

Experiments of Machine Learning for Neurodiagnosis

by

Alexandre Savio

Submitted to the department of
Computer Science and Artificial Intelligence
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

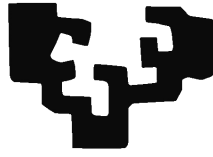
PhD Advisor:

Prof. Dr. Manuel Graña Romay
at The University of the Basque Country

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Donostia - San Sebastián

2013



Experiments of Machine Learning for Neurodiagnosis

By

Alexandre Savio

Submitted to the department of Computer Science and Artificial Intelligence in partial
fulfillment of the requirements for the degree of
Doctor of Philosophy

PhD Advisor:

Prof. Dr. Manuel Graña Romay
at The University of the Basque Country

Universidad del País Vasco
Euskal Herriko Unibertsitatea
Donostia - San Sebastián

2013

Acknowledgements

Special and kind thanks to my research mentor professor Manuel Graña. Without his mentoring, friendship, paternalism and insanity this work would have lost half the fun. His constructive ideas, support and crude social logics were crucial for the development of this thesis and very helpful for whatever comes next.

Thanks to my group colleagues for being there when I wanted to have a beer or coffee with good company and/or bad scientific discussion.

My deepest gratitude to Prof. Stephen Smith, Dr. Saad Jbabdi and Prof. Tim Behrens for giving me the opportunity of doing an internship in the FMRIB Centre at the University of Oxford. It was a great experience. Thanks to all the great people I've met in research and still keep in touch.

I would also like to thank all my beloved families for the support and encouragement. To my dad for the side mentoring and experience and my mom for the solace, pragmatism and good sense. To my grandparents for the good holidays, specially to my grandfather Prof. Octavio Manhães de Andrade Jr. for his support, Math books and because probably if he wasn't who he is I wouldn't have done this. I dedicate this thesis to him and my grandmother who is happy in whatever dimension she might be.

Thanks to all my friends and people close to me for the needed drops of reality during these research years. Thanks to Almudena for the understanding and acceptance, indeed it took longer than what I told you.

My heartfelt thanks to Miren for suffering me full-time with a smile during this last year.

A blissful thank you!

Alexandre Manhães Savio

Experiments of Machine Learning for Neurodiagnosis

by

Alexandre Savio

Submitted to the Department of Computer Science and Artificial Intelligence, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Abstract

The application of Machine Learning algorithms to Neuroscience data has two main goals in this thesis. First, the construction of Computer Aided Diagnostic (CAD) systems to help alleviate the burden of increasing amounts of data for diagnosis. Second, the identification of image biomarkers corresponding to anatomical locations of the features selected for classification. This thesis is an empirical exploration of these ideas in the case of three neurological diseases. We have developed sound methodological frameworks, avoiding circularity effects in the validation process. For feature selection we have been working with supervised methods which produce good classification results and the ability to determine the spatial location of the features in the brain, among them an evolutionary wrapper selection method based on Extreme Learning Machines. The thesis covers experiments with a wide spectrum of classifiers for comparison. Finally, the approach has been tested on different MRI modalities showing its general applicability. A critical issue for multivariate modalities is the definition of appropriate scalar measures that may be useful for feature selection and extraction. The thesis tests the most appropriate for each modality: FA and MD for diffusion data, functional activity measures such as ReHo for functional data and measures from deformation maps resulting from non-linear registration of anatomical data. The results obtained show that the approach is useful for CAD systems in a variety of neurological diseases.

Keywords: *Medical Image Processing, Computer Aided Diagnostics, Image Biomarkers, Feature Selection, Alzheimer's Disease, Schizophrenia*

“Este no es un grupo serio.”

“This is not a serious group.”

Manuel Graña

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Motivation | 1 |
| 1.2 | Summary of the Thesis Contributions | 5 |
| 1.3 | Publications | 6 |
| 1.4 | Contents of the Thesis | 9 |
| 2 | Medical Background and Related Studies | 13 |
| 2.1 | Alzheimer’s Disease | 13 |
| 2.1.1 | Related Studies | 14 |
| 2.2 | Myotonic Dystrophy Type 1 | 16 |
| 2.2.1 | Related Studies | 19 |
| 2.3 | Schizophrenia | 19 |
| 2.3.1 | Related Studies | 20 |
| 3 | Pre-processing | 23 |
| 3.1 | Image Registration basics | 24 |
| 3.2 | Voxel-based Morphometry | 26 |
| 3.3 | Deformation-based morphometry | 29 |
| 3.4 | Diffusion Weighted Imaging (DWI) | 31 |
| 3.4.1 | Pre-processing | 33 |
| 3.5 | Resting state functional MRI (fMRI) | 34 |
| 3.5.1 | Pre-processing | 34 |
| 3.5.2 | Voxel activity measures from resting-state fMRI | 36 |
| 4 | Feature Selection | 37 |
| 4.1 | General procedure | 37 |
| 4.2 | Voxel significance measures | 38 |
| 4.2.1 | Pearson’s correlation | 38 |
| 4.2.2 | Bhattacharyya’s distance | 39 |
| 4.2.3 | Welch’s t-test | 39 |

| | | |
|----------|---|-----------|
| 4.2.4 | Spearman's correlation | 40 |
| 4.3 | Voxel-based Morphometry features | 40 |
| 4.4 | Atlas based GM feature selection | 42 |
| 4.5 | Deformation-based Morphometry | 43 |
| 4.6 | Scalar Features from Diffusion Tensors | 45 |
| 4.7 | Evolutionary wrapper feature selection | 49 |
| 5 | Classification | 53 |
| 5.1 | Problem statement | 53 |
| 5.2 | Support Vector Machines (SVM) | 54 |
| 5.3 | Artificial Neural Networks (ANN) | 57 |
| 5.3.1 | Multi Layer Perceptron trained with Backpropagation | 57 |
| 5.3.2 | Radial Basis Function Networks | 58 |
| 5.3.3 | Probabilistic Neural Networks | 59 |
| 5.3.4 | Learning Vector Quantization | 60 |
| 5.4 | Random Forests (RF) | 61 |
| 5.5 | Bootstrapped Dendritic Classifiers | 62 |
| 5.6 | Extreme Learning Machines | 62 |
| 5.6.1 | Basic ELM | 63 |
| 5.6.2 | Ensembles of ELM | 63 |
| 5.6.3 | Hybrid Extreme Rotation Forest | 64 |
| 5.7 | Adaptive Boosting | 65 |
| 5.8 | Ensemble of independent classifiers | 66 |
| 5.8.1 | Model selection and validation | 66 |
| 5.8.1.1 | Combination of boosted SVMs | 66 |
| 6 | Results | 69 |
| 6.1 | Performance measures | 69 |
| 6.2 | Experimental design | 71 |
| 6.3 | Alzheimer's Disease | 72 |
| 6.3.1 | OASIS98 | 72 |
| 6.3.1.1 | Experiment 1 | 72 |
| 6.3.1.2 | Experiment 2 | 77 |
| 6.3.2 | OASIS | 79 |
| 6.3.2.1 | Experiment 1 | 79 |
| 6.3.2.2 | Experiment 2 | 85 |
| 6.3.3 | Data from Hospital de Santiago | 87 |
| 6.4 | Myotonic Dystrophy Type 1 | 93 |
| 6.5 | Schizophrenia | 98 |
| 6.5.1 | NAMIC: Brain Multimodality | 98 |

| | | |
|----------|---|------------|
| 6.5.2 | COBRE Schizophrenia | 102 |
| 6.5.2.1 | Experiment 1 | 102 |
| 6.5.2.2 | Experiment 2 | 106 |
| 6.5.2.3 | Experiment 3 | 112 |
| 7 | Conclusions | 119 |
| 7.1 | Methodological conclusions | 119 |
| 7.2 | Disease-related conclusions | 119 |
| 7.2.1 | Alzheimer's Disease (AD) | 120 |
| 7.2.2 | Schizophrenia | 121 |
| 7.3 | Operational conclusions | 122 |
| A | Databases | 125 |
| A.1 | Open Access Series of Imaging Studies (OASIS) | 125 |
| A.1.1 | OASIS98 | 126 |
| A.1.2 | Whole OASIS | 126 |
| A.2 | NAMIC: Brain Multimodality | 127 |
| A.3 | COBRE Schizophrenia | 128 |
| A.4 | Myotonic Dystrophy Type 1 | 129 |
| A.5 | Hospital de Santiago | 131 |
| B | Software | 133 |
| B.1 | NeuroDebian Repository[163] | 133 |
| B.2 | SPM | 133 |
| B.3 | FSL | 134 |
| B.4 | AFNI | 135 |
| B.5 | FreeSurfer | 135 |
| B.6 | LibSVM | 135 |
| B.7 | SVM-Perf | 136 |
| B.8 | C-PAC | 136 |
| B.9 | Python Libraries | 138 |
| B.9.1 | Numpy and Scipy | 138 |
| B.9.2 | NiBabel | 138 |
| B.9.3 | Scikit-Learn | 138 |
| B.9.4 | Author's software releases | 139 |
| B.9.4.1 | Aizkolari | 139 |
| B.9.4.2 | Other Python tool-sets | 140 |
| | Bibliography | 141 |

List of Figures

| | | |
|-----|---|----|
| 1.1 | Structure of the thesis | 4 |
| 1.2 | Specification of the classification experiments performed during this thesis with colleagues and collaborators. | 10 |
| 3.1 | VBM computational pipeline. | 27 |
| 3.2 | SPM results: clusters of significant voxels with increased grey matter density in the controls relative to the patient subjects, detected by the VBM process in the OASIS data. | 29 |
| 3.3 | Pipeline of the image pre-processing steps. | 30 |
| 3.4 | Example of deformation measure maps from the OASIS database (a) Jacobian determinant, (b) modulated GM, (c) displacement norm, (d) Jacobian matrix trace and (e) geodesic anisotropy. | 32 |
| 3.5 | Image preprocessing pipeline for the preprocessing the DWI data in order to obtain the FA and MD scalar measures of diffusivity. | 33 |
| 3.6 | Resting-state fMRI preprocessing workflow. | 35 |
| 4.1 | Flow diagram of the VBM based feature extraction process from the subjects GM segmentation volumes. | 41 |
| 4.2 | Flow diagram of the feature extraction and ensemble of classifiers. | 42 |
| 4.3 | Pipeline of the computation of the deformation scalar measures. | 44 |
| 4.4 | Pipeline of the calculation of the correlation volumes. | 44 |
| 4.5 | A summary of the pipelines of the feature selection methods. Each path from left to right specifies a feature selection process tested in the computational experiments. | 45 |
| 4.6 | FA and MD maps of one subject. | 47 |
| 4.7 | Image processing pipeline for the spatial normalization of the FA and MD scalar maps from DWI data. | 48 |
| 4.8 | Evolutionary wrapper feature selection computational pipeline. | 49 |

| | | |
|------|---|----|
| 5.1 | Flow chart of the 10-fold cross validation procedure followed in the experiments reported in this thesis. | 54 |
| 6.1 | Pipeline for circularity free validation of supervised classification with supervised feature selection. | 71 |
| 6.2 | Average and standard deviation of Accuracy, Precision, Recall (Sensitivity) and ROC area over the 10-fold cross-tests results of a linear SVM trained to minimize the error rate, features selected using Pearson's correlation . Note that the cross-validation population subsets were all the same for all the experiments. . . | 81 |
| 6.3 | Average and standard deviation of Accuracy, Precision, Sensitivity and ROC area over the 10-fold cross-tests results of linear SVM trained to minimize error rate, features selected using Bhattacharyya's distance . Note that the cross-validation population subsets were all the same for all the experiments. | 82 |
| 6.4 | Mean and standard deviation of Accuracy, Precision, Recall (Sensitivity) and ROC area over the 10-fold cross-tests results of linear SVM trained minimizing error rate, features selected using Welch's t-test . Note that the cross-validation population subsets were all the same for all the experiments. | 83 |
| 6.5 | Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson's correlation (left), Bhattacharyya's distance (middle) and Welch's t-test (right) computed on the modulated GM (<i>modgm</i>). | 84 |
| 6.6 | Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson's correlation (left), Bhattacharyya's distance (middle) and Welch's t-test (right) computed on the Jacobian maps (<i>jacs</i>). | 85 |
| 6.7 | Slices of the MNI standard template showing the <i>a priori</i> maps of affected regions in the left temporal area (left), corresponding to mild AD (middle) and to normal AD (right) used for ensemble decision. | 86 |
| 6.8 | Slices of the MNI standard template where the 20 best training AUC (left) and F1-score (right) ROIs are coloured. | 86 |
| 6.9 | Slices of the MNI standard template where the ROIs selected by ELM (left) and BDC (right) are coloured. | 87 |
| 6.10 | Slices of the MNI standard template where the ROIs selected by RF (left) and HERF (right) are coloured. | 88 |
| 6.11 | Voxel sites (red dots) corresponding to FA features selected setting a 99.5% percentile of the e.c.d.f. of the absolute value of the Pearson's correlation of FA data with the class variable. | 88 |
| 6.12 | Average accuracy of AD patient classification obtained with SVM and 1-NN classifiers varying number of features for FA and MD maps computed from DWI data. | 90 |

| | | |
|------|--|-----|
| 6.13 | Average sensitivity of AD patient classification obtained with SVM and 1-NN classifiers varying number of features for FA and MD maps computed from DWI data. | 91 |
| 6.14 | Average specificity obtained with SVM and 1-NN classifiers varying number of features for FA and MD data. | 92 |
| 6.15 | SPM results showing GM differences between patients and controls in the VBM of the Myotonic Dystrophy database. | 93 |
| 6.16 | SPM results shown as the most significant VBM clusters found in the Myotonic Dystrophy database superimposed on the GM averaged image (obtained from subjects under study, 8 mm FWHM isotropic Gaussian kernel). | 96 |
| 6.17 | Classification performance using TPF-GSR ReHo data and the Bhattacharyya distance. | 106 |
| 6.18 | Classification performance using GSR ReHo data and the Bhattacharyya distance. | 107 |
| 6.19 | Classification performance using TPF-GSR ReHo data and the Pearson correlation. | 108 |
| 6.20 | Voxels selected from the TPF-GSR ReHo data using Bhattacharyya distance thresholded at 95%. The colormap indicates the voxel selection frequency during cross-validation. | 108 |
| 6.21 | Voxels selected from the GSR ReHo data using Bhattacharyya distance thresholded at 95%. The colormap indicates the voxel selection frequency during cross-validation. | 109 |
| 6.22 | Voxels selected from the TPF-GSR ReHo data using Pearson correlation thresholded at 90%. The colormap indicates the selection frequency during cross-validation of each voxel. | 110 |
| 6.23 | Effect of the number of ELM and the number of hidden units in the (a) Accuracy, (b) Sensitivity, (c) Specificity of ELM ensemble for the ReHo functional connectivity map. | 113 |
| 6.24 | Effect of the number of ELM and the number of hidden units in the (a) Accuracy, (b) Sensitivity, (c) Specificity of ELM ensemble for the ALFF functional connectivity map. | 114 |
| 6.25 | ReHo features localization: (a) 2113 features selected by EW over the raw f.c.m, (b) Pearson's 2113 features, (c) 500 selected by EW over Pearson's features. . . | 117 |
| 6.26 | ALFF features localization: (a) 2198 features selected with EW over the f.c.m., (b) Pearson's 2198 features, (c) 500 features selected by EW over the Pearson's features. | 118 |
| A.1 | DM1 patients excluded from the study: sagittal T1 and transversal T2 images. 1.a. and 1.b. Grade 5 ventricular dilation (44-year-old male with 333 CTG and 41-year-old female with 667 CTG respectively), 1.c. grade 3 ventricular dilation (28-year-old female with 500 CTG), 2. Frontal hyperostosis (36-year-old female with 400 CTG) and 3. Frontal calcification (38-year-old female with 833 CTG). | 131 |

| | | |
|-----|--|-----|
| A.2 | Patients with white matter lesions; transversal T2 images: 1) 52-year-old female with 333 CTG and both subcortical and deep focal lesions. 2) 44-year-old male with 667 CTG and confluent subcortical lesions. | 132 |
|-----|--|-----|

List of Tables

| | | |
|-----|---|----|
| 4.1 | Pearson's correlation significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the feature selection to the OASIS data. | 46 |
| 4.2 | Bhattacharyya significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the deformation feature selection to the OASIS data. | 46 |
| 4.3 | Welch's t-test significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the deformation feature selection to the OASIS data. | 46 |
| 6.1 | Classification results with a linear kernel (lk) and a non-linear kernel (nlk). No covariates have been taken into account in the VBM. The values of $\gamma = (2\sigma^2)^{-1}$ for non linear kernel were 0.5, 0.031, 0.0078 for each feature extraction process, respectively. | 73 |
| 6.2 | Classification results with a linear kernel (lk) and a non-linear kernel (nlk). The normalized brain volume (nWBV) covariate has been taken into account in the VBM. The values of γ for nlk were 0.5, 2.7, 0.004 for GM proportion, MSD and VV features, respectively. | 73 |
| 6.3 | Classification results of 40 AD patients vs. 49 control subjects with the SVM and a RBF kernel, 9 possible outliers were taken out from the AD patients subset. | 74 |
| 6.4 | Classification results with a BP network with resilient back-propagation. Mean (Standard deviation) of 10 cross-validations. | 75 |
| 6.5 | Classification results with a RBF network. Mean (Standard deviation) of 10-fold cross-validation. | 75 |
| 6.6 | Classification results with a PNN network. Mean (Standard deviation) of 10-fold cross-validation. | 75 |
| 6.7 | Classification results with a LVQ1 network. Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 150 epochs, goal: 0.10 and learning rate: 0.010. Mean (Standard deviation) of 10-fold cross-validation. | 76 |

| | | |
|------|---|----|
| 6.8 | Classification results with a LVQ2 network. Network training parameters: <i>MSD</i> : 200 epochs, goal: 0.01 and learning rate: 0.01; <i>VV</i> : 50 epochs, goal: 0.01 and learning rate: 0.005. Mean (Standard deviation) of 10-fold cross-validation. . . | 76 |
| 6.9 | Majority voting classification results with linear kernel (lk) and non-linear kernel (nlk) SVM built independently for each VBM cluster. | 76 |
| 6.10 | Weighted individual SVM per cluster classification results. The value of the RBF kernels for the nonlinear (nlk) classifiers were searched for the best fit to the training set. | 76 |
| 6.11 | Diverse AdaBoost SVM classification results. | 76 |
| 6.12 | Results using linear SVM on DM features obtained from the 0.995 percentile of the Spearman's correlation. | 77 |
| 6.13 | Results using linear SVM on DM features obtained from the 0.999 percentile of the Spearman's correlation. | 78 |
| 6.14 | Results using linear SVM on normalized JD features obtained from the 0.995 percentile of the Spearman's correlation measures. | 78 |
| 6.15 | Results using linear SVM on DM features over the 0.995 percentile of the Pearson's correlation saliency measures. | 78 |
| 6.16 | Results using linear SVM on JD features over the 0.990 percentile of the Pearson's correlation saliency measures. | 79 |
| 6.17 | Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile. Note that the cross-validation population subsets were all the same for all the experiments. | 84 |
| 6.18 | Accuracy, sensitivity, specificity and ROC area of the leave-one-out cross-validation of SVM classification results. | 85 |
| 6.19 | Ensembles classification performance. | 87 |
| 6.20 | SVM classification results of Myotonic Dystrophy for MSD features extracted from clusters found by t-test VBM, FWE=0.05. | 94 |
| 6.21 | SVM classification results of Myotonic Dystrophy of type I for VV features extracted from clusters found by t-test VBM, FWE=0.05. | 94 |
| 6.22 | SVM classification results of Myotonic Dystrophy of type I for MSD features extracted from clusters found by F-test VBM, FWE=0.05. | 95 |
| 6.23 | SVM classification results of Myotonic Dystrophy of type I for VV features extracted from clusters found by F-test VBM, FWE=0.05. | 95 |
| 6.24 | Databases considered, percentile on the correlation distribution and size of the feature vectors. | 98 |
| 6.25 | Average 10-fold cross-validation classification performance results for diverse databases and classifiers on the FA features of the NAMIC database. | 99 |

| | | |
|------|---|-----|
| 6.26 | Average 10-fold cross-validation classification performance results for diverse databases and classifiers on the FA features of the NAMIC database. | 100 |
| 6.27 | Maximum accuracies and variances across feature selection thresholds for each experiment on the TPF-GSR data. | 103 |
| 6.28 | Maximum accuracies and variances across feature selection thresholds for each experiment on the GSR data. | 103 |
| 6.29 | Maximum accuracies and variances across feature selection thresholds for each experiment on the TPF data. | 104 |
| 6.30 | Maximum accuracies and variances across feature selection thresholds for each experiment on the data without TPF and GSR. | 104 |
| 6.31 | Cortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/), the Harvard-Oxford Cortical Structural Atlas and a cluster threshold of 95%. | 105 |
| 6.32 | (cont.) Cortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/), the Harvard-Oxford Cortical Structural Atlas and a cluster threshold of 95%. | 109 |
| 6.33 | Subcortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/), the Harvard-Oxford Subcortical Structural Atlas and a cluster threshold of 95%. | 110 |
| 6.34 | Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over ReHo functional connectivity map. Bold entries are above 80%. | 111 |
| 6.35 | Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over ALFF functional connectivity map. Bold entries are above 80%. | 111 |
| 6.36 | Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over fALFF functional connectivity map. Bold entries are above 80%. | 111 |
| 6.37 | Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over VMHC functional connectivity map. Bold entries are above 80%. | 111 |
| 6.38 | Feature vector size for the selected functional connectivity maps. | 115 |
| 6.39 | 10-fold cross-validation of Accuracy of V-ELM over the different feature selection methods. By columns, evolutionary wrapper over raw functional connectivity map (f.c.m.). | 115 |

| | | |
|------|---|-----|
| 6.40 | Effect on the ensemble size and number of hidden units on the cross-validation Accuracy for the different feature selection and feature sizes. x/x means the feature vector size per functional connectivity map. Comparison of EW over f.c.m., Pearson selection and EW over Pearson's features | 116 |
| A.1 | Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best). | 127 |
| A.2 | Age and Diagnosis Characteristics of the Data Set. *Male/Female | 127 |
| A.3 | Summary of subject demographics and dementia status. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best). Handedness is a continuous measure from 1 (right) to 0 (left). | 128 |
| A.4 | Summary of subject demographics. | 129 |
| A.5 | Summary of control subjects diagnoses. | 129 |
| A.6 | Summary of patient subjects diagnoses. | 129 |
| A.7 | Summary of subject demographics and MD1 status. ¹ Muscular Impairment Rating Scale. | 130 |

Chapter 1

Introduction

This chapter provides the general motivation of the thesis in section 1.1, an overview of the thesis contents and contributions in section 1.2, a listing of the publications obtained as a result of the research associated with this thesis studies in section 1.3 and the structure of the thesis in section 1.4.

1.1 Motivation

Machine learning offers a new paradigm for Neuroscience. In this new paradigm one of the emphasis is in prediction instead of hypothesis tests leading to findings related to the subject of study. Thus, this thesis is an exploration of this paradigm effect on discoveries that can be made using machine learning on data produced by the neuroscientific community. Several diseases and Magnetic Resonance Imaging (MRI) modalities are objects of research within this thesis. There are some salient aspects of the contribution of machine learning to this realm:

1. Setting a computational and experimental framework for the discovery and validation of hypothesis. It emphasises on the reproducibility of results by independent researchers, implying the publication of datasets and of code for the realization of the experiments. Although tuning the experiments parameters will remain an art and despite careful descriptions reproducibility still poses difficulties.
2. The development of Computer Aided Diagnostic (CAD) systems, that may alleviate the increasing burden of data processing to the medical experts. CAD systems are not intended to substitute the human judgement, but to help perform initial triages to focus the human work on cases that are either dubious or have a great cost. Considering longitudinal studies and treatment of patients, CAD systems may be able to detect and call the attention to early stages, prodromal cases.
3. The discovery of effects (image biomarkers) is indirect. In this thesis we look for voxels containing discriminant information and then we validate the classification system based

on these features, and the classification results tell us about their predictive power. The question then is posed as follows: if we can make predictions based on these data, could that mean that the corresponding anatomical locations, functional networks, etc, have some causal relation to the disease? Exploring this question requires careful collaboration between medical experts and computer scientists. In the thesis we have had few opportunities to perform such interactions, however we are far to have a fluid relationship, and the whole potential of this approach still remains unfulfilled.

