.
- [88] 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.
- [89] 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* (2008), pp. 29–34.
DOI: [10.1109/BIOTECHNO.2008.24](https://doi.org/10.1109/BIOTECHNO.2008.24).
- [90] GN Srinivasan and G Shobha. "Statistical texture analysis". In: *Proceedings of world academy of science, engineering and technology*. Vol. 36. 2008, pp. 1264–1269.
- [91] 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.
- [92] Klaus Tatsch. "Extrapyramidal syndromes: PET and SPECT". In: *Diseases of the Brain, Head & Neck, Spine*. Springer, 2008, pp. 234–239.
- [93] John Ashburner. "A fast diffeomorphic image registration algorithm". In: *Neuroimage* 38.1 (2007), pp. 95–113.

- [94] 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.
- [95] 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](https://doi.org/10.1016/s1474-4422(07)70178-3).
- [96] 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](https://doi.org/10.1016/j.cnr.2007.05.001).
- [97] 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.
- [98] Jeffrey T Leek and John D Storey. "Capturing heterogeneity in gene expression studies by surrogate variable analysis". In: *PLoS Genet* 3.9 (2007), e161.
- [99] 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](https://doi.org/10.1016/S1578-200X(07)70062-1).
- [100] 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.
- [101] 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).
- [102] 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).
- [103] 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.
- [104] 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/nrn1611](https://doi.org/10.1038/nrn1611).
- [105] 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.
- [106] 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.
- [107] Xiao Han, Jorge Jovicich, David Salat, Andre van der Kouwe, Brian Quinn, Silvester 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.
- [108] Jorge Jovicich, Silvester 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.
- [109] John Rice. *Mathematical statistics and data analysis*. Nelson Education, 2006.
- [110] 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.

- [111] Nobuyuki Nishitani, Martin Schürmann, Katrin Amunts, and Riitta Hari. "Broca's region: from action to language". In: *Physiology* 20.1 (2005), pp. 60–69.
- [112] Christoph Scherfler, Klaus Seppi, Eveline Donnemiller, Georg Goebel, Christian Brenneis, Irene Virgolini, Gregor K Wenning, and Werner Poewe. "Voxel-wise analysis of [^{123}I]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).
- [113] 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.
- [114] 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).
- [115] 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.
- [116] 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).
- [117] S. Bruggink K. P. Ebmeier N. J. Dougall. "Systematic review of the diagnostic accuracy of $^{99\text{m}}\text{Tc}$ -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.
- [118] 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).
- [119] 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.

- [120] 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.
- [121] 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).
- [122] 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.
- [123] I. Guyon and A. Elisseeff. "An Introduction to Variable and Feature Selection". In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182.
- [124] Aapo Hyvärinen and Yutaka Kano. "Independent component analysis for non-normal factor analysis". In: *New developments in psychometrics*. Springer, 2003, pp. 649–656.
- [125] 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).
- [126] 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.
- [127] 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.
- [128] 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.

- [129] 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.
- [130] 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).
- [131] Marti J Anderson and John Robinson. "Permutation tests for linear models". In: *Australian & New Zealand Journal of Statistics* 43.1 (2001), pp. 75–88.
- [132] 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).
- [133] 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.
- [134] 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.
- [135] 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).
- [136] John Ashburner and Karl J Friston. "Voxel-based morphometry—the methods". In: *Neuroimage* 11.6 (2000), pp. 805–821.
- [137] 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](https://doi.org/10.1109/ICPR.2000.906006).

- [138] 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](https://doi.org/10.1016/S0893-6080(00)00026-5).
- [139] 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>.
- [140] 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.
- [141] 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](https://doi.org/10.1006/nimg.1998.0395).
- [142] A. Hyvärinen. "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis". In: *IEEE Transactions on Neural Networks* 10.3 (1999), pp. 626–634.
- [143] 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.
- [144] 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.
- [145] P Szatmari. "Heterogeneity and the genetics of autism." In: *Journal of psychiatry & neuroscience : JPN* 24 (2 Mar. 1999), pp. 159–165. ISSN: 1180-4882.
- [146] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. New York: Academic Press, 1999. ISBN: 9781597492720.
- [147] David Wechsler. *Wechsler abbreviated scale of intelligence*. Psychological Corporation, 1999.

- [148] 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.
- [149] V. N. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, Inc., New York, 1998.
- [150] 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.
- [151] A. Hyvärinen and E. Oja. "A fast fixed-point algorithm for independent component analysis". In: *Neural Computation* 9 (1997), pp. 1483–1492.
- [152] 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.
- [153] 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](https://doi.org/10.1006/nimg.1996.0019).
- [154] 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](https://doi.org/10.1016/0031-3203(95)00067-4).
- [155] 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.
- [156] 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>.
- [157] Juliet Popper Shaffer. "Multiple hypothesis testing". In: *Annual review of psychology* 46.1 (1995), pp. 561–584.

- [158] 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](https://doi.org/10.1109/83.366472).
- [159] 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](https://doi.org/10.1159/000051283).
- [160] 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.
- [161] 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](https://doi.org/10.1007/BF02172145).
- [162] 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.
- [163] 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.
- [164] RM Haralick and LG Shapiro. *Robot and Computer Vision (vols. 1 and 2)*. 1992.
- [165] J Talairach and P Tournoux. *Co-Planar Stereotactic Atlas ofthe Human Brain*. 1988.
- [166] 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.
- [167] 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.
- [168] David Freedman and David Lane. “A nonstochastic interpretation of reported significance levels”. In: *Journal of Business & Economic Statistics* 1.4 (1983), pp. 292–298.
- [169] “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.
- [170] L. A. Shepp and Y. Vardi. “Maximum Likelihood Reconstruction for Emission Tomography”. In: *IEEE Transactions on Medical Imaging* 1.2 (1982), pp. 113–122.
- [171] 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.