The realization of machine learning experiments can not be made in the void. Data do not come ready for the direct application of the classifiers, much unlike the conventional experiments reported in the literature where authors download specific experimental (small size) databases and perform exhaustive computational experiments for parameter tuning and validation. First, we are dealing with Big Data, meaning that information for each experiment ranges in the tens or hundreds of gigabytes. In some experiments, dealing with the data to perform the computations is a difficult task, moving us to the time of the early computers that needed to slice matrices, store and reload parts of them, and proceed in a very slow manner. Second, data is not clean, it needs some pre-processing in order to have meaningful results. Fortunately, the state of the art is well established in many cases and the corresponding pipelines, though laborious, are available in free software bundles (which are described in Appendix B).

Historical perspective Every research work has a context and an evolution. We think that it is relevant to recall a draft of this history to highlight the value of the work done. Starting studies on medical image processing in the GIC research group precede the arrival of the author, but were sparse and directed to basic filtering and segmentation procedures. Upon arrival of Maite García-Sebastián and the author, the group started to consider the construction of CAD systems for Alzheimer's Disease, on the basis of public available data at the time, the OASIS database. Downloading the database was time consuming and risking the prosecution by the system administration staff. We started to work on a reduced and balanced database, whose feature datasets are available in the GIC wiki. Other branch of work, that coincides with a internship of the author at the FMRI and the arrival of Maite Termenón, dealt with another public database for schizophrenia containing diffusion data (NAMIC). The group has been gathering experience on the pre-processing of the data through lengthy experiments. The studies call the attention of the Unidad de Neurología of the Hospital Donostia (now Biodonostia) working on Myotonic Dystrophy, starting a collaboration between medical and computer scientists carried out mostly by Maite García. Next contact was the Unidad de Imagen Funcional of the Universidad Jaume I, which provided data with cocaine addicted subjects that has been mostly carried out by Maite Termenón. At some time, people from the McLean Hospital and Harvard Medical School, Boston, provided fMRI data for the discrimination of Schizophrenia patients with and without auditory hallucinations, which has been treated by Darya Chyzhyk. Another source of data has been the Alzheimer's Disease (AD) and Bipolar Disorder dataset provided by Hospital

Santiago de Vitoria, processed by Maite Termenón. Still, the author has resorted to the public databases (OASIS and COBRE) because their protocols are well standardized and the results are appearing in the literature. Overall, the group has been growing in expertise and understanding of the available tools, dealing with errors in data capture and other issues, without a scheduled planning. The role of the author of this thesis has been to technical computation support, e.g. building and configuring a computer cluster, to explore new processing pipelines and to conduct most of the experiments that are reported in the thesis, if not said otherwise in the correspondent experiment description.

Magnetic Resonance Imaging Data Magnetic Resonance Imaging (MRI) is a medical imaging process based on the phenomenon known as Nuclear Magnetic Resonance (NMR). It has already widespread acceptance for a large variety of medical explorations. It is one of the most commonly used techniques in radiology to visualize the body soft tissues with great contrast, including the brain. Some reasons for this great interest are:

- It is not an invasive imaging procedure, it does not produce ionization and it is, thus, safe to perform repetitive and lengthy observations.
- It is very ductile, the change of capture parameters allows to obtain very diverse image modalities to measure diverse effects, from anatomical imaging to functional, motion, diffusion, etc. The field is in constant evolution, often producing new sequences with new applications.

Thesis structure Figure 1.1 provides a visualization of the thesis structure and the topics covered by it. The image gives a direct impression of the complexity and comprehensiveness of the experimental work performed in this thesis. The different layers from left to right (bottom up) correspond to the main parts in the thesis, summarizing their contents.

The baseline layer is the data modalities related to applications to different diseases, where we have been working with anatomical data for Alzheimer’s Disease and Myotonic Dystrophy, diffusion and resting state functional data for Schizophrenia.

The next layer corresponds to the spectrum of pre-processing procedures and scalar measures extracted from the multivariate data for further feature selection and classification. Pre-processing involves spatial normalization by image registration and specific denoising procedures for each data modality. From the deformation fields of anatomical data registration we have extracted scalar measures such as the deformation Jacobian determinant, from diffusion data we have extracted Fractional Anisotropy (FA) and Mean Diffusivity (MD) measures, and from functional data we have extracted functional activity measures such as the Regional Homogeneity (ReHo) and others.

The third layer corresponds to the feature selection procedures that we have tested. Feature selection in this thesis means the selection of voxel sites that have some discriminant power

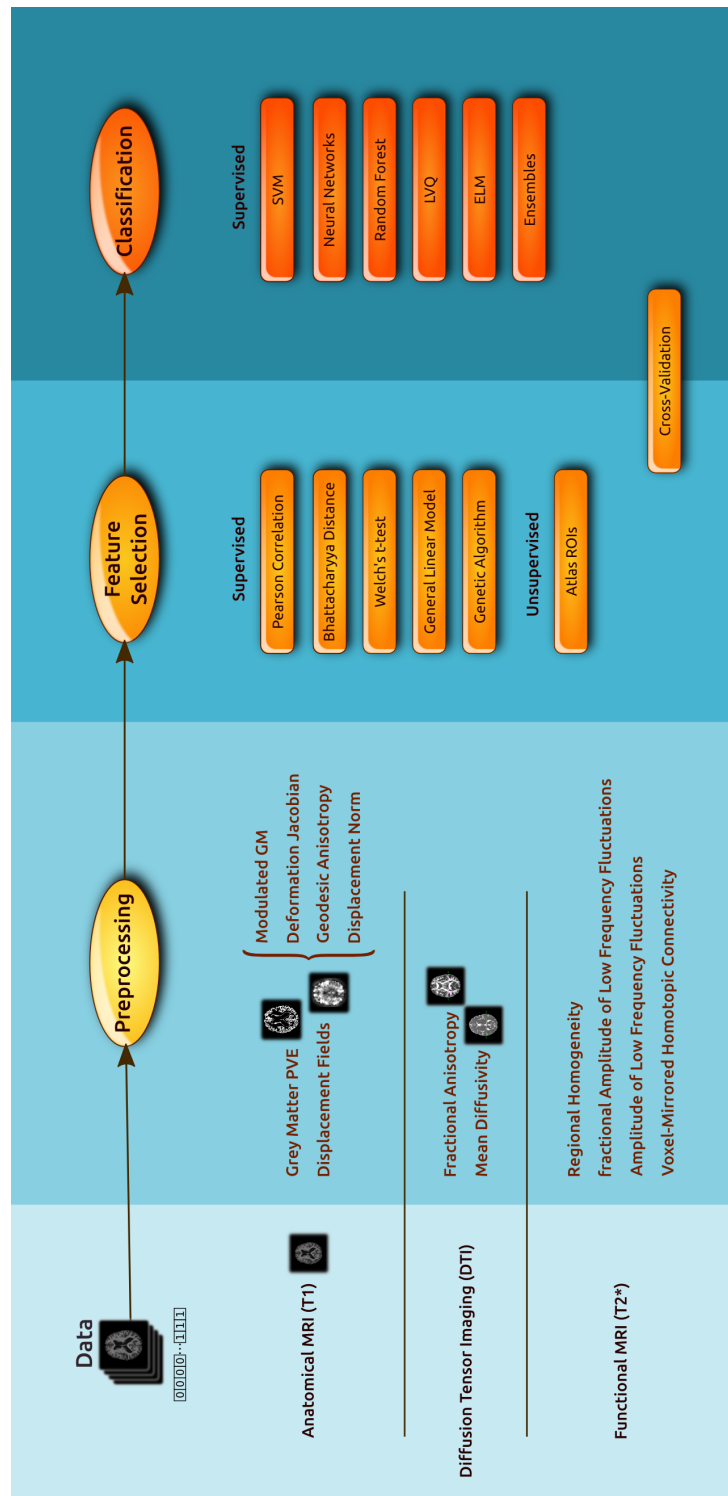


Figure 1.1: Structure of the thesis

and can be used as features for the classifiers. These features have the additional meaning of image biomarkers if we take into account their anatomical localization. Supervised feature selection methods employ information about the subject class, i.e. patient vs. control, in the computation of the voxel saliency measure by means some correlation or class discrimination measure, such as the Pearson's correlation coefficient, the Bhattacharyya's univariate Gaussian distance, the Welch's t-test and the GLM performed by the VBM process. We have tested also an evolutionary wrapper feature selection. Unsupervised feature selection means the features are selected without regarding class information. We have considered the ROIs provided by the AAL atlas for the extraction of such features.

The last layer of the computational processes corresponds to the classification building algorithms that have been tested in different computational experiments reported in the thesis. The most frequently used classifier is the SVM which has turned into a *de facto* standard in the application of machine learning to neuroscientific data. However, SVM is not easily tuned and has some shortcomings dealing with specific databases, such as imbalanced databases. Hence, in diverse cases we have tested alternative classifiers, such as the Extreme Learning Machine (ELM) or varieties of Ensembles of Classifiers.

A subtle interaction between the feature selection processes and the training of classifiers may happen in cross-validation schemes, specially in the case of supervised feature selection where special care must be taken to avoid circularity effects.

The combination of all the elements reflected in this figure has produced a rich variety of computational experiments that are reported in the experimental chapter of the thesis, hence its relatively large size.

1.2 Summary of the Thesis Contributions

The main contributions of the thesis are in the methodology of the application of machine learning to neuroscience data, specifically MRI of diverse modalities. There are also some operational contributions steaming from the actual developments needed for the realization of the computational experiments.

Methodological contributions

- We have performed extensive experimental demonstration of the application of machine learning techniques to neuroscience data, with the two-fold aim of demonstrating CAD systems performance and obtaining relevant image biomarkers. We have tested the approach on data of several diseases and diverse MRI modalities.
- We have proposed and tested a feature selection general approach based on voxel saliency, which can be applied to a variety of scalar maps derived from diverse data modalities.

- We have proposed and tested a variety of voxel saliency measures for feature selection, performing experimental evaluation of their relative merits.
- We have found image biomarkers for several diseases that are in good agreement with the medical expertise, thus validating the approach from the medical practice point of view.
- We have contributed a new evolutionary wrapper feature selection based on ELM which provides compact features and a measure of their relevance.
- We have shown the value of specific scalar measures as features for CAD systems
 - Scalar deformation measures
 - Diffusivity scalar measures: FA and MD
 - Functional activity measures: ReHo, ALFF, fALFF, VMHC.

Classification algorithms We have contributed new variations of AdaBoost for SVM and ensembles of independent SVM, which provide competitive results with state of the art algorithms.

Operational contributions We have dealt with a wealth of software and data resources. The appendices of thesis contain descriptions of

- the software resources available to the researchers in the neuroscience domain, with special emphasis in free open source software. The descriptions include tips that may be useful.
- the databases employed in the thesis.

Besides the contents of the thesis, we have published in the website of the research group some of the datasets and codes. Specifically, we have made public the feature datasets obtained after pre-processing and feature selection; this resource is available to researchers interested in testing their classification algorithms on these problems. Besides, software scripts specifying computational pipelines have also been made public.

1.3 Publications

1. Darya Chyzyk, Alexandre Savio, Manuel Graña, “Computer Aided Diagnosis of Schizophrenia by Ensembles of ELM on resting state fMRI data”, Applied Soft Computing, (submitted)
2. Borja Ayerdi, Alexandre Savio, Manuel Graña, “Meta-ensembles of classifiers for Alzheimer’s disease detection using independent ROI features”, editors: J.M. Ferrández et al., IWINAC 2013, Part II, vol. 7931 of LNCS, pages 122-130. Springer, Heidelberg (2013).

3. Alexandre Savio, Manuel Graña, “An ensemble of classifiers guided by the AAL brain atlas for Alzheimer’s disease detection”, I. Rojas, G. Joya, and J. Cabestany (Eds.), IWANN 2013, Part II, vol. 7903 of LNCS, pages 107-114, Springer, Heidelberg (2013).
4. Alexandre Savio, Manuel Graña, “Deformation based feature selection for Computer Aided Diagnosis of Alzheimer’s Disease”, Expert Systems with Applications, available online 20 September 2012, vol. 40, issue 5, April 2013, pages 1619–1628. JCR (2012): 1.854; 5-Year JCR: 2.339; Eigfactor: 0.04223; AI score: 0.497.
5. Darya Chyzyk, Alexandre Savio and Manuel Graña, “Evolutionary ELM wrapper feature selection for Alzheimer’s disease CAD on anatomical brain MRI”, ELM 2012, Neurocomputing (in press), JCR (2012) 1.634 (T1, Q2 CSAI).
6. Alexandre Savio, Manuel Graña, “Supervised classification using deformation-based features for Alzheimer’s disease detection on the OASIS cross-sectional database”, Advances in Knowledge-Based and Intelligent Information and Engineering Systems. Frontiers in Artificial Intelligence and Applications (FAIA) series, vol. 243, pages 2191 - 2200, 2012, editors: Manuel Graña, Carlos Toro, Jorge Posada, Robert J. Howlett and Lakhmi C. Jain.
7. S. Jbabdi, S.N. Sotiropoulos, A. Savio, M. Graña, T.E.J. Behrens, “Model-based analysis of multi-shell diffusion MR data for tractography: How to get over fitting problems”, Magnetic Resonance in Medicine, vol. 68, issue 6, pages 1846–1855, December 2012, DOI: 10.1002/mrm.24204. JCR (2012): 3.267; 5-Year JCR: 3.948; Eigfactor: 0.03755; AI score: 1.159.
8. A. Bourisly, A. Savio, M. Graña, “A Meta-Analysis of fMRI Biomarkers for Alzheimer’s Disease”, OHBM 2012, June 10-14, 2012 Beijing, China.
9. Darya Chyzyk, Manuel Graña, Alexandre Savio, Josu Maiora, “Hybrid Dendritic Computing with Kernel-LICA applied to Alzheimer’s Disease detection in MRI”, Neurocomputing, 2012, vol. 75, issue 1, pages 72-77. JCR (2012): 1.634; 5-Year JCR: 1.811; Eigfactor: 0.01694; AI score: 0.535.
10. A. Savio, A. Bourisly, M. Graña, “On Current Biomarkers for Alzheimer’s Disease detected by Machine Learning techniques”, ISMRM SCIENTIFIC WORKSHOP: Neuroimaging Biomarkers of Psychiatric Disorders: What are They?, October 13-16, 2011, Montabaur, Germany.
11. M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J. M. Pérez, A. Besga, “Computer Aided Diagnosis system for Alzheimer Disease using brain Diffusion Tensor Imaging features selected by Pearson’s correlation”, Neuroscience letters, Volume 502, Issue 3, September 20, 2011, pages 225-229, on line. JCR (2012): 2.026; 5-Year JCR: 2.146; Eigfactor: 0.05221; AI score: 0.618.

12. Darya Chyzhyk, Ann K. Shinn, and Manuel Graña, “Exploration of LICA Detections in Resting State fMRI”, 4th. International work-Conference on the Interplay between Natural and Artificial Computation (IWINAC 2011), La Palma, Canary Islands, Spain. 30 May - 3 June, 2011. Published in: IWINAC’11 Proceedings of the 4th international conference on Interplay between natural and artificial computation: new challenges on bioinspired applications - Volume Part II, pages 104-111.
13. Alexandre Savio, Manuel Graña, Jorge Villanúa, “Deformation based features for Alzheimer’s disease detection with linear SVM”, Hybrid Artificial Intelligence Systems, 6th International Conference (HAIS 2011), HAIS 2011, Part II, vol. 6679 of LNAI proceedings, pages 336-343, Springer, Heidelberg (2011).
14. Alexandre Savio, Juliette Charpentier, Maite Termenón, Ann K. Shinn, Manuel Graña, “Neural classifiers for schizophrenia diagnostic support on diffusion imaging data”, Neural Network World 20(7), pages 935-949. JCR (2012): 0.362; 5-Year JCR: 0.381; Eigfactor: 0.00033; AI score: 0.082.
15. Darya Chyzhyk, Maite Termenón, and Alexandre Savio, “A Comparison of VBM results by SPM, ICA and LICA”, Hybrid Artificial Intelligent Systems, Part II, editors: Emilio Corchado, Manuel Graña, Alexandre Savio, vol. 6077 of (LNCS/LNAI), Springer-Verlag Berlin Heidelberg, 2010, pages 429 - 435 vol. 2.
16. Alexandre Savio, Maite García-Sebastián, Darya Chyzhyk, Carmen Hernández, Manuel Graña, Andone Sistiaga, Adolfo Lopez de Munain, Jorge Villanua, “Neurocognitive disorder detection based on Feature Vectors extracted from VBM analysis of structural MRI”, Computers in Biology and Medicine 41 (2011), pages 600-610. JCR (2012): 1.162; 5-Year JCR: 1.359; Eigfactor: 0.00422; AI score: 0.356.
17. Alexandre Savio, Maite García-Sebastián, Andone Sistiaga, Darya Chyzhyk, Esther Fernández, Fermín Moreno, Elsa Fernández, Manuel Graña, Jorge Villanúa, Adolfo López de Munain, “Machine Learning Approach for Myotonic Dystrophy Diagnostic Support from MRI”, in Juan M. Górriz, Elmar W. Lang and Javier Ramírez (Eds), Recent Advances in Biomedical Signal Processing, 93 - 101 2010 Bentham Science Publishers.
18. Manuel Graña, Alexandre M. Savio, Maite García-Sebastián, Elsa Fernández, “A Lattice Computing approach for On-line fMRI analysis”, Image and Vision Computing 28(7), pages 1155-1161, July 2010. JCR (2012): 1.959; 5-Year JCR: 1.952; Eigfactor: 0.01024; AI score: 0.851.
19. Darya Chyzhyk and Alexandre Savio, “Feature extraction from structural MRI images based on VBM: data from OASIS database”, Technical Report GIC-UPV-EHU-RR-2010-10-14, Grupo de Inteligencia Computacional UPV/EHU, 2010.

20. Maite García-Sebastián, Alexandre Savio, Manuel Graña, Jorge Villanúa, “On the use of Morphometry based features for Alzheimer’s Disease detection on MRI”, *Bio-Inspired Systems: Computational and Ambient Intelligence / IWANN 2009 (Part I)*, Joan Cabestany, Francisco Sandoval, Alberto Prieto, Juan M. Corchado (Editors), LNCS 5517, Springer Verlag 2009, pages 957-964; ISBN: 978-3-642-02477-1; DOI: 10.1007/978-3-642-02478-8_120.
21. Alexandre Savio, Maite García-Sebastián, Manuel Graña, Jorge Villanúa, “Results of an Adaboost approach on Alzheimer’s Disease detection on MRI”, *Methods and Models in Artificial and Natural Computation / IWINAC 2009 (Part II)*, José Mira, José Manuel Fernández, José R. Álvarez, Félix de la Paz, F. Javier Toledo (Editors), LNCS 5602, Springer Verlag 2009, pages 114-123, ISBN: 978-3-642-02266-1, DOI: 10.1007/978-3-642-02267-8_13.
22. Alexandre Savio, Maite García-Sebastián, Carmen Hernández, Manuel Graña, Jorge Villanúa, “Classification Results of Artificial Neural Networks for Alzheimer’s Disease Detection; Intelligent Data Engineering and Automated Learning - IDEAL 2009, LNCS 5788, Springer Verlag 2009, pages 641-648, ISBN: 978-3-642-04393-2, DOI: 10.1007/978-3-642-04394-9_78.
23. Maite García-Sebastián, Alexandre M. Savio, M. Graña, “Comments on an Evolutionary Intensity Inhomogeneity Correction Algorithm”, *Proceedings CEC 2008*, pages 4146-4150, ISBN: 978-1-4244-1822-0.
24. Alexandre Manhães Savio, “Capítulo de libro en las Actas de las II Jornadas de Inteligencia Computacional” - JIC’08, Ramón Moreno Jiménez, Alexandre Manhães Savio (Editors), Servicio editorial de la UPV/EHU, ISBN: 978-84-296-2316-1.

1.4 Contents of the Thesis

The contents of the thesis are organized as follows:

- Chapter 2 contains a detailed medical background information about the diseases that we have been considered along the thesis: Alzheimer’s Disease, Myotonic Dystrophy and Schizophrenia. We provide the definition, some medical relevant facts, and the state of the art of classification approaches to build CAD systems for their detection.
- Chapter 3 reviews the pre-processing pipelines for the diverse MRI data modalities that have been treated. Pre-processing is a critical step in the overall process with many specificities for each modality. Besides, the chapter provides the definition of some scalar maps that are used later for feature selection and classification, such as diffusivity measures and functional activity maps.

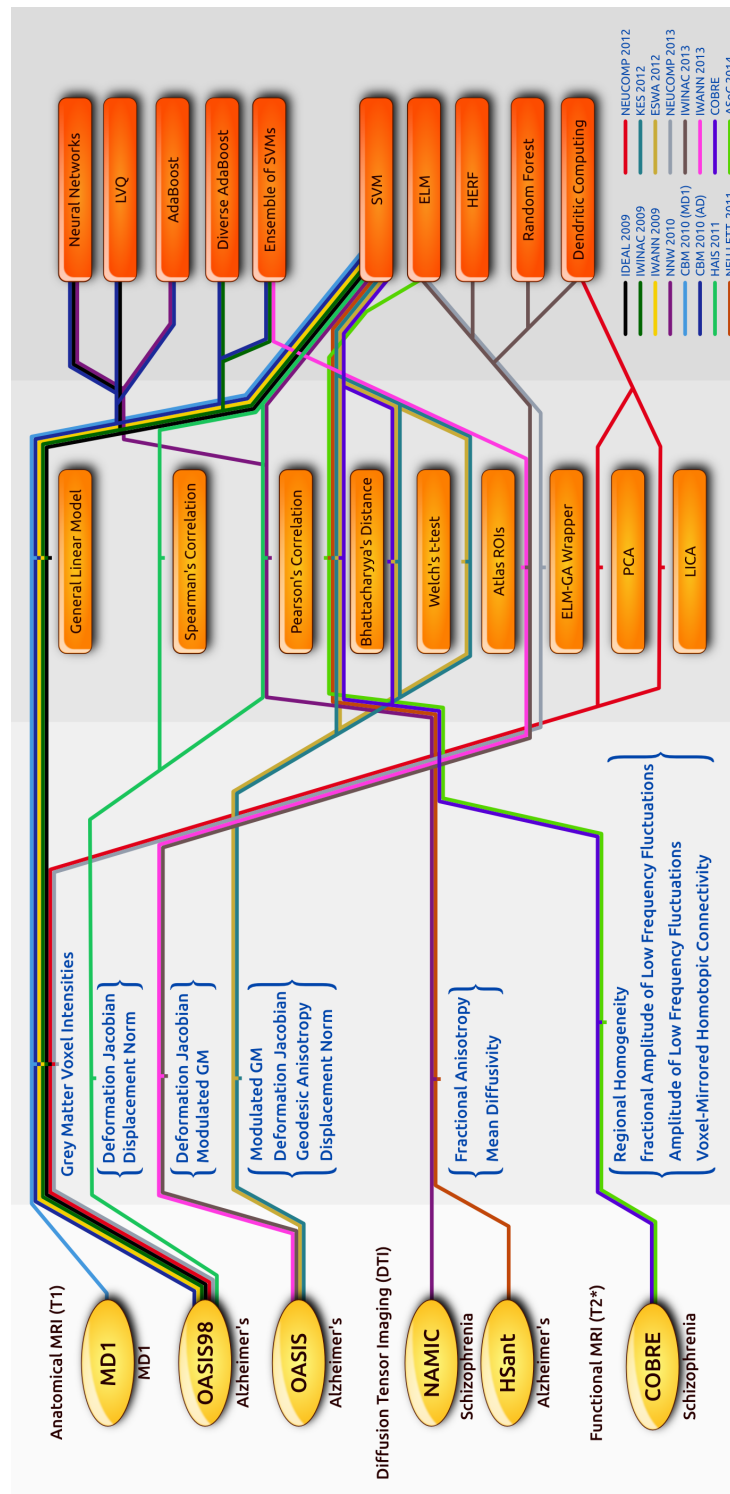


Figure 1.2: Specification of the classification experiments performed during this thesis with colleagues and collaborators.

- Chapter 4 provides a definition of feature selection as performed in the thesis. We have followed the general practice of computing voxel relevance measures relative to the class, so that our general approach to feature selection is supervised and prone to circularity effects if not performed carefully.
- Chapter 5 provides the definition of several classification algorithms used in the thesis, from bio-inspired to statistical or lattice computing based approaches.
- Chapter 6 provides the computational experiment results of the studies developed in the thesis. Some as personal work, others as collaboration with Maite Termenón, Borja Ayerdi and Darya Chyzhyk. It is the culmination of the vast amount of experimental work performed. The chapter is very long, but we have not found a reasonable way to split it. It is structured around the diseases and the databases that we have used to build and test CAD systems for them.
- Chapter 7 provides the conclusions of the thesis. Notice that most chapters do not have conclusions, as they are exposition of tools or current practices leading to the experiment and conclusions chapters.

Chapter 2

Medical Background and Related Studies

This chapter briefly describes the diseases related to the studies in this thesis and similar studies found in the literature. This information serves as motivation for the work described in this thesis.

Machine learning methods have become very popular to classify functional or structural brain images to discriminate them into two classes: normal or a specific neurodegenerative disorder [52]. The development of automated detection procedures based in Magnetic Resonance Imaging (MRI) and other medical imaging techniques [59] is of high interest in clinical medicine. It is important to note that these techniques are aimed to help clinicians with more statistical evidence for the diagnosis, it is not intended to substitute any other existing diagnosis procedure.

2.1 Alzheimer's Disease

Alzheimer's Disease (AD) is one of the most important causes of disability in the elderly and taking into account the increasing proportion of elderly in many populations the number of dementia patients will also rise [14]. AD is a progressive neurodegenerative disease showing gradual deterioration in cognition, function and behavior. Due to the socioeconomic importance of the disease in occidental countries there is a strong international effort focus in AD. This degenerative disorder presents a cognitive and behavioral impairment that interferes with the daily life of the individual and its social network, imposing a high economical and psychological cost. Although research has revealed a great deal about AD, with the exception of certain inherited forms of the disease, the causes of AD remain unknown. Early diagnosis is crucial to mitigate its effects. The diagnosis can be done after the exclusion of other forms of dementia but a definitive diagnosis can only be made after a post-mortem study of the brain tissue. Ante-mortem approaches for diagnosis are under development, but they require neuropathologic confirmation

of the characteristic amyloid plaques and neurofibrillary tangles [150]. There has been a strong effort directed to the assessment of medical imaging biomarkers identifying the disease at various stages [101, 112, 114, 145]. In addition, the pharmaceutical companies have already recognized that imaging techniques especially MRI and Positron Emission Tomography (PET) provide "surrogate" information concerning the pattern and rate of neurodegeneration, which can be used to monitor the effects of treatments which slow the progression of neurodegeneration. Therefore, there is high interest in the development of automated detection procedures based on MRI and other medical imaging techniques[59].

One of the current lines of research involves the application of machine learning algorithms to features extracted from brain MRI. It is important to note that these techniques are aimed to help clinicians with more statistical evidence for the diagnosis, it is not intended to substitute any other existing diagnosis procedure. In the early stages of AD brain atrophy may be subtle and spatially distributed over many brain regions, including the entorhinal cortex, the hippocampus, lateral and inferior temporal structures, as well as the anterior and posterior cingulate. Recently, two very important initiatives to build public available databases of clinical and imaging data related to AD to improve experimental research on biomarkers for this disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI) [149] and AddNeuroMed [149]. ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations. AddNeuroMed, a part of InnoMed (Innovative Medicines in Europe) is an Integrated Project funded by the European Union Sixth Framework program.

2.1.1 Related Studies

Computer Aided Diagnosis (CAD) systems apply machine learning techniques to build classifier systems that may help the clinician to come up with a diagnosis on the basis of available data. Most CAD systems applied to imaging data, consist of a spatial normalization (registration) of the images to a standard template, the extraction of features from the image and the classification of these features into the target classes.

Machine learning methods have become very popular to classify brain images to discriminate them into two classes: normal or a specific neurodegenerative disorder [52]. Support Vector Machines (SVM) [203] have become a standard for the construction of such CAD classifiers. SVMs either with linear [220, 116, 205] or non-linear [67, 126] kernels are used to discriminate AD patients from controls based on Positron Emission Tomography (PET), Single-Photon Emission Tomography (SPECT) functional volumes [48, 81, 88, 106, 165], anatomical MRI data [59, 84, 173, 117], Diffusion Tensor Imaging (DTI) [90] and studies that combine structural and functional information where anatomical MRI and PET volumes are used [66]. In some works on functional imaging obtained by SPECT [171], findings in temporo-parietal region, posterior gyri, precune and temporal lobe were reported. Further studies on SPECT data have reported

high AD classification accuracy [47, 133, 157, 166, 182].

Many of the classification studies on the detection of AD were done over populations mixing men and women. However, it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [127]. This could lead to biases in the classification results.

Feature extraction There are different ways to extract features from MRI for SVM classification: based on morphometry methods [173, 59, 90, 69], based on ROIs (region of interest) [132], grey matter (GM) voxels in automated segmentation images [117] or cortical measures [210]. Work has also been reported on the selection of a small set of the most informative features for classification, such as the SVM-Recursive Feature Elimination [67], selection based on statistical tests [132, 164], the wavelet decomposition of the RAVENS maps [126], among others. It is important to note that these techniques are aimed to help clinicians with more statistical evidence for the diagnosis, it is not intended to substitute any other existing diagnosis procedure.

A study with 509 subjects drawn from ADNI database achieved high accuracies up to 81% of sensitivity and 95% of specificity using different feature selection approaches based on tissue segmentations, cortical and hippocampal measures and a hold out method for validation [57]. In [209] they combined regional segmentation and a manual hippocampus delineation features from AddNeuroMed were used for detection of AD patients with a sensitivity of 90% and a specificity of 94% using orthogonal partial least squares (OLPS).

Voxel-Based Morphometry Voxel-based Morphometry (VBM) is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images [12]. It has become a common tool for computational brain anatomy studies. The whole VBM procedure is detailed in Chapter 3. For instance, VBM has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of AD patients vs. control subjects [41, 179, 78]. Another work using ANNs and VBM for AD detection have been reported in [102], where a single three-layer, feed-forward ANN trained with a backpropagation algorithm was used as a classifier over a small set of unpublished proprietary MRI data. They perform data dimensionality reduction applying a Principal Component Analysis (PCA) to improve the classification performance.

Deformation-based Morphometry Deformation-based Morphometry (DBM) [84, 198] and Tensor-based Morphometry (TBM) [32, 128] to guide the feature extraction process. These morphometry methods analyse displacement vectors resulting from non-linear registration procedures with high number of degrees of freedom. A similar study [161] with 50 subjects obtained 92% of accuracy when discriminating AD subjects from healthy controls using features extracted from displacement fields and different classification methods with SVM, Bayesian statistics, and voting feature intervals (VFI). In addition, another study [198] obtained 83% of

accuracy using similar approaches to detect subjects with mild cognitive impairment. Although their results can not be reproduced, this work confirms that the approach that we follow is a promising area of research.

Ensemble methods Most published classification methods working on MRI data train a single classifier. However, it is challenging to train only a global classifier that can be robust enough to achieve good classification performance, mostly due to noise, small sample size and huge dimensionality of neuroimaging data. Some studies using ensemble of classifiers on brain anatomical MRI can be found in the literature. In [50] they propose a classification method via aggregation of regression algorithms fed with histograms of deformations generated from the Open Access Series of Imaging Studies (OASIS) database obtaining a 0.04 test error rate. Another study shows a local patch-based subspace ensemble method which builds multiple individual classifiers based on different subsets of local patches with the sparse representation-based classifier obtaining an accuracy of 90.8% on the ADNI database [131]. In [204] subsets of ranked features from neuroimaging data are used to in an ensemble of linear SVMs obtaining 0.94 of Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) when detecting AD patients vs. control subjects.

Diffusion imaging Multiple white matter (WM) abnormalities in AD have been found in DTI studies, such as differences between AD patients and healthy controls found in the splenium of the corpus callosum and in the temporal lobe [191], temporal lobe WM atrophy with WM microscopic damage in the thalamic radiations and in the corpus callosum [184]. Some studies on FA show effects in the bilateral posterior cingulate gyri and bilateral superior longitudinal fascicles [19]. However, there have been few whole-brain voxel-wise studies in which no a priori hypothesis was made regarding anatomical localization of white matter abnormalities [63].

2.2 Myotonic Dystrophy Type 1

Nowadays, there is much research effort devoted to the development of ways to provide automatized diagnostic support tools that may help the clinician to perform their work with additional assessment data to meet the ever increasing demands of primary attention of a rising population of patients with neurological disorders.

Myotonic dystrophy (MD) is a neuromuscular disease characterised by the presence of varying degrees of muscle weakness and myotonia, as well as a plethora of extra-muscular symptoms including the formation of cataracts, heart conduction problems and endocrine, ophthalmological and central nervous system (CNS) disorders. From a genetic point of view there are two types of MD: type 1, which involves mainly distal muscular atrophy and weakness (DM1); and type 2 (DM2 or PROMM), where there is no atrophy and a more proximal involvement. Here, we will only consider DM1 as this is the most common and best understood form. The unusual

molecular substrate of this disease results in considerable clinical variability, which means that different DM1 classifications have been proposed [96].

With an incidence of 1 in 8000 cases and a prevalence of between 69 and 90 cases per million, DM1, is the most common neuromuscular disease in humans [64, 148]. However, its prevalence may be much higher in some locations, such as the Canadian population of Saguenay-Lac St Jean [33] and in the province of Guipuzcoa in the Spanish Basque Country [60]. Indeed, despite the availability of molecular diagnosis-based genetic counselling since 1993, the prevalence in Guipuzcoa has increased, particularly at the expense of oligosymptomatic carriers of this disease.

DM1 is transmitted by autosomal dominant inheritance, which means that the risk of transmission to offspring is 50%, with no gender-based differences in the risk of transmission. Furthermore, there is an anticipation phenomenon whereby the disease onset may be earlier in subsequent generations.

The molecular basis of DM1 is related to the expansion of a repetitive sequence of the CTG triplet that may vary in length (5-35 repetitions) and that is stably transmitted across generations in normal individuals. However, in sufferers of DM1 the number of triplets increases to between 50 and up to 3000, and it displays unstable inter-generational transmission. This CTG triplet is located at the non-translated 3' end of the gene that codes for a protein kinase (DMPK) on the long-arm of chromosome 19 [37, 96]. The phenomenon of anticipation is related to the inter-generational increase of the pathological expansion, which is itself related to the sex of the transmitting parent. Generally speaking, in low ranges of pathological expansion (less than 150 repeats) the increase is greater when the transmitter is male, whereas above this number and in the rest of the spectrum the transmission is greater when the transmitter is female. This difference is one of the main reasons why almost all congenital cases are transmitted via the mother.

The mechanism by which the pathological expansion produces the clinical symptoms involves translation of the CTG triplets into CUG-type residues, which induce the nuclear build-up of RNAs for various genes that affect the splicing process of numerous proteins, thereby explaining the multisystemic character of DM1's symptoms [199].

Minimum or partial DM1: Characterised mainly by the presence of cataracts, with no or very few neuromuscular symptoms, and an onset after 50 years of age. Sufferers usually carry an expansion of between 35 and 100 repeats that are unstably transmitted across generations.

- Classic or adult DM1: Myotonia and progressive weakness generally present during adult life. Between 100 and 900 repeats.
- Childhood DM1: Shares many of its symptoms with the congenital form, although these are not apparent at birth.
- Congenital DM1: Symptoms are present from birth or even in uterus. These include respiratory failure, hypotonia and retarded development. This form usually involves more

than 1,000 repeats. Cases with larger expansions are unstable in both the somatic and germ cells, which means that more repeats result in greater instability [142].

Cognitive involvement in DM1 was noted in some of the earliest studies of this disease[6], particularly in congenital sufferers and individuals with early onset. Thus, as well as reduced intelligence levels, the very earliest descriptions from authors such as Stolba or Tomasen (1948) referred to patients lacking initiative or being apathetic, amongst other similar terms.

The clinical data available suggest that patients difficulties are focused on executive functions [105, 187], visuospatial/constructive abilities [136, 187], memory [168] and even on facial emotion recognition [213]. Although there is still no general consensus regarding the existence of DM1-related emotional and personality disorders[39, 214], depression and anxiety [82], apathy [167] and avoidance[146] are features and symptoms most frequently associated with DM1. In addition, more than 50% of patients with classic DM1 are referred to a consultant because of excessive daytime sleepiness [167, 159, 125].

As for neuroimaging studies, cortical atrophy, increased ventricular size with periventricular hypodensity and the involvement of the subcortical white matter, and calcification of the basal ganglia have been detected in a large proportion of DM1 patients by structural magnetic resonance ([45, 58, 97]. Volumetric analyses based on these studies and using a technique known as voxel-based morphometry (VBM), identified atrophy of the grey matter located mainly in the frontal and bilateral parietal lobes, the bilateral medial temporal gyrus and on the left side of the superior and occipital temporal [11]. There are also some studies that use diffusion tensor imaging (DTI) analyses in order to detect abnormalities associated with regional cortical atrophy ([155]. They observe significantly lower fractional anisotropy and higher mean diffusivity values in the genu, rostral body, anterior midbody, posterior midbody and splenium in patients compared to healthy control subjects.

The neuropathological effect of these structural anomalies is less well known. A loss of cerebral cortex tissue, with neuron inclusion-bodies (especially in the thalamus, the caudate and some nuclei in the brain stem), reduced myelin sheath thickness and an increase in the presence of neurofibrillary tangles have been reported [154]. However, although neurofibrillary tangles have been identified in the limbic system and/or brain stem, hippocampus, entorhinal cortex and temporal regions, no senile plaques or β -amyloid deposits have been found [156]. Molecular studies performed on these brains and on brain tissue from transgenic animals have shown anomalies in tau splicing and protein phosphorylation that may be involved in a group of neurodegenerative disorders known as tauopathies. The clinical expression of these disorders includes frontotemporal cognitive deterioration and other neurodegenerative symptoms [7]. Indeed, some authors have found a relationship between the size of the CTG expansion and the altered tau expression in DM1 patients, which may indicate that DM1 is a special type of tauopathy [183]. This may shed light on a previously unexplored connection between the cognitive alterations present in DM1 and this group of disorders, possibly suggesting, as noted previously, that the cognitive deterioration inherent to DM1 is not as stable as was originally thought [212].

A correlation between β -amyloid and tau levels in the cerebrospinal fluid of DM1 patients and their cognitive performance, expansion size and MRI findings has also been reported recently [213]. DM1 patients displayed structural differences with respect to the controls when studied by MRI. These differences involved smaller brain tissue volume and some specific regions where GM loss was particularly notable, despite five of the 35 patients being excluded prior to the GM volumetric analysis on the basis of anomalies that could have affected this analysis. Some of these anomalies, such as ventricular dilation, have previously been associated with DM1 [5].

2.2.1 Related Studies

Cortical atrophy has frequently been reported in DM1 [45, 97, 20]. Here, we observed that DM1 patients had a lower GM volume than the controls [11, 85]. Likewise, the patients had smaller WM and larger cerebro-spinal fluid (CSF) volumes than the controls. There was no significant difference in TIV between patients and controls, probably due to the fact that the larger volume of the CSF in patients compensates for the decrease in the WM and GM.

Previous studies suggested premature brain aging in DM1 patients, with a faster loss of GM with age in patients than in controls [11]. However, our results do not show such a difference. Indeed, a significant association between age and GM is found in the controls rather than in the patients. It is likely that whereas age is responsible for most of the variation in the GM in controls, other as yet unknown factors that have a stronger influence than age are at play in the patients. It should be noted that the patients have a smaller GM volume than the controls from the outset, which could partly explain why the age-related decrease is smaller in the patients.

Although previous studies have not found any association between brain volume and the patients' molecular defect [20, 85, 141], this could be one of the factors that influences the variation in GM in DM1 patients. However, the number of triplet repeats only tends to be associated with GM volume rather than with that of other tissues.

2.3 Schizophrenia

There is growing research effort devoted to the development of automated diagnostic support tools that may help clinicians perform their work with greater accuracy and efficiency. In medicine, diseases are often diagnosed with the aid of biological markers, including laboratory tests and radiologic imaging. The process of diagnosis becomes more difficult, however, when dealing with psychiatric disorders, in which diagnosis relies primarily on the patient's self-report of symptoms and the presence or absence of characteristic behavioural signs. Schizophrenia is a disabling psychiatric disorder characterized by hallucinations, delusions, disordered thought/speech, disorganized behaviour, emotional withdrawal, and functional decline [15]. Currently, diagnosis is made almost exclusively on subjective measures like self-report, observation, and clinical history.

The most commonly demonstrated DTI abnormalities in schizophrenia are decreased FA in the uncinate fasciculus (a tract connecting temporal and frontal regions and involved in decision-making, emotions, and episodic memory), the cingulum bundle (a tract interconnecting limbic regions which involved in attention, emotions, and memory), and the arcuate fasciculus (a tract connecting language regions) [123]. Lower anisotropic diffusion within white matter may reflect loss of coherence of WM fibre tracts, to changes in the number and/or density of interconnecting fibre tracts, or to changes in myelination [121, 124, 10, 122].

2.3.1 Related Studies

A large number of magnetic resonance imaging (MRI) morphological studies have shown subtle brain abnormalities to be present in schizophrenia. Structural studies have found enlargement of the lateral ventricles, particularly the temporal horn of the lateral ventricles [185]; reduced volumes of medial temporal structures (hippocampus, amygdala, and parahippocampal gyrus) [23, 100, 186], superior temporal gyrus [100], prefrontal cortex [92, 193], and inferior parietal lobule [181, 87]; and reversal of normal left greater than right volume in male patients with schizophrenia [153, 76]. In 1984, Wernicke [208] proposed that schizophrenia might involve altered connectivity of distributed brain networks that are diverse in function and that work in concert to support various cognitive abilities and their constituent operations. Consistent with the “dysconnectivity hypothesis”, studies have found correlations between prefrontal and temporal lobe volumes [211, 34] and disruptions of functional connectivity between frontal and temporal lobes in schizophrenia [144]. These findings strongly point to widespread problems of connectivity in schizophrenia.

Schizophrenia biomarkers Cortical thickness and regional brain volume have been used in two experiments of schizophrenia detection with ($n = 66$ and $n = 26$) correctly classifying 96.7% of the females and 86.1% of the males [194]. Since 1984, the works of Wernicke proposed that schizophrenia might involve altered connectivity of distributed brain networks that are diverse in function and that work in concert to support various cognitive abilities and their constituent operations [208]. Consistent with this “dysconnectivity hypothesis”, functional connectivity studies have found correlations between prefrontal and temporal lobe volumes [211, 34] and disruptions of functional connectivity between frontal and temporal lobes in schizophrenia [144].

Classification related works We found literature on the application of ML algorithms to the discrimination of schizophrenia patients from healthy subjects. A minimum recognition error of 17.8% using geometry features and FA of DTI from a database of 36 healthy subjects and 34 patients with schizophrenia was reported in [206]. A study of the effect of principal component analysis (PCA) and discriminant PCA (DPCA) was carried on FA volumes reaching a minimum one-leave-out validation classification error 20% using Fisher linear discriminant (FLD) in [44]. Good classification results were also obtained in anatomical MRI studies [216, 68].

Experiments on the classification of Schizophrenia patients based on functional MRI data have been reported with small datasets achieving good classification accuracies, e.g., Tang et al. showed a 93.2% [196] with 44 matched subjects and Yu et al. [217] experiment with healthy patient siblings achieves 62% accuracy. A novel kernel approach (BDopt) to Support Vector Machines (SVM) and global complex network measures has been reported [71] to classify a 18 subjects schizophrenia vs. controls dataset with 100% accuracy.

Chapter 3

Pre-processing

Data pre-processing or data cleaning is a very critical step in order to obtain meaningful results. In this thesis we have been working with different modalities of MRI data, each requiring diverse pre-processing procedures. A central procedure is the spatial normalization (registration) of the data volumes in order to ensure that each voxel has an equivalent meaning in all volumes. Specifically, it is desired that their voxels have the same anatomical correspondence, hence the registration is often performed against an anatomically meaningful template. We introduce some basic ideas of image registration to illustrate the general approach, though dealing with an specific approach. Most neuroimaging software packages include a customized version of registration techniques and this step is included in most processing pipelines, moreover when there is a group analysis or inter-subject process. Other typical processes are skull removal, smoothing, tissue segmentation and motion correction in multivariate image modalities. In general, the final result of the pre-processing step is a data instance which is amenable to the ensuing feature selection and extraction processes that will be discussed in Chapter 4. For anatomical (T1-weighted) MRI, we explore two kinds of processes, the Voxel-based Morphology and the Deformation-based Morphometry (DBM). The latter includes the computation of scalar values from 3-dimensional deformation measures. For Diffusion-weighted Imaging (DWI) we produce specific scalar diffusivity measures after appropriate registration and specific cleaning. Finally, we have also pre-processing pipelines tailored to for resting-state fMRI that produce voxel based functional activity measures providing a kind of functional parcelation of the brain and can be used as classification features.

Section 3.1 reviews the image registration fundamentals underlying all spatial normalizations performed in the thesis. Section 3.2 describes the Voxel-based Morphometry (VBM). Section 3.3 describes some features from Deformation-based Morphometry (DBM) that are useful for classification. Section 3.4 describes the pre-processing of Diffusion-weighted Imaging data. Finally, section 3.5 comments on the pre-processing of resting state fMRI and the voxel functional measures computed to perform the ensuing feature selection and classification processes.

3.1 Image Registration basics

There are many image registration methods reported in the neuroimaging literature [115], the work reported here follows a small deformation model. A general expression of the non-linear registration coordinate transformation in the small deformation model is as follows [188, 108]:

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} + \begin{bmatrix} d_x(x, y, z) \\ d_y(x, y, z) \\ d_z(x, y, z) \\ 1 \end{bmatrix} \quad (3.1)$$

where M is an affine transformation matrix, denoting the linear part of the transformation involving scaling, rotation and translation. The term $D = [d_x(x, y, z), d_y(x, y, z), d_z(x, y, z), 1]^T$ denotes the non-linear components of the transformation, the non-linear displacement field, giving the displacement of the registered volume voxel location (x, y, z) in each direction that matches optimally with the reference template volume. Let us denote $g(x, y, z)$ the MRI intensity value, defined as long as x , y and z fall within the original image sampling grid. Therefore, $g(x', y', z')$ denotes the value of g for a set of coordinates after the transformation given by eq. (3.1). Function $g(x, y, z)$ involves some interpolation procedure allowing to compute values of coordinates that did not coincide with original MRI data voxel sites.

Each component $d_i(x, y, z)$ of the displacement field, where $i \in \{x, y, z\}$ denotes the axis, is built as a linear combination of basis functions, i.e.

$$d_i(x, y, z) = \sum_{l, m, n} c_{l, m, n} B_{l, m, n}(x, y, z), \quad (3.2)$$

where $B_{l, m, n}(x, y, z)$ is a basis function, i.e. a 3D cubic spline, value at spatial location $[x, y, z]$. It is notationally convenient to gather the basis coefficients of each displacement field into a vector \mathbf{w}_i , further concatenating all of them into a global parameter vector:

$$\mathbf{w} = [\mathbf{w}^{(x)T}, \mathbf{w}^{(y)T}, \mathbf{w}^{(z)T}]^T. \quad (3.3)$$

Then, $g_{xyz}(\mathbf{w})$ denotes the value of g at (x', y', z') under the transform determined by \mathbf{w} , where M is implicit. This general notation is independent of our choice of basis-function and would be equally valid for e.g. a discrete cosine basis set. The gradient of g at point (x', y', z') is defined as

$$\nabla g_{xyz}(\mathbf{w}) = \left[\left. \frac{\partial g_{xyz}}{\partial x} \right|_{\mathbf{w}}, \left. \frac{\partial g_{xyz}}{\partial y} \right|_{\mathbf{w}}, \left. \frac{\partial g_{xyz}}{\partial z} \right|_{\mathbf{w}} \right] \quad (3.4)$$

where the partial derivative with respect to e.g. x , $\partial g_{xyz} / \partial x|_{\mathbf{w}}$, denotes the rate of change of g at (x', y', z') as one translates the sampling point in the x -direction. The exact expression of the partial derivatives depends on the interpolating function invoked to compute $g(x', y', z')$, and on

the basis functions used to model the displacement fields. Assuming that parameter $w_i^{(x)}$ is the coefficient for the lmn -th 3D spline basis function composing the x -displacement field d_x we can write

$$\left. \frac{\partial g_{xyz}}{\partial w_i^{(x)}} \right|_{\mathbf{w}} = \left. \frac{\partial g_{xyz}}{\partial x} \right|_{\mathbf{w}} B_{lmn}(x, y, z), \quad (3.5)$$

where the \mathbf{w} subscript indicates that the derivative has been calculated at point \mathbf{w} in parameter space, and i identifies the lmn -th 3D spline basis function. To obtain a matrix expression, the vector of signal values obtained after registration with the deformation fields determined by \mathbf{w} is:

$$\mathbf{g}(\mathbf{w}) = [g_{x,y,z}(\mathbf{w}); x \in \{1, \dots, X\}, y \in \{1, \dots, Y\}, z \in \{1, \dots, Z\}]^T. \quad (3.6)$$

Analogously,

$$\left. \frac{\partial g_{xyz}}{\partial x} \right|_{\mathbf{w}} = \left[\left. \frac{\partial g_{x,y,z}}{\partial x} \right|_{\mathbf{w}}; x \in \{1, \dots, X\}, y \in \{1, \dots, Y\}, z \in \{1, \dots, Z\} \right]^T \quad (3.7)$$

and

$$\mathbf{B}_{lmn} = [B_{lmn}(x, y, z); x \in \{1, \dots, X\}, y \in \{1, \dots, Y\}, z \in \{1, \dots, Z\}]^T. \quad (3.8)$$

We can write \mathbf{B}_{lmn} as the vector representation of the i -th basis function \mathbf{B}_i . Combining equations (3.5), (3.7) and (3.8) we can define the matrix

$$\mathbf{J}_x(\mathbf{w}) = \left[\left. \frac{\partial g}{\partial x} \right|_{\mathbf{w}} \odot \mathbf{B}_1 \quad \left. \frac{\partial g}{\partial x} \right|_{\mathbf{w}} \odot \mathbf{B}_2 \dots \left. \frac{\partial g}{\partial x} \right|_{\mathbf{w}} \odot \mathbf{B}_{LMN} \right], \quad (3.9)$$

where \odot denotes element-wise Hadamard matrix product. Each element of $\mathbf{J}_x(\mathbf{w})$ is of the form given by equation (3.5). Finally, the Jacobian matrix \mathbf{J} of the mapping $\mathbf{g}(\mathbf{w}) : \mathbb{R}^{3LMN} \rightarrow \mathbb{R}^{XYZ}$ is composed as:

$$\mathbf{J}(\mathbf{w}) = [\mathbf{J}_x(\mathbf{w}) \quad \mathbf{J}_y(\mathbf{w}) \quad \mathbf{J}_z(\mathbf{w})], \quad (3.10)$$

where $\mathbf{J}_y(\mathbf{w})$ and $\mathbf{J}_z(\mathbf{w})$ are computed analogously to $\mathbf{J}_x(\mathbf{w})$, $\mathbf{J}(\mathbf{w})$ is of size $XYZ \times 3LMN$, and LMN and XYZ are the number of 3D spline basis functions and the number of samples in the volume, respectively.

The estimation of the parameters \mathbf{w} determining the displacement fields $d_i(x, y, z)$ that transform the image native space into a reference space is performed minimizing an objective function

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} O(\mathbf{w}) \quad (3.11)$$

measuring the matching between the warped image $\mathbf{g}(\mathbf{w})$ and the reference template $\mathbf{f} \in \mathbb{R}^{XYZ}$:

$$O(\mathbf{w}) = \frac{1}{XYZ} (\mathbf{g}(\mathbf{w}) - \mathbf{f})^T (\mathbf{g}(\mathbf{w}) - \mathbf{f}). \quad (3.12)$$

This minimisation is performed by a Levenberg-Marquardt procedure, using the Gauss-Newton approximation of the Hessian matrix of the objective function

$$H_{ij}(\mathbf{w}) = \left. \frac{\partial^2 O}{\partial w_i \partial w_j} \right|_{\mathbf{w}} \quad (3.13)$$

given by

$$H(\mathbf{w}) = \frac{2}{XYZ} \mathbf{J}^T(\mathbf{w}) \mathbf{J}(\mathbf{w}).$$

Details on the registration of the OASIS database The subjects in the database were already linearly registered to a MNI152 template [137]. Taking this into account, we need to non linearly transform them to a common template in order to obtain the deformation fields which will be used as starting point in the feature extraction process. For this non-linear registration step we could have used again the MNI152 standard template, but the registration algorithm used in this study [16] could not cope well with the “large” deformations required to register some subjects with enlarged ventricles. For this reason, we tried two approaches: 1) using FMRIB Software Library’s (FSL) FNIRT [188] instead [175] and 2) creating a custom brain template volume with all the subjects of the database with SPM8 [178]. This custom template was subsequently non-linearly registered to all the study subjects. As a result from the registration, displacement vector volumes for each subject are obtained. These volumes describe the effects of deformation of the template brain to the subject’s.

3.2 Voxel-based Morphometry

Morphometry analysis has become a common tool for computational brain anatomy studies. It allows a comprehensive measurement of structural differences within a group or across groups, not just in specific structures, but throughout the entire brain. Voxel-based Morphometry (VBM) is a computational approach to neuroanatomy that measures differences in local concentrations of brain tissue, through a voxel-wise comparison of multiple brain images [12]. For instance, VBM has been applied to study volumetric atrophy of the grey matter (GM) in areas of neocortex of Alzheimer’s disease (AD) patients vs. control subjects [41, 179, 78].

The computational pipeline of VBM is illustrated in figure 3.1. The procedure involves the spatial normalization of subject anatomical MRI images into a standard space, segmentation of tissue classes using *a priori* probability maps, the resulting segmentation image is modulated to account for effects of the spatial normalization, smoothing to correct noise and small variations, and voxel-wise statistical tests. Smoothing is done by convolution with a Gaussian kernel whose

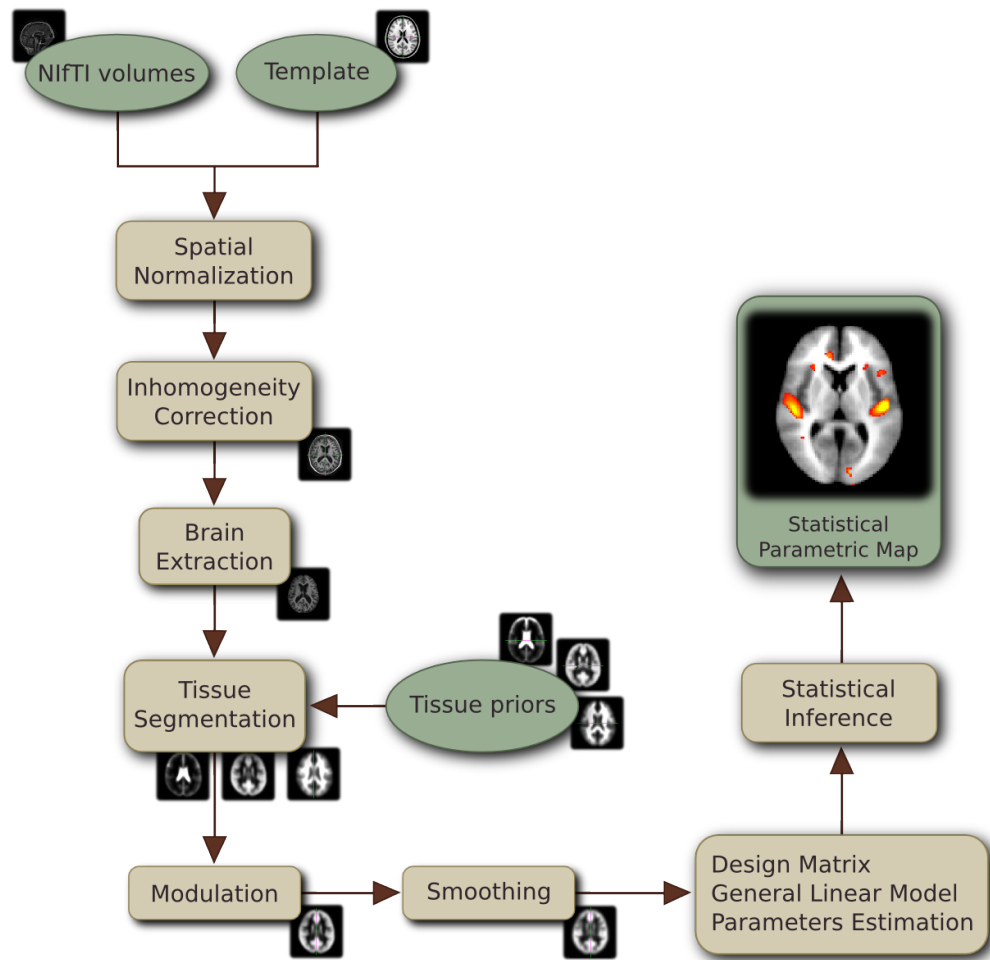


Figure 3.1: VBM computational pipeline.

Full-Width at Half-Maximum (FWHM) is tuned to the data of each case study. Statistical analysis is based on the General Linear Model (GLM) to describe the data in terms of experimental and confounding effects, and residual variability, applied to each voxel independently. Statistical inference is used to test hypotheses that are expressed as linear functions of the GLM estimated regression parameters. These linear scalar functions are called *contrasts* in VBM terminology, they specify diverse statistical tests (i.e. t-test, F-test). Different versions of VBM appear in different software bundles (i.e. SPM or FSL). In SPM, the value of the contrast at each voxel is interpreted as a Statistical Parametric Map (SPM), which is optimally thresholded according to the Random Field theory. In FSL, the detection of significant voxels is performed applying a randomization test which builds an empirical cumulative distribution function of the contrast under the null hypothesis.

Image processing and VBM for OASIS98 We have used the average MRI volume for each subject, provided in the OASIS98 data set [173]. Images were reoriented into a right-handed coordinate system to work with SPM. These images are already registered and re-sampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [137], thus the tissue segmentation step does not need to perform bias correction. SPM [3] was used to compute the VBM which gives us the spatial mask to obtain the classification features. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the FWHM of the Gaussian kernel to 10mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data is modulated, but it is close enough for the mask to be reasonable. We designed the statistical analysis as a two-sample t-test in which the first group corresponds with AD subjects. In SPM software jargon: the contrast has been set to [-1 1] right-tailed (Controls > AD patients) with $p\text{-value} < 0.05$ Familywise Error rate (FWE) correction. The VBM detected clusters are used for feature extraction. Statistical significance was determined using an extent threshold of 0 adjacent voxels for two sample comparisons. The clusters of significant voxels detected by the VBM analysis are displayed in figure 3.2, they agree with the findings reported in the literature [41, 179, 78].

Image processing and VBM for MD1 The SPM [3] was used to compute the VBM which gives us the spatial mask to obtain the classification features [173]. Images were reoriented into a right-handed coordinate system to work with SPM8. The tissue segmentation step does not need to perform bias correction. We performed the modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before performing the voxel-wise statistics, setting the FWHM of the isotropic Gaussian kernel to 8mm, 9mm, 10mm,

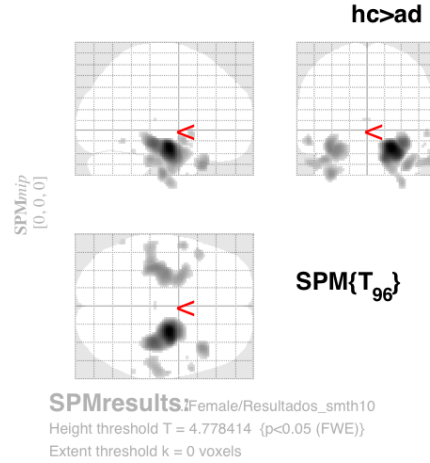


Figure 3.2: SPM results: clusters of significant voxels with increased grey matter density in the controls relative to the patient subjects, detected by the VBM process in the OASIS data.

11mm and 12mm . For all comparisons we have applied a significance probability threshold of $p < 0.05$ FWE corrected for multiple comparisons at a voxel level. Statistical significance was determined using spatial extent thresholds of 0, 100 and 200 adjacent voxels for two sample comparisons.

3.3 Deformation-based morphometry

An alternative morphometric analysis is based on the deformation fields obtained from non-linear registration. Once this registration is performed in multiple images, the resultant displacement fields contain information on the estimated correspondences with the reference volume. The term deformation-based morphometry (DBM) will be used to refer generally to approaches using this concept. Figure 3.3 illustrates the process of obtaining scalar maps characterizing the registration deformation at each voxel. These scalar maps will be used for feature extraction in ensuing processes.

For each voxel i , the displacement field for one subject have a vector (x_i, y_i, z_i) representing the ending point of voxel i in the registration process. The spatial derivative of a three-dimensional deformation field is a three-by-three matrix, or second order tensor, known as the Jacobian. The term tensor-based morphometry (TBM) is often used to name the analysis of a measure derived from the Jacobian. The determinant of this Jacobian matrix is the most commonly used scalar measure of deformation for TBM analyses [128]. Given the non-linear defor-

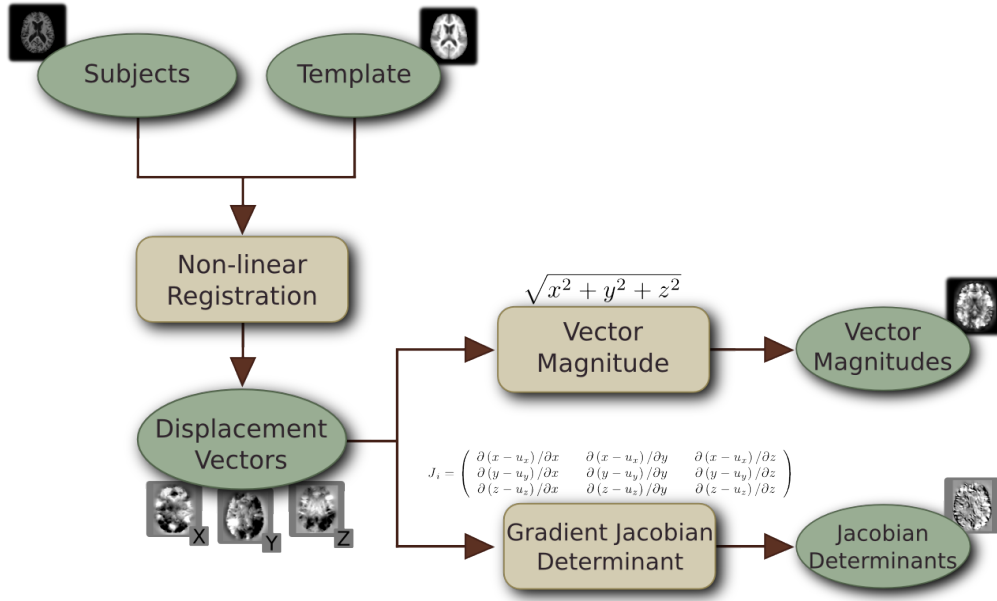


Figure 3.3: Pipeline of the image pre-processing steps.

mation field $D = [d_x(x, y, z), d_y(x, y, z), d_z(x, y, z), 1]^T$, its Jacobian matrix is defined as:

$$J = \begin{bmatrix} \frac{\partial(x-d_x)}{\partial x} & \frac{\partial(x-d_x)}{\partial y} & \frac{\partial(x-d_x)}{\partial z} \\ \frac{\partial(y-d_y)}{\partial x} & \frac{\partial(y-d_y)}{\partial y} & \frac{\partial(y-d_y)}{\partial z} \\ \frac{\partial(z-d_z)}{\partial x} & \frac{\partial(z-d_z)}{\partial y} & \frac{\partial(z-d_z)}{\partial z} \end{bmatrix}. \quad (3.14)$$

The Jacobian matrix describes the velocity of the deformation in the neighbouring area of each voxel. Let us denote $J(\mathbf{x})$ the local Jacobian matrix at voxel site \mathbf{x} . Its determinant, the Jacobian, is used to analyse the deformation needed for the spatial registration. A value $\det(J(\mathbf{x})) > 1$ implies that the neighbourhood of the voxel \mathbf{x} was stretched to match the template (i.e., local volumetric expansion), while $\det(J(\mathbf{x})) < 1$ is associated with local shrinkage. In TBM the Jacobian map obtained computing the Jacobian at each voxel site is used to compute group statistics at each voxel, in the template reference frame, to identify group differences in anatomical shape.

The following scalar measures are used in the computational experiments

1. The Jacobian map (*jacs*) given by $\det(J(\mathbf{x}))$ at each voxel site \mathbf{x} .
2. The modulated GM (*modgm*) segmentation data computed multiplying the GM segmentation probability by the Jacobian at each voxel.
3. The trace of the Jacobian matrix trace($J(\mathbf{x})$) at each voxel site \mathbf{x} (*trace*).
4. The magnitude of the displacement field (*norms*) at each voxel site $m(\mathbf{x}) = \|[d_x(\mathbf{x}), d_y(\mathbf{x}), d_z(\mathbf{x})]\|$

5. The Geodesic Anisotropy [26] (*geodan*) defined as the geodesic distance between a strain tensor $\mathbf{S}_i = (\mathbf{J}_i^T \mathbf{J}_i)^{1/2}$ and the isotropic tensor within the affine-invariant metric framework [128]:

$$\text{GA}(S) = \left(\text{Trace}(\log S - \langle \log S \rangle I)^2 \right)^{1/2}$$

where

$$\langle \log S \rangle = \frac{\text{Trace}(\log S)}{3}.$$

6. The displacement vector magnitudes (DM)

$$DM_i = \sqrt{d_x^2 + d_y^2 + d_z^2},$$

In this thesis, we use these maps as voxels raw data for the feature selection procedures that will be described later. Figure 3.4 shows instances of these maps.

3.4 Diffusion Weighted Imaging (DWI)

Diffusion-Weighted Imaging (DWI) is a method that can provide structural tissue information using the phenomenon of diffusion of water molecules to give contrast to images. DWI depends upon the motion of water molecules to provide structural information *in vivo* [160, 25]. By allowing estimation of the water diffusion profile in different tissue regions, we can study the integrity of the White Matter (WM) fibres [30].

Diffusion Tensor Imaging (DTI) Diffusion Tensor Imaging (DTI) [24] captures the potential anisotropic diffusion of the tissues and for each voxel its signal will be:

$$\mathcal{S}(\mathbf{g}, \tau) = \mathcal{S}(0) \cdot \exp(-b \cdot \mathbf{g}^T \cdot \mathcal{D} \cdot \mathbf{g}),$$

where S_0 is the signal intensity without the diffusion weighting, \mathbf{g} is a unit vector which in the direction of the acquisition gradient, τ is the effective diffusion time, b the b-value and \mathcal{D} the diffusion tensor. This tensor \mathcal{D} , a 3×3 symmetric positive definite matrix, can be diagonalized and its eigenvectors $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ and eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ calculated. A diffusion ellipsoid, whose axial orientations coincide with the eigenvectors, reveals the anisotropy degree of \mathcal{D} . Rotation invariant scalar parameters can be extracted from the diffusion tensor. The Fractional Anisotropy (FA),

$$FA = \sqrt{\frac{3 \sum_{i=1}^3 (\lambda_i - \bar{\lambda})^2}{2 \sum_{i=1}^3 \lambda_i^2}},$$

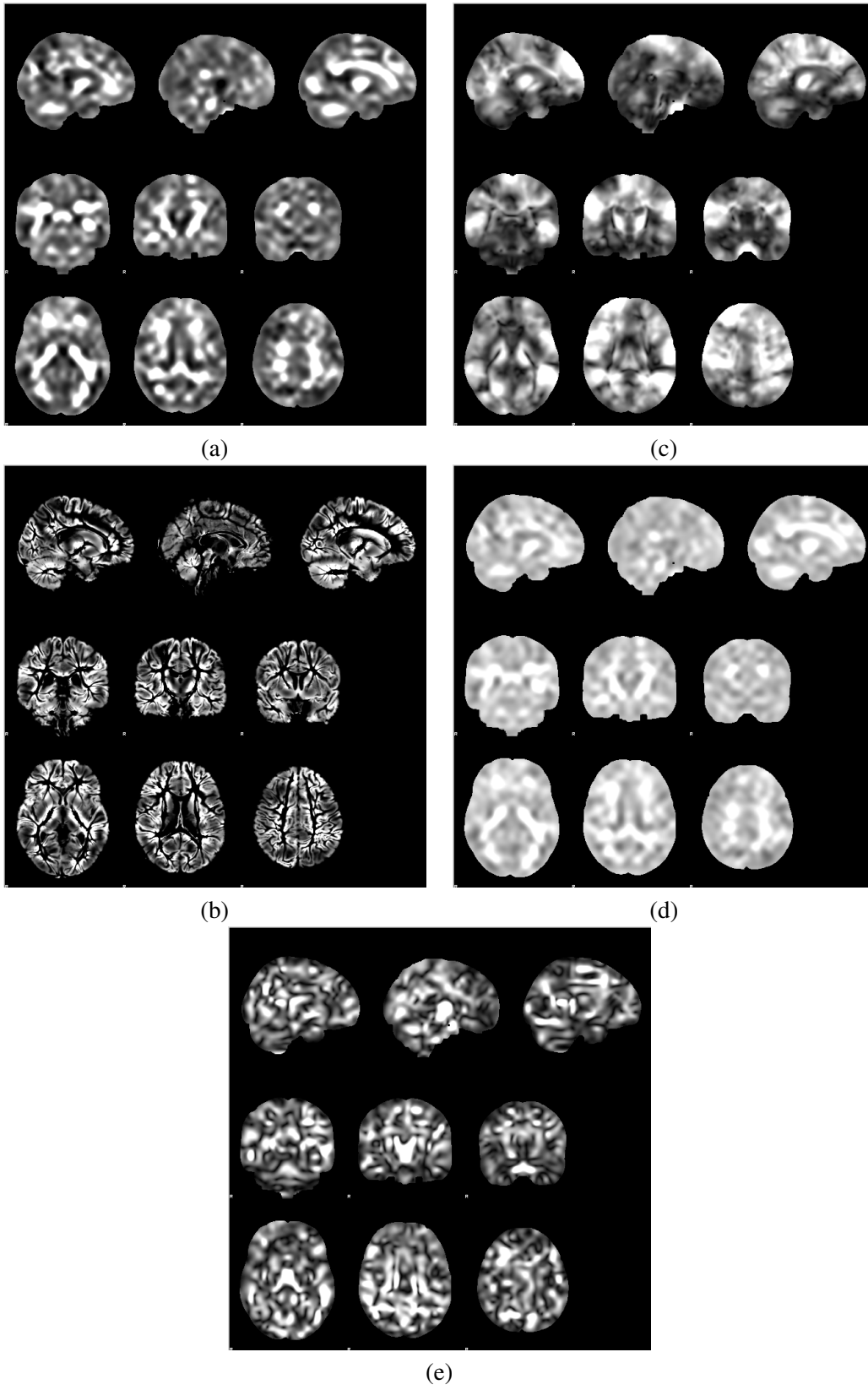


Figure 3.4: Example of deformation measure maps from the OASIS database (a) Jacobian determinant, (b) modulated GM, (c) displacement norm, (d) Jacobian matrix trace and (e) geodesic anisotropy.

Diffusion tensor image pre-processing

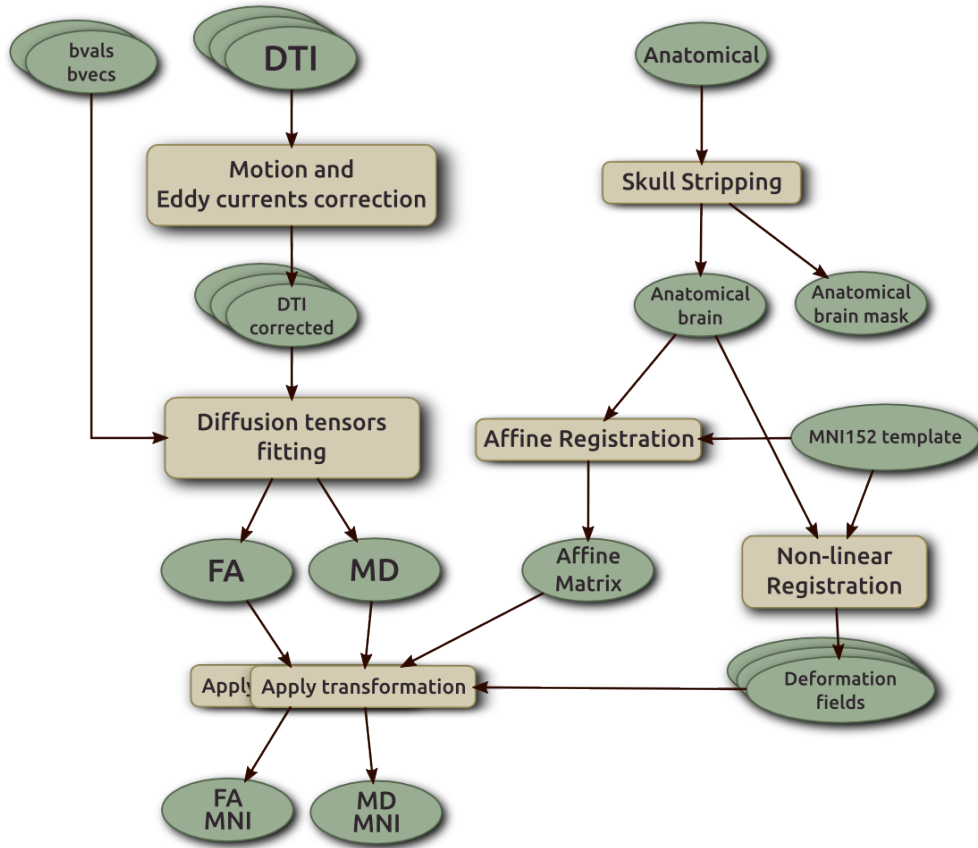


Figure 3.5: Image preprocessing pipeline for the preprocessing the DWI data in order to obtain the FA and MD scalar measures of diffusivity.

is a normalized expression of the variance of the tensor eigenvalues. It has high values in WM and low for grey matter and CSF regions. The Mean Diffusivity (MD),

$$MD = (\lambda_1 + \lambda_2 + \lambda_3) / 3$$

is the mean of the tensor eigenvalues.

3.4.1 Pre-processing

The careful pre-processing of the data is of paramount importance. Figure 3.5 provides a diagram illustrating the pre-processing pipeline to obtain the spatially normalized FA and MD volumes. All processes, except the non-linear registration, have been performed using FSL software (<http://www.fmrib.ox.ac.uk/fsl/>). First step was converting the file format from DICOM images

to NIfTI format. Second, T1-weighted MRI volumes were skull stripped and affine registered to the Montreal Neurological Institute (MNI152) standard template. Third, we computed their non-linear diffeomorphic registration (fast symmetric normalization) [16] to MNI152 template using ANTS software (<http://www.picsl.upenn.edu/ANTS>) applying a probabilistic correlation similarity metric of window radius 4 and gradient step length 1. The similarity optimization applied a three resolutions schedule of 100, 100, and 10 iterations at coarse, medium and full resolution, respectively. Gaussian regularization ($\sigma = 3$) on the deformation field was applied. Finally, the transformation model's step size was set to 0.15. To obtain the FA and MD volumes, first we apply the correction for eddy currents to the DWI data, then we extract the brain mask of the volumes and perform a least squares estimate of the diffusion tensor to obtain the FA and MD measures [27]. We spatially normalize the FA and MD maps to ensure the anatomical correspondence of voxels across all subjects applying the affine transformations and non-linear deformations computed to register the structural volumes to the MNI152 template.

3.5 Resting state functional MRI (fMRI)

Slow fluctuations in activity measured by the BOLD signal in functional MRI of the resting brain (rs-fMRI) allows to find correlated activity between brain regions and determine resting state networks. Measures on the correlation of these fluctuations provide functional connectivity maps that may serve as biomarkers or discriminant features for individual variations or dysfunction. The processing of fMRI data has two aspects in this thesis: first, the spatial normalization and filtering of the data to fit it to the appropriate analysis; second, the computation of functional measures at the voxel level. Figure 3.6 provides the pipeline of processes that give the voxel functional measures.

3.5.1 Pre-processing

Individual functional and anatomical acquisitions have been registered to the MNI152 template [73], provided by the Montreal Neurological Institute, using FSL FNIRT [109, 107]. In addition, AFNI SkullStrip (see section B.4) and FSL FAST [221] have been used for brain extraction and tissue segmentation. The first 6 fMRI volumes were discarded for scanner calibration leaving 144 time volumes. The data preprocessing included slice timing, head motion correction (Friston's 24 parameters motion model [162, 80, 172, 201]) and nuisance corrections (principal components regression[28] and linear detrending). Four different preprocessing results can be obtained depending on the combination of band-pass temporal filtering (TPF) between 0.01 and 0.1Hz [13] and the global signal regression (GSR) [75] which is performed on the data.

The processing sequence has been carried out using an open source software pipeline known as Configurable Pipeline for the Analysis of Connectomes (C-PAC) (<http://fcp-indi.github.io/>), built upon AFNI [56], FSL (the FMRIB Software Library) [108] and FreeSurfer [72] further introduced in Appendix B.

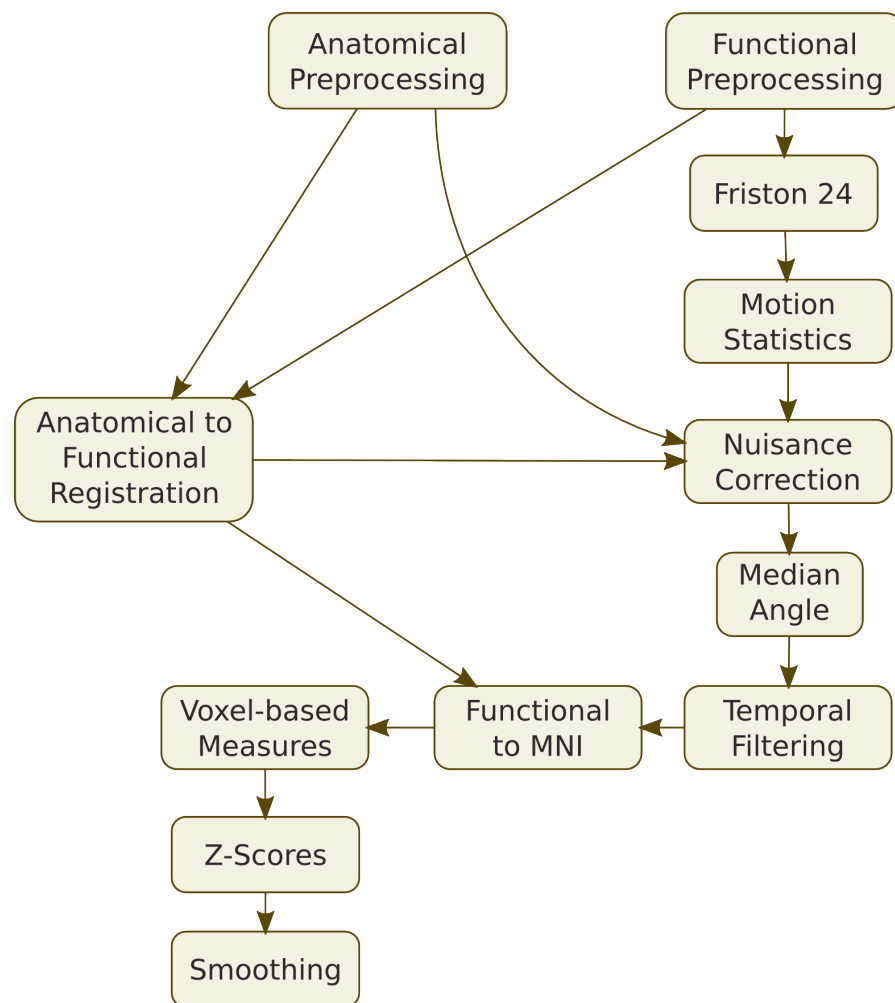


Figure 3.6: Resting-state fMRI preprocessing workflow.

3.5.2 Voxel activity measures from resting-state fMRI

The majority of rs-fMRI studies incorporate functional connectivity to analyse their data, thus looking into temporal relationships between shifts observed in different brain regions. Nonetheless, functional connectivity delivers little insight about local properties of spontaneous brain activity observed in singular regions. Several local measurements of brain activity are complementary to functional connectivity analyses [130]. Measures on the slow fluctuations in activity in the resting brain may serve as markers or discriminant features for individual variations or dysfunction, as they can vary between brain regions and between subjects.

ALFF and fALFF: Amplitude of Low Frequency Fluctuations (ALFF) [218] and fractional Amplitude of Low Frequency Fluctuations (fALFF) [222] are considered measures of amplitude for low frequency oscillations (LFOs) of the BOLD signal. ALFF is an index for the strength or intensity of the LFO defined as the total power within the frequency range between 0.01 and 0.1 Hz. fALFF is the relative contribution of specific LFO to the power of whole frequency range, defined as the power within the low-frequency range (0.01-0.1 Hz) split by the total power in the entire detectable frequency range [223].

VHMC: Voxel-Mirrored Homotopic Connectivity (VMHC) quantifies functional homotopy through a voxel-wise measure of connectivity between hemispheres, assuming the synchronization of spontaneous activity between homotopic (geometrically corresponding) regions at each hemisphere. The strength of these homotopic patterns can vary between regions [192], providing a fingerprint of the brain functional connectivity. An estimation of this connectivity is calculated between each voxel in one hemisphere and its mirrored counterpart in the other [224], assuming morphology symmetry between them. To ensure this property, a symmetric anatomical volume is created averaging the anatomical volume with its mirrored version. The fMRI data is registered to the symmetric anatomical volume.

ReHo: Regional Homogeneity (ReHo) is a voxel-based measure of brain activity which estimates the similarity between the time series of a given voxel and a specific number of its nearest neighbours [219], requiring no *a priori* specification of ROIs. Similarities within a voxel cluster is computed as the Kendall's coefficient of concordance (KCC) [113]. In this paper the cluster size has been set to 27 neighbouring voxels. The KCC values are standardized and smoothed (4mm FWHM) to build a voxel-based map for each subject.

Chapter 4

Feature Selection

The feature selection process is central in the studies of this thesis. It provides the input data for the classifiers that will try to implement a Computer Aided Diagnosis system. Besides, we are following a general procedure that provides the anatomical locations where the data is extracted including, when possible, relevance rankings of anatomical locations, based on the classification results. The overall procedure involves as well the computation of scalar measures of the data at each voxel site, in case of multivariate modalities, such as DTI or fMRI, the computation of voxel significance measures based on the given categorical information and the selection of voxel sites according to a percentile of the empirical distribution of this significance measure.

Section 4.1 describes the general feature selection procedure. Section 4.2 introduces the voxel significance measures used in most of the works in this thesis. Section 4.3 presents the use of VBM for feature selection. Section 4.4 describes the feature extraction based on ROIs given by an atlas. Section 4.5 presents some specifics of the feature extraction from deformation information obtained from image registration. Section 4.6 presents the feature selection from diffusion data. Section 4.7 presents the evolutionary wrapper approach that has been tested on resting state fMRI data.

4.1 General procedure

Considering each voxel site independently, we build a vector \mathbf{v}_j composed of the scalar valued voxel signal intensities at the j -th voxel site across all the subjects derived from the data. We compute the voxel saliency measure $r_{\mathbf{v}_j, \mathbf{y}}$ involving this vector and the categorical variable, $y_i = 0$ for Healthy Control and $y_i = 1$ for Patient, obtaining independent volumes of such measures when we combine diverse scalar voxel intensities and/or significance measures, as described below. Using the significance measure volume we compute the empirical distribution of the

absolute correlation values in this volume

$$F_r(R) = \frac{\#\{j \mid |r_{v_j, y}| \leq R\}}{\#\{j\}}, \quad (4.1)$$

where $\#\{\cdot\}$ denotes the set cardinality. Then, for a given percentile θ of this empirical distribution, we select the voxel sites whose absolute correlation value has cumulative frequency above it:

$$J(\theta) = \{j \mid F_r(|r_{v_j, y}|) \geq \theta\}. \quad (4.2)$$

These voxel sites in $J(\theta)$ correspond to discriminant features of the data volumes. Finally, for a given percentile we compose the feature vector \mathbf{x}_i for each subject as follows:

$$\mathbf{x}_i = [v_{i,j}; j \in J(\theta)]. \quad (4.3)$$

Different feature vectors \mathbf{x}_i are extracted from each raw data, i.e. the FA and MD data in DWI image processing provide disjoint feature vectors.

4.2 Voxel significance measures

On the raw voxel data we calculate the significance measure of each voxel site. This significance measure has the following forms in this thesis:

- The absolute value of the Pearson's correlation (PC) between voxel data and the subject class labels. Big absolute values imply high voxel significance.
- The Bhattacharyya distance (BD) between classes. This is a class separability measure, so that high values mean good class separation, and thus high voxel significance.
- Welch's t-test (WT) between both groups testing if the data belong to the same distribution. Its value indicates the difference between the mean of two normally distributed populations when the variances of the two populations are not assumed to be equal, based on two independent samples.
- The Spearman's correlation coefficient is often thought of as being the Pearson's correlation coefficient between ranked variables.

4.2.1 Pearson's correlation

The Pearson's product-moment correlation coefficient [54] (PMCC or PC), typically denoted by r , is a measure of the correlation (linear dependence) between two random variables \mathbf{x} and \mathbf{y} , taking values in the interval $r \in [-1, 1]$. It is widely used in the sciences as a measure of the strength of linear dependence between two variables. It was developed by Karl Pearson

from a similar but slightly different idea introduced by Francis Galton in the 1880s. A value of $r = 1$ means that the two variables are in complete agreement. A value $r = -1$ means that the two variables take opposite values. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables normalized by the product of their standard deviations:

$$r(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}}, \quad (4.4)$$

where μ_x and μ_y are the means of variables \mathbf{x} and \mathbf{y} , respectively. In our case, considering each voxel site independently, we build a vector \mathbf{x}_j composed of the voxel intensities at the j -th voxel site across all the subjects in the training set (fully independent cross-validation method) or in the whole dataset (partially independent cross-validation method). Vector \mathbf{y} contains the subjects class labels, defined as:

$$y_i = \begin{cases} 1 & \text{if } i\text{-th subject is patient} \\ 0 & \text{otherwise} \end{cases}$$

Then for each voxel site $\text{abs}(r)$ is the significance measure for feature extraction.

4.2.2 Bhattacharyya's distance

The Bhattacharyya's distance [29] is a measure of dissimilarity between two discrete or continuous probability distributions. Assuming that the data from the same voxel site values in patients (\mathbf{p}) and controls (\mathbf{q}) follow two univariate Gaussian distributions, the Gaussian univariate Bhattacharyya's distance has the following form [152][111]

$$b(\mathbf{p}, \mathbf{q}) = \frac{1}{4} \frac{(\mu_p - \mu_q)^2}{\sigma_p^2 + \sigma_q^2} + \frac{1}{2} \ln \frac{\sigma_p^2 + \sigma_q^2}{2\sigma_p\sigma_q}, \quad (4.5)$$

where (μ_p, σ_p^2) and (μ_q, σ_q^2) are the sample mean and sample variance of the intensity values of patients and controls, respectively.

4.2.3 Welch's t-test

Welch's t-test [207] is an adaptation of Student's t-test to compare two samples that may have unequal variances. The t-statistics is computed as follows

$$t(\mathbf{p}, \mathbf{q}) = \frac{(\mu_p - \mu_q)}{\sqrt{\frac{\sigma_p^2}{n_p} + \frac{\sigma_q^2}{n_q}}}, \quad (4.6)$$

where (μ_p, σ_p^2, n_p) and (μ_q, σ_q^2, n_q) are the sample mean, sample variance and sample size of the intensity values of patients and controls, respectively. The conventional way of working with these will be to compute the degrees of freedom of the t-distribution, performing a two-tailed test on the statistic to reject or accept the null hypothesis. Instead, we will use the t-statistic's absolute value as the voxel site significance measure.

4.2.4 Spearman's correlation

The Spearman's correlation coefficient is often thought of as being the Pearson's correlation coefficient between ranked variables. In practice, however, a simpler procedure is normally used to calculate ρ . The n raw scores X_i, Y_i are converted to ranks x_i, y_i and the differences $d_i = x_i - y_i$ between the ranks of each observation on the two variables are calculated.

If there are no tied ranks, then ρ is given by [139]:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (4.7)$$

If tied ranks exist, Pearson's correlation coefficient between ranks should be used for the calculation 4.4. One has to assign the same rank to each of the equal values. It is an average of their positions in the ascending order of the values.

4.3 Voxel-based Morphometry features

Voxel-based Morphometry (VBM) To perform the VBM analysis we have used the average MRI volume for each subject. These images are already registered and re-sampled into a 1-mm isotropic image in atlas space and the bias field has been already corrected [138]. The Statistical Parametric Mapping (SPM8) [3] was used to compute the VBM which gives us the spatial mask to obtain the classification features. A bias correction step has been performed with the tissue segmentation step and also a modulation normalization for GM, because we are interested in this tissue for this study. We performed a spatial smoothing before executing the voxel-wise statistics, exploring the Full-Width at Half-Maximum (FWHM) of the Gaussian kernel in a range from 8mm up to 12mm isotropic. A GM mask was created from the average of the GM segmentation volumes of the subjects under study. Thresholding the average GM segmentation, we obtain a binary mask that includes all voxels with probability greater than 0.1 in the average GM segmentation volume. This interpretation is not completely true, since the data is modulated, but it is close enough for the mask to be reasonable.

Feature extraction Here we focus on the feature extraction process derived from VBM cluster findings. A flow chart of the whole classification and detection process is shown in figure 4.1. The VBM process is illustrated in the box embedded in the figure. The VBM is applied to the set of anatomical MRI volumes used for the experiment. The loci of clusters of significant

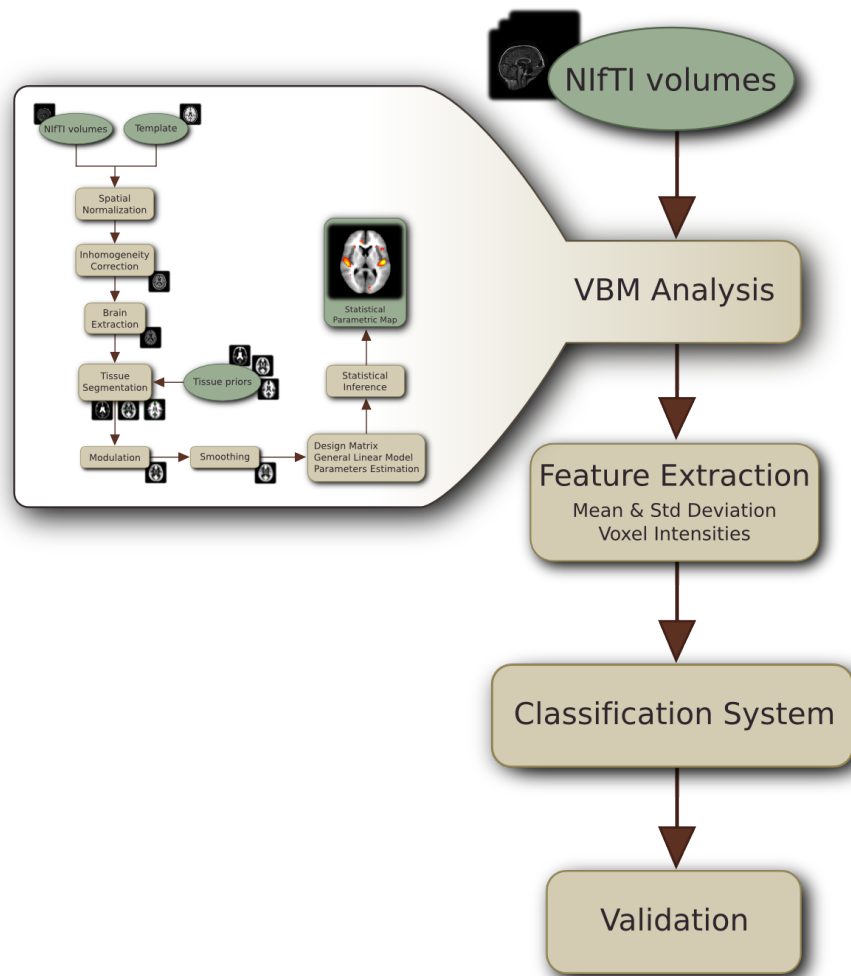


Figure 4.1: Flow diagram of the VBM based feature extraction process from the subjects GM segmentation volumes.

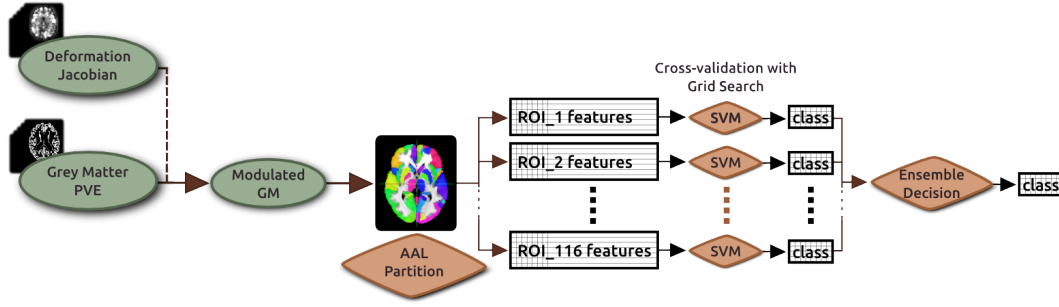


Figure 4.2: Flow diagram of the feature extraction and ensemble of classifiers.

voxels are used to extract classification feature vectors which are used to train and validate the classification systems build with the above mentioned Machine Learning techniques.

We have applied this approach to Alzheimer’s Disease data from the OASIS database and Myotonic Dystrophy Type 1 from the Neurology Department of the Donostia Hospital (San Sebastián, Spain) (see descriptions of the datasets in appendix A). The feature datasets are published in the research group web page ¹ to allow the independent validation of our results, and to allow new classification approaches to be tested on this data. The process consists in the realization of Voxel-based Morphometry analysis to detect clusters of voxels which can be discriminant about AD patients and controls. These clusters are used as an spatial mask to extract from the original volumes the values of the corresponding voxels. We have tested three different feature extraction processes, based on the voxel location clusters obtained from the VBM analysis:

1. The first feature extraction process computes the ratio of GM voxels to the total number of voxels of each voxel location cluster, so we have as many features as clusters.
2. The second feature extraction process computes two features *per* VBM cluster: the mean and standard deviation of the GM voxel intensity values of each voxel location cluster.
3. The third feature extraction process extracts a very high dimensional vector with all the GM segmentation values for the voxel locations included in each VBM detected cluster. The GM segmentation voxel values were ordered in this feature vector according to the coordinate lexicographic order

4.4 Atlas based GM feature selection

The procedure we have followed to obtain atlas based features from anatomical MRI data was:

1. Segment the subjects in 3 tissue volume estimation maps,
2. Nonlinearly register each subject to the MNI template,

¹<http://www.ehu.es/ccwintco/index.php/GIC-experimental-databases>

3. Calculate the Jacobian determinant of the corresponding displacement fields, with this,
4. Modulate the GM partial volume estimation maps using the Jacobians.
5. Extract from the GM voxels inside each region of interest (ROI) in the Automatic Anatomical Labeling (AAL) atlas a set of statistical values and finally,
6. Use the feature set of each ROI in a leave-one-out classification procedure with parameter grid search. In figure 4.2 we show a pipeline of the experiment procedure.

The implementation of the feature extraction, classification and result measures have been done in Python with scikit-learn [158]. The source code and preprocessing scripts are freely available for [download](http://www.ehu.es/ccwintco/index.php/Usuario:Alexsavio) in <http://www.ehu.es/ccwintco/index.php/Usuario:Alexsavio>.

Feature extraction The AAL atlas [200] is used to partition the GM maps into 116 brain anatomical regions. In this study we extract of each AAL anatomical region from each subject GM map 7 statistical measures: the maximum voxel value, the minimum, the mean, the variance, the median, the kurtosis and the skewness. Resulting in 116 sub-datasets of 416 subjects with 7 features each.

4.5 Deformation-based Morphometry

Deformation-based Morphometry (DBM) [128, 198] studies aim to detect morphological changes by statistical analysis of scalar values from deformation measures computed from the Jacobian matrix of the deformation field obtained by non-linear registration of the study's subjects to a template. DBM allows a comprehensive measurement of structural differences within or across groups, not only in specific structures but throughout the entire brain. In this thesis, deformation scalar measures are the raw data for the construction of the classifier systems. Figure 4.3 shows a general pipeline of the computation of them. Feature selection is performed as explained in Sections 4.1 and 4.2

The feature values were extracted from the scalar DM and JD maps computed on the displacement vectors resulting from the registration processes, as described in figure 4.3. For each voxel site i we extracted one vector \mathbf{v}_i with n components being the value of the voxel i of each one of the n subjects of this experiment. Afterwards correlation measures of these vectors with the control variable, specified by the vector containing the subject class label (-1 for control subject or 1 for patient) were performed. Pearson's and Spearman's correlation were used in this study. Volume masks containing the voxel sites whose correlation values were above some specified percentile (i.e. 0.990, 0.995) for each combination of map and correlation measure were applied to the corresponding DM or JD map to extract the feature vectors. Figure 4.4 illustrates the process. Features were normalized before training the SVM on them.

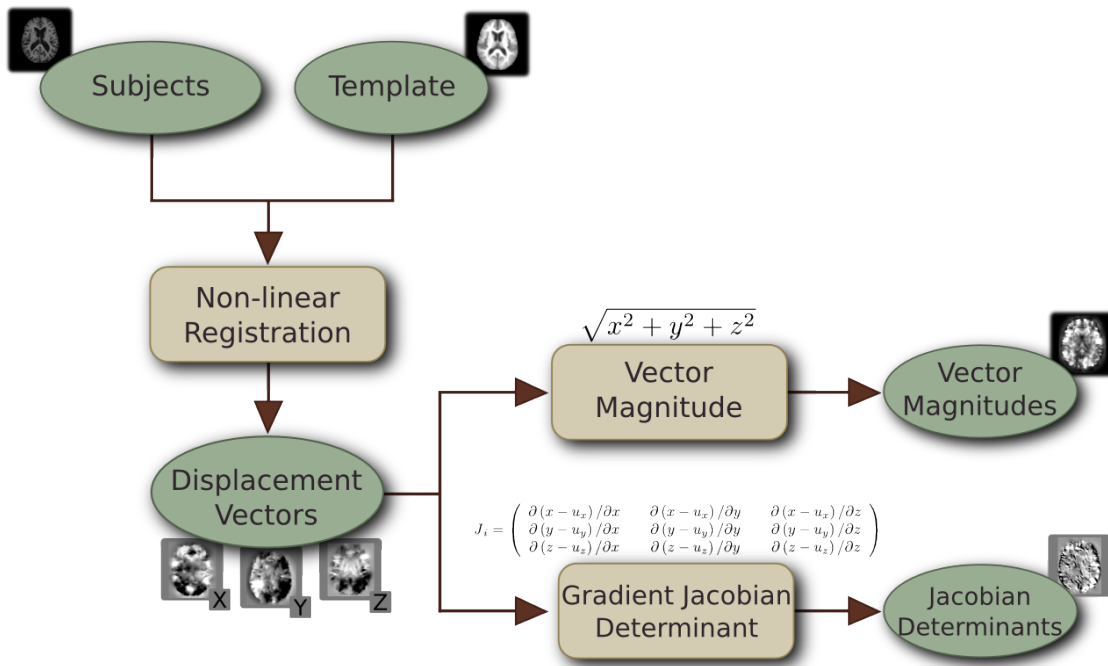


Figure 4.3: Pipeline of the computation of the deformation scalar measures.

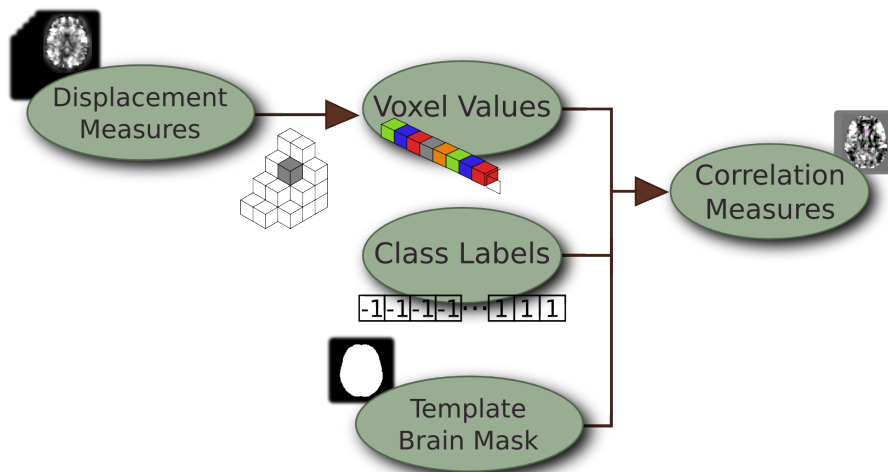


Figure 4.4: Pipeline of the calculation of the correlation volumes.

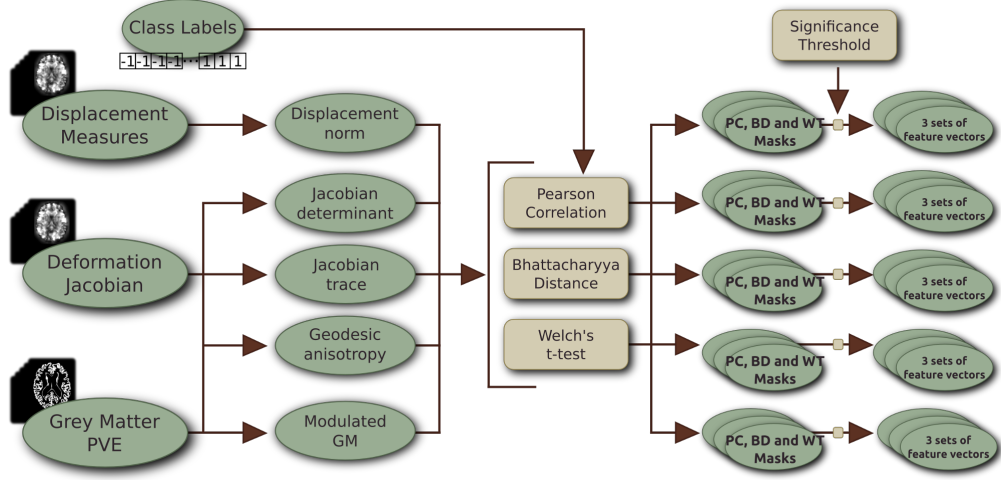


Figure 4.5: A summary of the pipelines of the feature selection methods. Each path from left to right specifies a feature selection process tested in the computational experiments.

Feature selection on the OASIS data The classification experiments reported in Chapter 6 on the OASIS database have been performed on a wide collection of feature vector specifications and selections. Figure 4.5 shows the a graph summarizing the diverse pipelines that lead to the actual feature vectors employed in the cross-validation experiments. From left to right, raw MRI data provide the deformation Jacobian and the grey matter segmentation that may be used to compute either the Jacobian maps or the modulated GM map. From these measures, computing Pearson’s correlation, Battacharyya’s separability measure or Welch’s t-test we obtain the significance maps. Significance thresholds set according to various percentiles of the significance map e.c.d.f we obtain the feature voxel sites where data is extracted to build the feature vectors. Tables 4.1, 4.2 and 4.3 contain the number of features extracted for each setting of the percentile threshold. Feature selection is performed in each cross-validation folder, thus the standard deviation of the number of features is provided. Only two of the five scalar deformation measures (modulated GM and Jacobian maps) are shown in these tables.

4.6 Scalar Features from Diffusion Tensors

In DTI, each voxel has an associated diffusion tensor defined as a 3×3 positive-definite symmetric matrix D , which can be represented by its decomposition as

$$D = \lambda_1 g_1 g_1^T + \lambda_2 g_2 g_2^T + \lambda_3 g_3 g_3^T, \quad (4.8)$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and g_1, g_2, g_3 are the eigenvalues and eigenvectors of D , respectively. Two scalar measures were extracted [27] from the voxels diffusion tensors:

| | jacs | modgm |
|-------|--------------------|--------------------|
| None | 812498.4 (6343.65) | 775086.8 (2318.37) |
| 80% | 141705.7 (3682.77) | 128598.3 (1567.37) |
| 90% | 82261.3 (1971.43) | 77540.0 (768.25) |
| 95% | 60427.6 (1020.87) | 57981.0 (413.33) |
| 99% | 45338.3 (238.56) | 44718.9 (178.46) |
| 99.5% | 43564.7 (174.22) | 43207.5 (160.54) |
| 99.9% | 42136.3 (123.47) | 42032.0 (130.96) |

Table 4.1: Pearson's correlation significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the feature selection to the OASIS data.

| | jacs | modgm |
|-------|-------------------|-------------------|
| None | 2102154.4 (0.49) | 2098251.5 (77.98) |
| 80% | 87644.8 (1775.03) | 62757.5 (516.61) |
| 90% | 62642.4 (772.39) | 50604.6 (204.75) |
| 95% | 51749.6 (403.02) | 45889.1 (83.38) |
| 99% | 43757.5 (164.41) | 42657.5 (32.21) |
| 99.5% | 42792.2 (151.90) | 42281.5 (32.17) |
| 99.9% | 42039.3 (156.62) | 41991.8 (25.25) |

Table 4.2: Bhattacharyya significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the deformation feature selection to the OASIS data.

| | jacs | modgm |
|-------|---------------------|---------------------|
| None | 1289656.4 (6343.79) | 1323163.8 (2313.34) |
| 80% | 84545.1 (1108.19) | 71914.6 (652.86) |
| 90% | 59979.4 (579.14) | 54066.6 (247.73) |
| 95% | 50228.2 (220.14) | 47383.8 (138.54) |
| 99% | 43432.1 (78.49) | 42904.0 (82.05) |
| 99.5% | 42637.7 (71.82) | 42397.6 (88.42) |
| 99.9% | 42004.6 (80.34) | 41987.7 (82.18) |

Table 4.3: Welch's t-test significance measure feature set sizes mean(std) for varying percentile threshold obtained applying the deformation feature selection to the OASIS data.

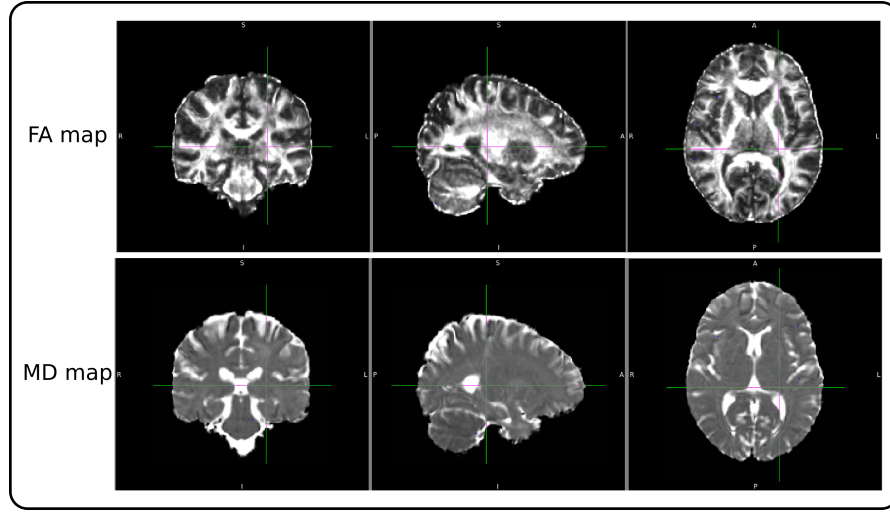


Figure 4.6: FA and MD maps of one subject.

- The mean diffusivity (MD) corresponding to the average eigenvalue:

$$MD = \frac{\text{Tr}(D)}{3} = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \quad (4.9)$$

- The fractional anisotropy (FA)

$$FA = \sqrt{\frac{1}{2} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}}. \quad (4.10)$$

The FA measures the fraction of the magnitude of D that can be related to anisotropic diffusion in a mean-squared sense (i.e. the extent of deviation from isotropic diffusivity in all direction). Its magnitude is also rotationally invariant, and independent from sorting of the eigenvalues. Thus, isotropic diffusion is imaged as zero value and FA maximum value is one. Figure 4.6 show slices of FA and MD volumes of one study subject.

The FA and MD volumes are subject to the feature selection and extraction described in Sections 4.1 and 4.2.

Image preprocessing Feature extraction requires that the diffusion related data is spatially normalized, in order to have the same anatomical correspondences for each voxel. Figure 4.7 illustrates the preprocessing pipeline to obtain the spatial normalization of FA and MD scalar maps. Our starting point was the nonlinear registration [16] of the T1-weighted MRI skull stripped volumes of each subject to the Montreal Neurological Institute (MNI152) standard

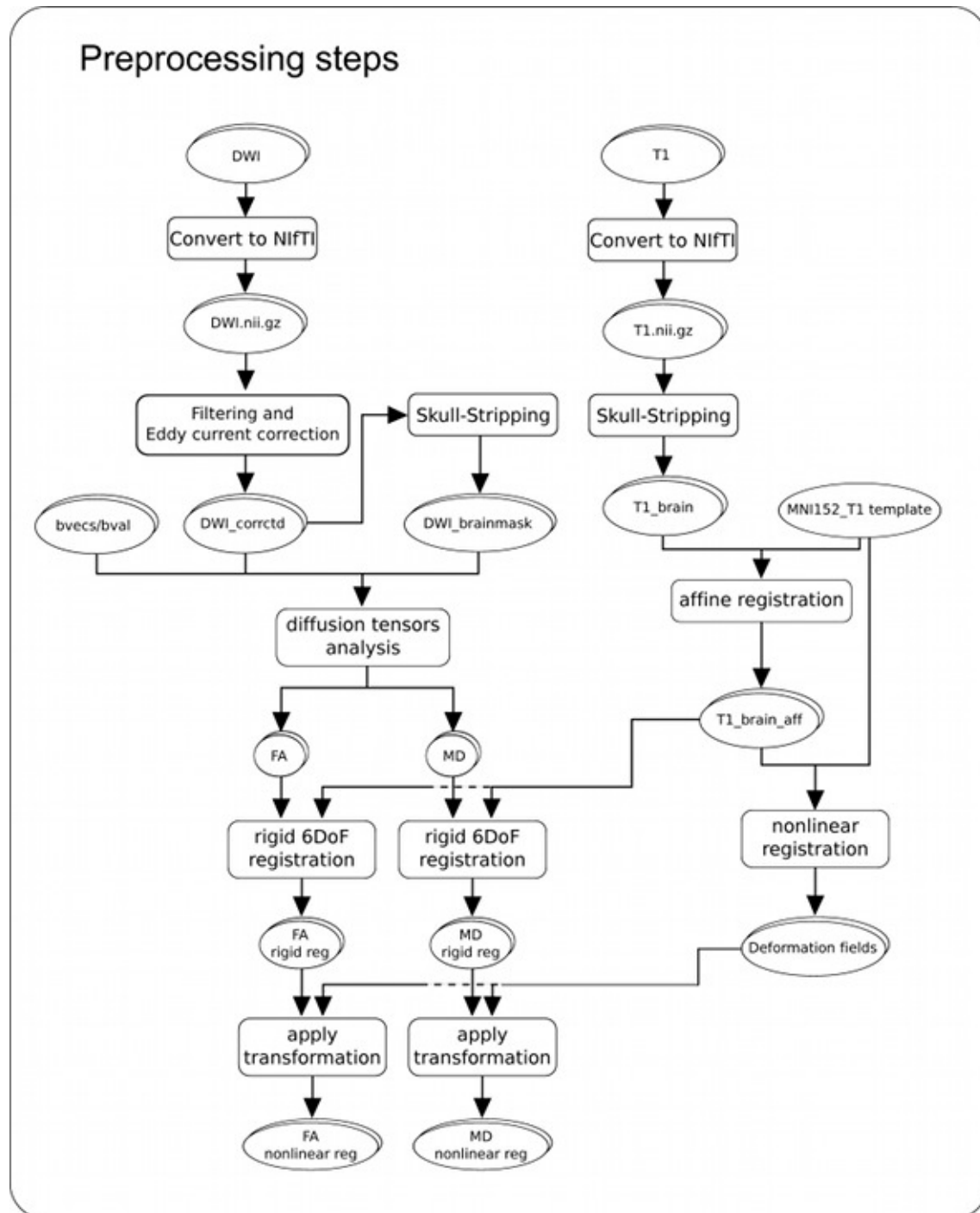


Figure 4.7: Image processing pipeline for the spatial normalization of the FA and MD scalar maps from DWI data.

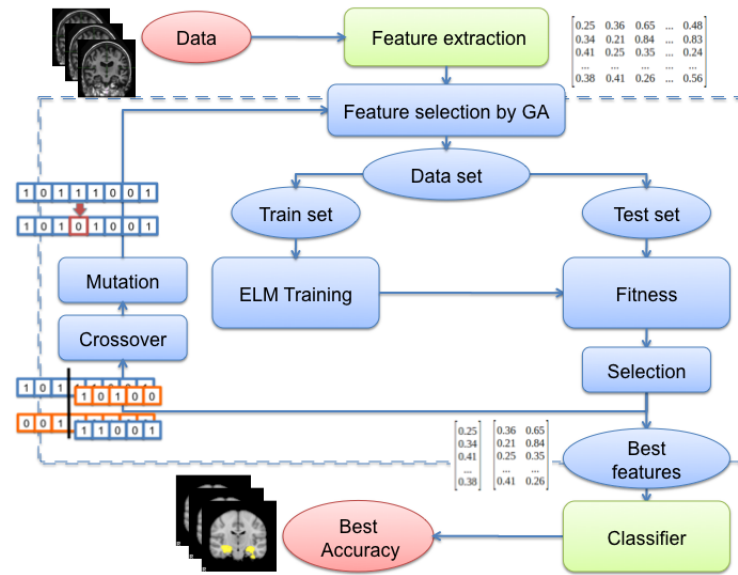


Figure 4.8: Evolutionary wrapper feature selection computational pipeline.

Algorithm 4.1 Evolutionary wrapper feature selection with ELM**Input** Raw data features extracted from the original data,**Output** Best features selected and accuracy**Initial** Generate random population**Apply** Genetic Algorithm for a fixed number of generations**Capture** statistics of feature activation**Estimate** feature saliency from GA feature statistics**Perform** cross-validation with estimated best features

template. The deformation fields of this registration were used afterwards for the spatial co-registration of the FA and MD volumes.

The DWI scans were already noise filtered and corrected for eddy currents and head motion by the group that originally acquired the scans. For image registration we use the ANTS (<http://www.picssl.upenn.edu/ANTS>) nonlinear elastic registration algorithm.

4.7 Evolutionary wrapper feature selection

Figure 4.8 and Algorithm 4.1 illustrate the computational workflow of the evolutionary wrapper feature selection algorithm presented in [52], where the raw features are the mean and stan-

standard deviations of the GM intensity values of the voxel intensities in the voxels belonging to the clusters detected by the VBM when performed on a subset of the OASIS database. The computational core of the process is a Genetic Algorithm (GA) [86] performing the search on the combinatorial space of feature selections. From the whole set of GA populations traversed during a realization of the evolutionary process, feature saliency is computed as the marginal probability of reaching a given accuracy threshold when the feature is selected. A final cross-validation experiment is performed on them.

Genetic Algorithm The general GA scheme is illustrated in Algorithm 4.2, with the following notation: N is the population size, T is the number of generations, K the number of crossings performed by generation, $H(t)$ the set of off-springs produced at each generation. The problem encoding is as follows: each individual is a binary valued vector $\mathbf{p} = \{p_j\}_{j=1}^n \in \{0, 1\}^n$ where $p_j = 0$ corresponds to the absence of the j -th feature and $p_j = 1$ to its presence in the set of features selected for classifier train and test. The population at generation t is a set of N individuals: $\mathbf{P}(t) = \{\mathbf{p}^i(t)\}_{i=1}^N$. In other words, each individual represents a feature selection, and the population a collection of feature selection hypothesis. Therefore, the problem falls in the classic category of binary valued individual GAs, so that standard mutation, crossover and selection operators can be applied from off-the-shelf implementations. In the wrapper feature selection process, the fitness function corresponds to the average accuracy provided by a 10-fold cross-validation process using the features indicated by coding of the corresponding individual. Let us denote the fitness of the i -th individual at generation t as $f_i(t) = \text{Acc} \mid_{\mathbf{p}^i(t)}$.

Salient feature selection After the realization of the GA we end up with an evolution trace consisting of a collection of individuals and corresponding fitness values: $\mathcal{T} = \{(\mathbf{p}^i(t), f_i(t)); i = 1, \dots, N; j = 1, \dots, T\}$. The saliency measure of each feature is intuitively defined as the probability of obtaining good accuracies when the feature is selected. To formalize this idea, let us define the j -th feature marginal probability distribution of the average accuracy as:

$$p(\text{Acc} \mid p_j = 1) = \frac{1}{N \cdot T} \left| \{f_i(t) \mid (f_i(t) = f) \wedge (\mathbf{p}_j^i(t) = 1)\} \right|. \quad (4.11)$$

Notice that the normalization factor is total number of experiment instances, therefore non-salient features will have a very low marginal probability mass. The selection of most salient features is accomplished by thresholding the accumulative marginal probability distribution. We select a feature as salient if

$$p(\text{Acc} > \theta \mid p_j = 1) > \varepsilon,$$

where selection parameters can be easily set by observation, because the difference between salient and non-salient features is sharp in practice. In the computational experiments reported in this paper $\theta = 90\%$ and $\varepsilon = 0.8$.

Algorithm 4.2 Canonical Genetic Algorithm

Let it be $P(0) = \{\mathbf{p}^1, \dots, \mathbf{p}^N\}$ the initial random population
for generations $t = 1, \dots, T$

1. for $k = 1, \dots, K$
 - (a) randomly select a pair of parents $(\mathbf{p}^i, \mathbf{p}^j)$
 - (b) obtain a pair off-springs $(\mathbf{h}^1, \mathbf{h}^2) = C(\mathbf{p}^i, \mathbf{p}^j)$
 - (c) apply mutation to each off-spring $(\mathbf{h}^{1'}, \mathbf{h}^{2'}) = M(\mathbf{h}^1, \mathbf{h}^2)$
 - (d) Add $(\mathbf{h}^{1'}, \mathbf{h}^{2'})$ to the set of $H(t)$ of off-springs in generation t
2. Compute the fitness value $f_i(t) = Acc|_{\mathbf{p}^i(t)}$ for each individual $\mathbf{p}^i \in P(t-1) \cup H(t)$
3. Select individuals for population $P(t)$ performing a roulette wheel selection on $P(t-1) \cup H(t)$ according to probabilities

$$s_i = \frac{f_i(t)}{\sum_{i=1}^{|P(t-1) \cup H(t)|} f_i(t)}.$$

Chapter 5

Classification

The final step in the computational pipeline is the classification of the data into corresponding classes. To demonstrate the predictive power of the classifiers, validation experiments must be carried out, where the issue of double dipping or circular analysis implies a methodological specification: training must be as independent as possible of the test data. The intuitive interpretation is clear: it is not possible to use for training unknown data that will arrive in the future [120]. However, during the validation phase, we have to separate both intentionally and carefully. In this chapter we give the description of the classifier building algorithms that have been tested along the thesis in diverse datasets. We recall their definition and comment on their properties at an intuitive level.

Section 5.1 gives the general problem statement, including the discussion of the validation issue. Section 5.2 describes Support Vector Machine (SVM) approaches, with some words on parameter tuning. Section 5.3 describes the Artificial Neural Network (ANN) architectures applied in the thesis: Multilayer Perceptron, Learning Vector Quantization, Radial Basis Function networks and Probabilistic Neural Networks. Sections 5.4 and 5.5 introduce two ensemble approaches: Random Forest and Bootstrapped Dendritic Computing. Section 5.6 introduces the Extreme Learning Machines and ensembles of them. Section 5.7 describes an AdaBoost of SVM classifiers. Finally, Section 5.8 describes some ensembles of independent classifiers.

5.1 Problem statement

We deal with two class classification problems, given a collection of training/testing input feature vectors $X = \{\mathbf{x}_i \in \mathbb{R}^n, i = 1, \dots, l\}$ and the corresponding labels $\{y_i \in \{-1, 1\}, i = 1, \dots, l\}$, which sometimes can be better denoted in aggregated form as a binary vector $\mathbf{y} \in \{-1, 1\}^l$. The algorithms described below build some classifier systems based on this data.

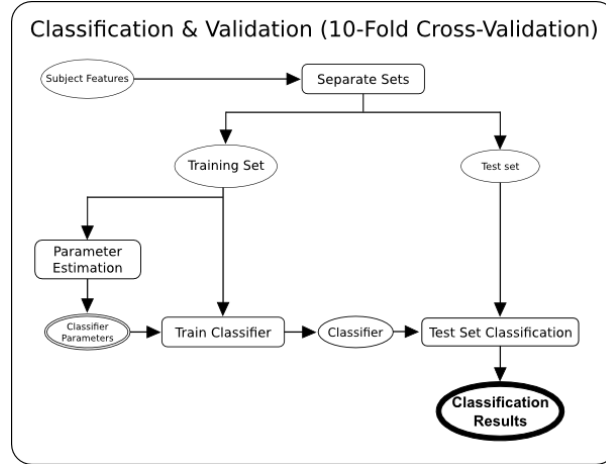


Figure 5.1: Flow chart of the 10-fold cross validation procedure followed in the experiments reported in this thesis.

The 1-NN classifier The simplest algorithm is the 1-NN which involves no parameter tuning and uses all the training data samples. The classification rule is of the form:

$$c(\mathbf{x}) = y_{i^*} \text{ where } i^* = \arg \min_{i=1, \dots, l} \{\|\mathbf{x} - \mathbf{x}_i\|\},$$

that is, the assigned class is that of the closest training vector.

Parameter tuning and validation A critical procedure from the statistical validation of the classification results is the strict separation between test and train process, so that no information from the testing data (unknown in real life) can be used in the training of the classifiers. In figure 5.1 we show the flow of the data for each procedure. The training set is used both for parameter tuning and for the training of the final classifier. In some instances, a small subset of the training dataset can be used for parameter tuning, while training of the selected classifier can be performed on the remaining data or on the whole training set. The statistical effect of these choices is not clarified.

5.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) [203] look for the set of support vectors that allow to build the optimal discriminating surface in the sense of providing the greatest margin between the classes¹. Support Vector Machines (SVMs) have attracted attention from the pattern recognition community [81, 197] owing to a number of theoretical and computational merits derived from

¹The implementation used in most studies is included in the libSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) software package. The implementation is described in detail in [46].

[202]. SVM separates a given set of binary labelled training data with a hyperplane that is maximally distant from the two classes (known as the maximal margin hyperplane). The objective is to build a discriminating function using training data that will correctly classify new examples (\mathbf{x}, y) . When no linear separation of the training data is possible, SVMs can work effectively in combination with kernel techniques using the kernel trick, so that the hyperplane defining the SVMs corresponds to a non-linear decision boundary in the input space that is mapped to a linearised higher-dimensional space [202]. In this way, the decision function can be expressed in terms of the support vectors only:

$$f(\mathbf{x}) = \text{sign} \left(\sum \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + w_0 \right),$$

where $K(.,.)$ is a kernel function, α_i is a weight constant derived from the SVM process and the \mathbf{s}_i are the support vectors [202].

Given training vectors $\mathbf{x}_i \in R_n, i = 1, \dots, l$ of the subject features of the two classes, and a vector $y \in R^l$ such that $y_i \in \{-1, 1\}$ labels each subject with its class, in our case, for example, patients were labelled as -1 and control subject as 1. To maximize the classification margin, the SVM approach solves the following optimization problem:

$$\min_{w, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i$$

subject to $y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq (1 - \xi_i)$, $\xi_i \geq 0, i = 1, 2, \dots, n$. The dual optimization problem is

$$\min_{\alpha} \frac{1}{2} \alpha^T \mathbf{Q} \alpha - \mathbf{e}^T \alpha,$$

subject to $\mathbf{y}^T \alpha = 0$, $0 \leq \alpha_i \leq C, i = 1, \dots, l$, where \mathbf{e} is the vector of all ones, $C > 0$ is the upper bound on the error, \mathbf{Q} is an $l \times l$ positive semi-definite matrix, $Q_{ij} \equiv y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$, and $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function that describes the behaviour of the support vectors. Here, the training vectors \mathbf{x}_i are mapped into a higher (maybe infinite) dimensional space by the function $\phi(\mathbf{x}_i)$. C is a regularization parameter used to balance the model complexity and the training error.

The kernel function chosen results in different kinds of SVM with different performance levels, and the choice of the appropriate kernel for a specific application is a difficult task. In this study two different kernels were tested: the linear and the radial basis function (RBF) kernel. The linear kernel function is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = 1 + \mathbf{x}_i^T \mathbf{x}_j,$$

this kernel shows good performance for linearly separable data. The RBF kernel is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right).$$

This kernel is best suited to deal with data that have a class-conditional probability distribution function approaching the Gaussian distribution [40]. The RBF kernel is largely used in the literature because it corresponds to the mapping into an infinite dimension feature space, and it can be tuned by its variance parameter σ .

SVM with non-linear performance measures as objective functions The SVM training algorithm can be generalized to optimize non-linear performance measures treating the learning problem as a multivariate prediction problem. Its implementation is included in the SVM-Perf² software package [110]. Defining the hypotheses functions \bar{h} that map a tuple $\bar{\mathbf{x}} \in \bar{\mathcal{X}}$ of n feature vectors $\bar{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ to a tuple $\bar{y} \in \bar{\mathcal{Y}}$ of n labels $\bar{y} = (y_1, \dots, y_n)$

$$\bar{h}: \bar{\mathcal{X}} \rightarrow \bar{\mathcal{Y}},$$

where $\bar{\mathcal{X}} = \mathcal{X} \times \dots \times X$ and $\bar{\mathcal{Y}} \subseteq \{-1, +1\}^n$ is the set of all admissible label vectors. A linear discriminant function is used for this multivariate mapping:

$$\bar{h}_{\mathbf{w}}(\bar{x}) = \arg \max_{\bar{y}' \in \bar{\mathcal{Y}}} \{ \mathbf{w}^T \Psi(\bar{\mathbf{x}}, \bar{y}') \},$$

where \mathbf{w} is a parameter vector and Ψ is a function that returns a feature vector describing the match between $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ and (y'_1, \dots, y'_n) . Ψ is restricted to be of the following simple form:

$$\Psi(\bar{\mathbf{x}}, \bar{y}') = \sum_{i=1}^n y'_i \mathbf{x}_i.$$

The prediction rule $\bar{h}_{\mathbf{w}}(\bar{\mathbf{x}})$ returns the tuple of labels \bar{y}' which scores highest according to a linear function.

From this, the multivariate SVM optimization problem is formulated as:

$$\begin{aligned} \min_{\mathbf{w}, \xi \geq 0} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C\xi \\ \text{s.t.} \quad & \forall \bar{y}' \in \bar{\mathcal{Y}} \setminus \bar{y}: \mathbf{w}^T [\Psi(\bar{\mathbf{x}}, \bar{y}) - \Psi(\bar{\mathbf{x}}, \bar{y}')] \geq \Delta(\bar{y}', \bar{y}) - \xi \end{aligned}$$

where Δ is the training loss function and its upper bound ξ , also called slack variable. This formulation allows to apply the training procedure to different loss functions based on the contingency tables. Efficient algorithms to perform this training in polynomial time are given in [110].

Grid search for SVM tuning Parameter setting for SVM is critical to obtain good performance. Thus, each learning experiment is wrapped with a grid search procedure [134]. Every time an algorithm is trained on a data set T , an internal stratified two-fold cross-validation (repeated 3 times) is performed for each combination of parameter values, looking for the best

²http://svmlight.joachims.org/svm_perf.html

parameter setting for T . The grid search mechanism selects the combination of parameters maximizing the F1-score for the SVM minimizing the error rate. The values examined for the regularization parameter are $C \in \{10^{-2}, 3 \cdot 10^{-2}, 10^{-1}, 3 \cdot 10^{-1}, 1\}$, the values for the RBF kernel parameter σ ranged from 10^{-3} to 1, varying the exponents in unit steps.

5.3 Artificial Neural Networks (ANN)

Artificial Neural Networks [169, 98, 93] introduced quantitative training algorithms following a biological inspiration, thus revolutionizing the field of Artificial Intelligence by shifting from logic based reasoning approaches to a numerical and statistics paradigm, which allows quantitative computational experiments for validation of the learned systems.

5.3.1 Multi Layer Perceptron trained with Backpropagation

Backward propagation of errors, or backpropagation (BP), [169, 98, 93] is a non-linear generalization of the squared error gradient descent learning rule for updating the weights of artificial neurons in a single-layer perceptron, which is extended to feed-forward network architectures, aka Multi-Layer Perceptron (MLP), despite the fact that they do not fit into the computational framework of the Perceptron. Backpropagation requires that the activation function used by the artificial neurons (or "nodes") is differentiable with its derivative being a simple function of itself. The backpropagation of the error allows to compute the gradient of the error function relative to the hidden units. It is analytically derived using the chain rule of calculus. During on-line learning, the weights of the network are updated at each input data item presentation. We have used the resilient backpropagation, which uses only the derivative sign to perform the weight updating.

We restrict our presentation of BP to train the weights of the MLP for the current two class problem. Let the instantaneous error E_p be defined as:

$$E_p(\mathbf{w}) = \frac{1}{2} (y_p - z_K(\mathbf{x}_p))^2, \quad (5.1)$$

where y_p is the p -th desired output y_p , and $z_K(\mathbf{x}_p)$ is the network output when the p -th training exemplar x_p is input to the MLP composed of K layers, whose weights are aggregated in the vector \mathbf{w} . The output of the j -th node in layer k is given by:

$$z_{k,j}(\mathbf{x}_p) = f \left(\sum_{i=0}^{N_{k-1}} w_{k,j,i} z_{k-1,i}(\mathbf{x}_p) \right), \quad (5.2)$$

where $z_{k,j}$ is the output of node j in layer k , N_k is the number of nodes in layer k , $w_{k,j,i}$ is the weight which connects the i -th node in layer $k-1$ to the j -th node in layer k , and $f(\cdot)$ is the sigmoid non-linear function, which has a simple derivative:

$$f'(\alpha) = \frac{df(\alpha)}{d\alpha} = f(\alpha)(1 - f(\alpha)). \quad (5.3)$$

The convention is that $z_{0,j}(\mathbf{x}_p) = \mathbf{x}_{p,j}$. Let the total error E_T be defined as follows:

$$E_T(\mathbf{w}) = \sum_{p=1}^l E_p(\mathbf{w}), \quad (5.4)$$

where l is the cardinality of X . Note that E_T is a function of both the training set and the weights in the network. The backpropagation learning rule is defined as follows:

$$\Delta w(t) = -\eta \frac{\partial E_p(\mathbf{w})}{\partial w} + \alpha \Delta w(t-1), \quad (5.5)$$

where $0 < \eta < 1$, which is the learning rate, the momentum factor α is also a small positive number, and w represents any single weight in the network. In the above equation, $\Delta w(t)$ is the change in the weight computed at time t . The momentum term is sometimes used ($\alpha \neq 0$) to improve the smooth convergence of the algorithm. The algorithm defined by equation (5.5) is often termed as *instantaneous backpropagation* because it computes the gradient based on a single training vector. Another variation is *batch backpropagation*, which computes the weight update using the gradient based on the total error E_T .

To implement this algorithm we must give an expression for the partial derivative of E_p with respect to each weight in the network. For an arbitrary weight in layer k this can be written using the Chain Rule:

$$\frac{\partial E_p(\mathbf{w})}{\partial w_{k,j,j}} = \frac{\partial E_p(\mathbf{w})}{\partial z_{k,j}(\mathbf{x}_p)} \frac{\partial z_{k,j}(\mathbf{x}_p)}{\partial w_{k,j,i}}. \quad (5.6)$$

Because the derivative of the activation function follows equation 5.3, we get:

$$\frac{\partial z_{k,j}(\mathbf{x}_p)}{\partial w_{k,j,i}} = z_{k,j}(\mathbf{x}_p) (1 - z_{k,j}(\mathbf{x}_p)) z_{k-1,j}(\mathbf{x}_p), \quad (5.7)$$

and

$$\frac{\partial E_p(\mathbf{w})}{\partial z_{k,j}(\mathbf{x}_p)} = \sum_{m=1}^{N_{k+1}} \frac{\partial E_p(\mathbf{w})}{\partial z_{k+1,m}(\mathbf{x}_p)} z_{k+1,m}(\mathbf{x}_p) (1 - z_{k+1,m}(\mathbf{x}_p)) w_{k+1,m,j},$$

which at the output layer corresponds to the output error :

$$\frac{\partial E_p(\mathbf{w})}{\partial z_K(\mathbf{x}_p)} = z_L(\mathbf{x}_p) - y_p. \quad (5.8)$$

5.3.2 Radial Basis Function Networks

Radial Basis Function networks (RBF) [49, 98] are a type of ANN whose activation function is a radial basis function. RBFs consist of a two layer neural network, where each hidden unit

implements a radial activated function. The output units compute a weighted sum of hidden unit outputs. Training consists of an unsupervised training of the hidden units followed by the supervised training of the output units weights. RBFs have their origin in the solution of a multivariate interpolation problem [38]. Arbitrary function $g(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ can be approximated by a map defined by a RBF network with a single hidden layer of K units:

$$\hat{g}_{\theta}(\mathbf{x}) = \sum_{j=1}^K w_j \phi(\sigma_j, \|\mathbf{x} - \mathbf{c}_j\|), \quad (5.9)$$

where θ is the vector of RBF parameters including $w_j, \sigma_j \in \mathbb{R}$, and $\mathbf{c}_j \in \mathbb{R}^n$; let us denote $\mathbf{w} = (w_1, w_2, \dots, w_p)^T$, then the vector of RBF parameters can be expressed as $\theta^T = (\mathbf{w}^T, \sigma_1, \mathbf{c}_1^T, \dots, \sigma_K, \mathbf{c}_K^T)$. Each RBF is defined by its centre $\mathbf{c}_j \in \mathbb{R}^n$ and width $\sigma_j \in \mathbb{R}$, and the contribution of each RBF to the network output is weighted by w_j . The RBF function $\phi(\cdot)$ is a non-linear function that monotonically decreases as \mathbf{x} moves away from its centre \mathbf{c}_j . The most common RBF used is the isotropic Gaussian:

$$\hat{g}_{\theta}(\mathbf{x}) = \sum_{j=1}^p w_j \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}_j\|^2}{2\sigma_j^2}\right).$$

The network can be thought as the composition of two functions

$$\hat{g}_{\theta}(\mathbf{x}) = W \circ \Phi(\mathbf{x}),$$

the first one implemented by the RBF units $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^K$ performs a data space transformation which can be a dimensionality reduction or not, depending on whether $K > n$. The second function corresponds to a single layer linear Perceptron $W : \mathbb{R}^K \rightarrow \mathbb{R}$ giving the map of the RBF transformed data into the class labels. Training is accordingly decomposed into two phases. First a clustering algorithm is used to estimate the Gaussian RBF parameters (centres and variances). Afterwards, linear supervised training is used to estimate the weights from the hidden RBF to the output. In order to obtain a binary class label output, a hard limiter function is applied to the continuous output of the RBF network.

5.3.3 Probabilistic Neural Networks

Probabilistic Neural Networks (PNN) [190] use a kernel-based approximation to form an estimate of the probability density function of categories in a classification problem. In fact, it is a generalization of the Parzen windows distribution estimation, and a filtered version of the 1-NN classifier. The distance of the input feature vector \mathbf{x} to the stored patterns is filtered by a RBF function. Let us denote the data sample partition as $X = X_1 \cup X_{-1}$, where $X_1 = \{\mathbf{x}_1^1, \dots, \mathbf{x}_{n_1}^1\}$ and $X_{-1} = \{\mathbf{x}_1^{-1}, \dots, \mathbf{x}_{n_{-1}}^{-1}\}$. That is, superscripts denote the class of the feature vector and $n_1 + n_{-1} = n$. Each pattern \mathbf{x}_j^i of training data sample is interpreted as the weight of the j -th neuron of the i -th class. Therefore the response of the neuron is computed as the probability of

the input feature vector according to a Normal distribution centred at the stored pattern:

$$\Phi_{i,j}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sigma^n} \exp \left[-\frac{\|\mathbf{x} - \mathbf{x}_j^i\|^2}{2\sigma^2} \right], \quad (5.10)$$

Therefore the output of the neuron is inside $[0, 1]$. The tuning of a PNN network depends on selecting the optimal sigma value of the spread σ of the RBF functions, which can be different for each class. In this paper an exhaustive search for the optimal spread value in the range $(0, 1)$ for each training set has been done. The output of the PNN is an estimation of the likelihood of the input pattern \mathbf{x} being from class $i \in \{-1, 1\}$ by averaging the output of all neurons that belong to the same class:

$$p_i(x) = \frac{1}{n_i} \sum_{j=1}^{n_i} \Phi_{i,j}(\mathbf{x}). \quad (5.11)$$

The decision rule based on the output of all the output layer neurons is simply:

$$\hat{y}(\mathbf{x}) = \arg \max_i \{p_i(\mathbf{x})\}, \quad i \in \{-1, 1\}. \quad (5.12)$$

where $\hat{y}(\mathbf{x})$ denotes the estimated class of the pattern \mathbf{x} . If the a priori probabilities for each class are the same, and the losses associated with making an incorrect decision for each class are the same, the decision layer unit classifies the pattern \mathbf{x} in accordance with the optimal Bayesian rule.

5.3.4 Learning Vector Quantization

Learning vector quantization (LVQ) [118, 189] provides a method for training competitive networks in a supervised manner. The system is composed of an unsupervisedly trained competitive layer which performs a partitioning of the input space. The supervisedly trained output layer provides the labelling of the input data according to its belonging to an input region (crisp clustering) or to its degree of membership (soft clustering). In the original proposition of the LVQ, the competitive units were cluster centres with the Euclidean distance as the similitude measure. Training of the competitive units can be performed by Kohonen's Self Organizing Map. Supervised training was simply the assignment of a label to a competitive unit according to a majority voting on the data samples falling in the partition corresponding to the unit. LVQ provides fine tuning of the competitive units using class information. The basic versions proposed by Kohonen are known as the LVQ1 and LVQ2. Both start with the unsupervised learning of the competitive units, and its initial majority voting labelling. In the LVQ1 a supervised training is performed as follows: for each data sample we compare its label with the one of its corresponding competitive unit, if the labels match (the data item is correctly classified) then the competitive unit is moved towards the input data sample, otherwise it is moved in the opposite direction. This rule

may cause an unstable and oscillatory behaviour if the discriminant boundary among classes is very complex. The LVQ2 approach involves determining the two input vector's closest weights. They are moved toward or away the input according to the matching of their classes.

LVQ1 LVQ as introduced by Kohonen [119] represents every class $c \in \{-1, 1\}$ by a set $W(c) = \{\mathbf{w}_i \in \mathbb{R}^n; i = 1, \dots, N_c\}$ of weight vectors (prototypes) which tessellate the input feature space. Let us denote W the union of all prototypes, regardless of class. If we denote c_i the class the weight vector $\mathbf{w}_i \in W$ is associated with, the decision rule that classifies a feature vector \mathbf{x} is as follows:

$$c(\mathbf{x}) = c_{i^*}$$

where

$$i^* = \arg \min_i \{\|\mathbf{x} - \mathbf{w}_i\|\}.$$

The training algorithm of LVQ aims at minimizing the classification error on the given training set, i.e., $E = \sum_j (y_j - c(\mathbf{x}_j))^2$, modifying the weight vectors on the presentation of input feature vectors. The heuristic weight updating rule is as follows:

$$\Delta \mathbf{w}_{i^*} = \begin{cases} \varepsilon \cdot (\mathbf{x}_j - \mathbf{w}_{i^*}) & \text{if } c_{i^*} = y_j \\ -\varepsilon \cdot (\mathbf{x}_j - \mathbf{w}_{i^*}) & \text{otherwise} \end{cases}, \quad (5.13)$$

that is, the input's closest weight is adapted either toward the input if their classes match, or away from it if not. This rule is highly unstable, therefore, the practical approach consists in performing an initial clustering of each class data samples to obtain an initial weight configuration using equation 5.13 to perform the fine tuning of the classification boundaries.

LVQ2 The LVQ2 rule is proposed to improve the learning, sometimes it is recommended to apply it *after* the LVQ1. In LVQ2, for each input data sample we find the two closest competitive units. If one correctly classifies the input and the other belongs to a wrong class, and the input data lies in a window around the mid-plane between them, then the correct class unit is moved towards the input and the incorrect unit is moved away from the input. We have used the simplest implementations.

5.4 Random Forests (RF)

The random forests (RF) algorithm is a classifier [35] that encompasses bagging [36] and random decision trees [9, 99], being used in a variety of applications [21]. RF became popular due to its simplicity of training and tuning while offering a similar performance to boosting. Consider a RF collection of tree predictors, that is, a RF is a large collection of decorrelated decision trees

$$h(\mathbf{x}; \psi_t), t = 1, \dots, T,$$

where \mathbf{x} is a d -dimensional random sample of random vector \mathbf{X} , ψ_t are independent identically distributed random vectors whose nature depends on their use in the tree construction, and each tree casts a unit vote for the most popular class of input \mathbf{x} . RF capture complex interaction structures in data, and are supposed to be resistant to over-fitting of data if individual trees are sufficiently deep.

Given a dataset of N samples, a bootstrapped training dataset is used to grow tree $h(\mathbf{x}; \psi_t)$ by recursively selecting a random subset of data dimensions \hat{d} such that $\hat{d} \ll d$ and picking the best split of each node based on these variables. Unlike conventional decision trees, pruning is not required. The independent identically distributed random vectors ψ_t determine the random dimension selection.

The trained RF can be used for classification of a new input \mathbf{x} by majority vote among the is the class prediction of the RF trees $C_u(x)$. The critical parameters of the RF classifier for the experiments reported below are set as follows. The number of trees in the forest should be sufficiently large to ensure that each input class receives a number of predictions.

5.5 Bootstrapped Dendritic Classifiers

Bootstrapped Dendritic Classifiers (BDC) is a collection of dendritic classifiers [51], $C(x; \psi_j)$, $j = 1, \dots, N$, where ψ_j are independent identically distributed random vectors whose nature depends on their use in the classifier construction. Each DC classifier casts a unit vote for the most popular class of input x . Given a dataset of n samples, a bootstrapped training dataset is used to train $C(x; \psi_j)$. The independent identically distributed random vectors ψ_j determine the result of bootstrapping. In conventional RF they also determine the subset of data dimensions \hat{d} such that $\hat{d} \ll d$ on which each tree is grown, in this paper we are not dealing with this kind of DC randomization, which will be studied elsewhere. The main parameters for the experimental evaluation of the BDC are the number of trees and the maximum depth of each DC given by the maximum number of dendrites allowed. Limiting the number of dendrites is a kind of regularization that weakens the classifier.

5.6 Extreme Learning Machines

Extreme Learning Machines (ELM) [104] is a fast training approach for the single layer feed-forward neural networks (SFLN), providing on average good quality classification and regression results whose computation burden is orders of magnitude lower than conventional back-propagation training. This method is based on the Moore-Penrose generalized inverse providing the minimum Least-Squares solution of general linear systems. Universal approximation properties have been proven for ELM [103]. In the short time since their proposal, ELMs have been successfully applied to a large number of problems such as face recognition [140]. However, an ELM major criticism is the uncertainty of its performance due the random generation of the

hidden layer weights. An approach to overcome that problem is the composition of elementary classifiers into ensembles, such as the Voting ELM (V-ELM) [43], which is a direct composition by majority voting of a collection of ELMs trained independently.

5.6.1 Basic ELM

For N arbitrary distinct samples (\mathbf{x}_i, t_i) , where input variables are $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ and target values are $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbb{R}^m$. The training of a standard SLFN with N hidden neurons and activation function $g(x)$ is mathematically modelled as solving the following equation to estimate the value of the SLFN parameters:

$$\sum_{i=1}^{hn} \beta_i \cdot g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, \dots, N. \quad (5.14)$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight vector connecting the i -th hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector connecting the i -th hidden neuron and the output neurons, and b_i is the threshold of the i -th hidden neuron. $\mathbf{w}_i \cdot \mathbf{x}_j$ denotes the inner product of \mathbf{w}_i and \mathbf{x}_j and hn is the number of hidden neurons. The activation function can be the identity for the so-called linear kernel approaches, sigmoid for the Multilayer Perceptron approaches, or Gaussian for Radial Basis Function approaches [215].

The equation (5.14) can be written in matrix form as:

$$\mathbf{H}\beta = \mathbf{T}, \quad (5.15)$$

where \mathbf{H} , of size $N \times hn$, is the output matrix resulting of the SLFN hidden layer activated by the input samples, β is the output weight matrix of size $hn \times m$, and \mathbf{T} is the target matrix with size $N \times m$. Training of SLFN is accomplished by computing the least-squares solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$, given by $\hat{\beta} = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moore-Penrose inverse of \mathbf{H} .

5.6.2 Ensembles of ELM

Ensembles of classifiers improve generalization performance by combining weak and diverse classifiers. Ensembles of ELM are trained on bootstrapped samples of the train data setting a limit on the number of hidden neurons in each classifiers, and their output is combined by majority voting. Such approach has been called V-ELM [43]. By construction, ELM with little hidden units are weak and diverse classifiers, so V-ELM reaps the maximum benefit from the ensemble approach. The V-ELM is a collection of ELMs,

$$C(x; \psi_j), j = 1, \dots, N,$$

where ψ_j are independent identically distributed random vectors denoting the random weight of hidden units. Each ELM classifier casts a unit vote for the most popular class of input x .

Algorithm 5.1 Cross-validation scheme for training the ELM ensembles

Let be $X = \{x_1, \dots, x_n\}$ input data $x_i \in \mathbb{R}^n$, and $T = \{t_1, \dots, t_n\}$ the input data class labels $t_i \in \{0, 1\}$, activation function $g(x)$, hidden neuron number N , number of ELM classifiers M .

1. for $i = 1 : 10$ (cross-validation folds)
 - (a) select disjoint train $X^e = \{x_{i_1}^e, \dots, x_{i_{n-n/10}}^e\} \subset X$, $T^e = \{t_{i_1}^e, \dots, t_{i_{n-n/10}}^e\} \subset T$ and test $X^s = \{x_{i_1}^s, \dots, x_{i_{n/10}}^s\} \subset X$, $T^s = \{t_{i_1}^s, \dots, t_{i_{n/10}}^s\} \subset T$ datasets .
 - (b) For $j = 1 : M$ (construct of classifiers)
 - i. Bootstrap a train dataset $X^{eb} = \{x_{i_1}^{eb}, \dots, x_{i_{n-2n/10}}^{eb}\} \subset X^e$, $T^{eb} = \{t_{i_1}^{eb}, \dots, t_{i_{n-2n/10}}^{eb}\} \subset T^e$. Out-of-bag error may be computed on the remaining training data and test $X^e - X^{eb}$, $T^e - T^{eb}$, disjunctions.
 - ii. Apply ELM to train classifier $C_j : \mathbb{R}^n \rightarrow \{0, 1\}$ on (X^{eb}, Y^{eb}) .
 - (c) end for. Optionally compute out of bag error
 - (d) Cross-validation test, For each $x \in X^s$
 - i. compute $C_1(g(x), x), \dots, C_M(g(x), x)$
 - ii. Majority voting, class $t = 0$ if $|\{j | C_j(g(x), x) = 0\}| > |\{j | C_j(g(x), x) = 1\}|$
 - (e) compute accuracy, sensitivity and specificity statistics
2. end fold i

Given a dataset of n samples, a bootstrapped training dataset is used to train $C(x; \psi_j)$. The main parameters for the experimental evaluation of the V-ELM are the number of ELMs and of hidden units. Finally, Algorithm 5.1 specifies the training and cross-validation scheme applied in the experiments.

5.6.3 Hybrid Extreme Rotation Forest

The Hybrid Extreme Rotation Forest (HERF) algorithm [17] is summarized as follows: Let $\mathbf{x} = [x_1, \dots, x_n]^T$ be a random vector composed of n feature variables, F is the feature variable set and X is the data set containing N data samples in a matrix of size $N \times n$. The i -th row of X , denoted by $X_i = [x_i^1, x_i^2, \dots, x_i^N]$ is the sample vector of the i -th feature variable F_i . Let Y be a vector containing the class labels of the data samples, $Y = [y_1, \dots, y_N]^T$, $y_i \in \Omega$. Denote by D_1, \dots, D_L the classifiers in the ensemble. As with most ensemble methods, we need to specify L in advance. All classifiers can be trained in parallel. To construct the training set for classifier D_i , we carry out the following steps:

1. Partition the set of feature variables F into K subsets of variables.

Algorithm 5.2 Diverse AdaBoostSVM

-
1. **Input:** a set of training samples with labels $\{(x_1, y_1), \dots, (x_N, y_N)\}$; the initial σ , σ_{ini} ; the minimal σ , σ_{min} ; the step of σ , σ_{step} ; the threshold on diversity DIV .
 2. **Initialize:** the weights of training samples: $w_i^t = 1/N$, for all $i = 1, \dots, N$
 3. **Do while** ($\sigma > \sigma_{ini}$)
 - (a) Calculate gamma: $\gamma = (2\sigma^2)^{-1}$.
 - (b) Use σ to train a component classifier h_t on the weighted training set.
 - (c) Calculate the training error of h_t : $\varepsilon_t = \sum_{i=1}^N w_i^t$, $y_i \neq h_t(x_i)$.
 - (d) Calculate the diversity of h_t : $D_t = \sum_{i=1}^N d_t(x_i)$, where $d_t(x_i) = \begin{cases} 0 & \text{if } h_t(x_i) = y_i \\ 1 & \text{if } h_t(x_i) \neq y_i \end{cases}$
 - (e) Calculate the diversity of weighted component classifiers and the current classifier: $D = \sum_{t=1}^T \sum_{i=1}^N d_t(x_i)$.
 - (f) If $\varepsilon_t > 0.5$ or $D < DIV$: decrease σ by σ_{step} and go to (a).
 - (g) Set weight of the component classifier h_t : $\alpha_t = \frac{1}{2} \ln(\frac{\varepsilon_t}{1-\varepsilon_t})$.
 - (h) Update the weights of training samples: $w_i^{t+1} = w_i^t \exp(-\alpha_t y_i h_t(x_i))$.
 - (i) Normalize the weights of training samples: $w_i^{t+1} = w_i^t \exp(-\alpha_t y_i h_t(x_i)) / (\sum_{i=1}^N w_i^t \exp(-\alpha_t y_i h_t(x_i)))$.
 4. **Output:** $f(x) = \text{sign}(\sum_{k=1}^C \alpha_k h_k(x))$.
-
2. For each subset of feature variables, extract the data from the training data set, and compute the rotation matrix performing the Principal Component Analysis (PCA) transformation.
 3. Compose the global rotation matrix reordering columns according to the original data.
 4. Transform the train and test data applying the same rotation matrix.

5.7 Adaptive Boosting

Adaptive Boosting (AdaBoost)[180, 77] is a meta-algorithm for machine learning that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classifiers are tweaked in favour of those instances misclassified by previous classifiers. AdaBoost is sensitive to noisy data and outliers. Otherwise, it is less susceptible to the over-fitting problem than most learning algorithms.

AdaBoost trains a weak classifier in a series of rounds $t = 1, \dots, T$. For each iteration the distribution of sample weights W_t is updated, indicating the importance of examples in the data

set for the classification. On each round, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples.

Following these ideas, we have also tested a combination of SVM classifiers along the ideas from the Diverse AdaBoost SVM [129], presented as Algorithm 5.7. In this approach we built a sequence of SVM classifiers of increasing variance parameter. The results of the classifiers are weighted according to their statistical error to obtain the response to the test inputs in the 10-fold validation process.

5.8 Ensemble of independent classifiers

In some instances we train independent classifiers on specific variables, as kind of generalization of the Naive Bayes approach. For instance, we use this approach on a set of features extracted from brain ROIs selected from the AAL atlas, described in section 4.4. For each ROI we train a separate classifier, which can be a single or ensemble classifier. The meta-ensemble classification performs a majority voting on the results of the ROI classifiers. To care for the effect of class unbalance we assume a threshold on the majority required for a decision, that is, the minimum number of ROIs that must agree in order to decide on the most frequent class. This majority threshold acts as a correction for the *a priori* probability (which is unknown to the model selection algorithm).

5.8.1 Model selection and validation

Algorithm 5.3 specifies the experimental design for the validation of the approach, including the model selection and validation for each independent classifier. Model selection is performed on a 20% of the data holding out 80% for validation. Model selection aims to select the subset of ROIs giving best classification performance and the majority threshold value. Model selection proceeds as follows, first the classifier performance on each ROI features is estimated by a leave-one-out (LOO) procedure. Second, the ROIs are ordered according to their independent performances. Third, a greedy search for the optimal subset of ROIs is performed, adding them according to the previous order and testing their LOO performance for varying majority threshold values. The maximum performance gives the optimal ROI subset and majority threshold value.

5.8.1.1 Combination of boosted SVMs

Algorithm 5.8.1.1 specifies a combination of independently trained classifiers weighted by the individual training errors, where the classifier weights are computed as in the AdaBoostSVM algorithm in [129], assuming an uniform weighting of the data samples. This algorithm is tested in Chapter 6 in the classification based on VBM features, for AD classification, hence the

Algorithm 5.3 Experiment cross-validation procedure

Let be $X = \{x_1, \dots, x_n\}$ input data $x_i \in \mathbb{R}^d$, and $Y = \{y_1, \dots, y_n\}$ the input data class labels $y_i \in \{0, 1\}$.

N is the number of samples.

K is the number of classifiers in the ensemble.

R is the number of ROIs.

1. Perform dataset partition $X = X^m \cup X^v$, $Y = Y^m \cup Y^v$, where (X^m, Y^m) contains the 20% of the dataset.

model selection using (X^m, Y^m)

1. for $i = 1 : R$
 - (a) Select from X^m the i -th ROI data, denoted X_i^m .
 - (b) Compute by LOO, the ROI accuracy, sensitivity and specificity. (a_i, ss_i, sp_i) of the ensemble classifier on X_i^m
2. end for
3. sort the ROIs according to decreasing value of $ss_i + sp_i$
4. for $i = 1 : R$
 - (a) Select from X^m the i -th ROI data, denoted X_i^m , adding it to the incremental model selection feature set $X_i^{inc} \leftarrow X_{i-1}^{inc} \cup X_i^m$, with $X_0^{inc} = \emptyset$.
 - (b) for $\theta = 1 : i$
 - i. Compute by LOO, the ROI accuracy, sensitivity and specificity. $(a_i^\theta, ss_i^\theta, sp_i^\theta)$ of the *meta-ensemble* classifier on X_i^{inc} applying majority threshold θ .
 - (c) endfor
5. Find the set of ROIs (*selectedROIs*) and majority threshold θ^* giving the highest performance $ss_i^\theta + sp_i^\theta$

validation using (X^v, Y^v)

1. for i in *selectedROIs*
 - (a) Select from X^v the i -th ROI data, denoted X_i^v , adding it to the final feature set $X^f \leftarrow \cup X_i^v$ used for testing.
2. endfor
3. Compute final accuracy, sensitivity and specificity by LOO on X^f using θ^* as the majority threshold.

Algorithm 5.4 Combining the independent SVM trained per cluster

1. **Input:** as many sets of training samples with labels as clusters in the statistical parametric map $T_k = \{(x_1, y_1), \dots, (x_N, y_N)\}, k = 1..C$, where N is the number of samples of each cluster.
 2. **Initialize:** the weights of training samples: $w_i^k = 1/N$, for all $i = 1, \dots, N$
 3. **For each k cluster do**
 - (a) Search the best γ for the RBF kernel for the training set T_k , we denote it as γ_k .
 - (b) Train the SVM with T_k and γ_k , we denote the classifier as h_k .
 - (c) Classify the same training T_k set with h_k .
 - (d) Calculate the training error of h_k : $\epsilon_k = \sum_{i=1}^N w_i^k, \quad y_i \neq h_k(x_i)$.
 - (e) Compute the weight of the cluster classifier h_k : $\alpha_k = \frac{1}{2} \ln(\frac{\epsilon_k}{1-\epsilon_k})$.
 4. **Output:** for each test data x its classification is $f(x) = \text{sign}(\sum_{k=1}^C \alpha_k h_k(x))$.
-

algorithm develops an independent classifier for each VBM cluster, estimating its training error and the corresponding weight for the linear combination giving the ensemble response.

Chapter 6

Results

This chapter contains the experimental corollary of the thesis. We have gathered the most salient results already published or that are in the way to publication. We have organized the chapter around the three main diseases described in Chapter 2 that we have tackled in our studies: Alzheimer's Disease, Myotonic Dystrophy, and Schizophrenia. For each of them, we present the results obtained on each database detailed in Appendix A, and the particular conclusions that we draw from them. Some of the experiments have been performed on public databases and others on data provided by medical or psychological collaborators who retains their ownership. We follow somehow the historical order of our studies, so that some of them are less exhaustive in the results reported, specially regarding brain localization of the selected features.

We start the Chapter with a recall of the definition of the performance measures employed to report the results in Section 6.1. Section 6.2 discusses the general experimental design of the computational experiments performed to validate the approaches. Next, Sections 6.3, 6.4, and 6.5 report the results from the studies on Alzheimer's Disease (AD), Myotonic Dystrophy of Type 1 (MD1), and Schizophrenia, respectively.

6.1 Performance measures

In statistics, the terms Type I error (also, α error, or false positive) and type II error (β error, or a false negative) are used to describe possible errors made in a statistical decision process. In 1928, Jerzy Neyman (1894-1981) and Egon Pearson (1895-1980), discussed the problems associated with "*deciding whether or not a particular sample may be judged as likely to have been randomly drawn from a certain population*": and identified "*two sources of error*", namely: (α) the error of rejecting a "correct" null hypothesis, and (β) the error of not rejecting a "false" null hypothesis [151].

In our classification performance evaluation setting, for a set of N binary supervised classification tests we count the number of:

- True positives (TP): positive samples correctly classified,
- False positives (FP): positive samples incorrectly classified,
- True negatives (TN): negative samples correctly classified, and
- False Negatives (FN) negative samples incorrectly classified as positive.

Note that $TP + FP + TN + FN = N$. In our case, positive samples are patients and negative samples are healthy controls. The following classification performance measures¹ have been used throughout this thesis:

Accuracy is the proportion of true results related to to the total number of tests.

$$Accuracy = \frac{TP + TN}{N} \quad (6.1)$$

Sensitivity (aka recall) measures the proportion of actual positives that are correctly classified as such. In the same way, specificity relates to the classifier ability to identify negative samples.

$$Sensitivity \text{ or } recall = \frac{TP}{TP + FN} \quad (6.2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6.3)$$

Precision is the ability of the classifier not to label as positive a sample that is negative.

$$Precision = \frac{TP}{TP + FP} \quad (6.4)$$

F1-score score can be interpreted as a weighted average of the precision and recall, where an F1-score reaches its best value at 1 and worst score at 0.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (6.5)$$

AUC The Area under the Receiver Operating Characteristic curve (AUC) can be interpreted as the probability that the classifier will assign a higher score to a randomly chosen positive example than to a randomly chosen negative example. It can be shown that the AUC is closely related to the Mann–Whitney U test, which tests whether positives are ranked higher than negatives. It is also equivalent to the Wilcoxon test of ranks [95]. The AUC is related to the Gini coefficient (G_q) by

¹In this chapter we sometimes present the performance measures as ratios (between 0 and 1) and sometimes as percentage values (between 0 and 100).

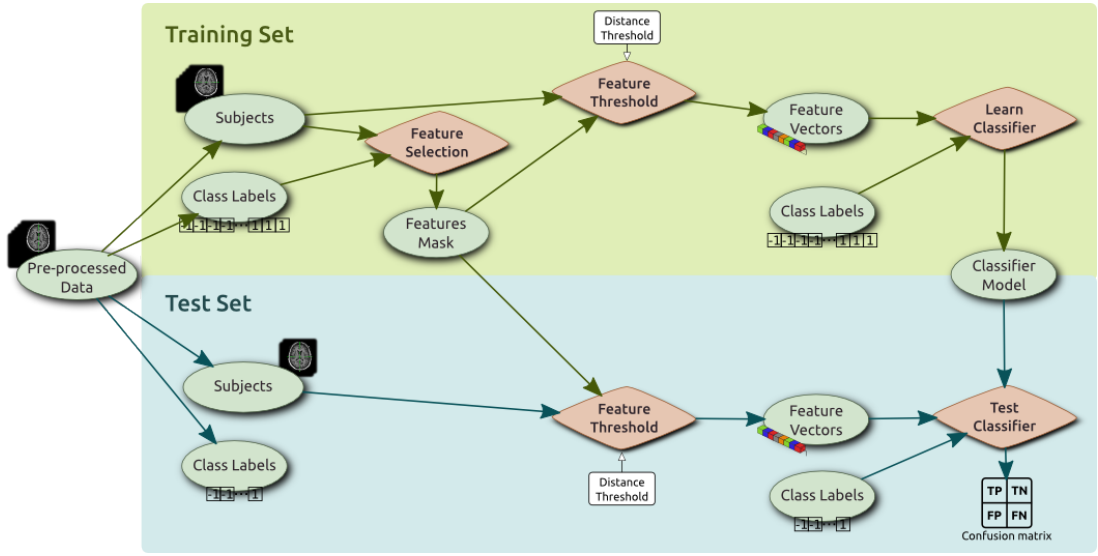


Figure 6.1: Pipeline for circularity free validation of supervised classification with supervised feature selection.

the formula $G_1 = 2 \times AUC - 1$, where

$$G_1 = 1 - \sum_{k=1}^n (X_k - X_{k-1}) (Y_k - Y_{k-1}), \quad (6.6)$$

where X is the increasing false positive rates such that element k is the false positive rate of predictions with score $\geq \text{thresholds}[k]$ and Y the increasing false positive rates such that element k is the true positive rate of predictions with score $\geq \text{thresholds}[k]$. In this case, thresholds are the decreasing classification score values. In this way, it is possible to calculate the AUC by using an average of a number of trapezoidal approximations [143].

6.2 Experimental design

Computational experiments are aimed to provide estimations of the predictive power of the classifiers built on specific features extracted from the data. We are working with supervised machine learning algorithms where validation must be performed on an independent testing dataset, disjoint from the training data in a, e.g. k-fold cross-validation schema. When the feature selection process is supervised as well we could introduce a positive bias in the results if we perform the feature selection process over all available data before carrying out the validation computational experiments. This is known as *double dipping* in the neurosciences [120] or as the introduction of *circular effects* in the analysis. Feature selection must be restricted to the

training data to avoid this bias, and thus, repeated in each fold of the cross-validation schema. We illustrate in Figure 6.1 the cross-validation pipeline applied in most of the studies in this thesis, where we have taken an effort to avoid circularity effects. In the training step, the class label information is used to perform the feature selection and the identified voxel sites are used in the testing step to perform the feature vector extraction. Feature extraction, as said in previous chapters involves the selection of the most salient sites according to set percentile thresholds on the empirical cumulative distribution function (e.c.d.f.) of the voxels saliency measure. The training phase produces a trained classifier that is fully independent from the testing data and the reported results are the average test classification performance values across all folds in the cross-validation process.

6.3 Alzheimer’s Disease

We follow the historical evolution of our experiments regarding machine learning application to AD datasets for the prediction of disease diagnosis based on image biomarkers extracted from MRI data. We have structured the description around the specific databases which are described in appendix A: OASIS98 refers to a subset of 49 matched female subjects (N=98), while OASIS refers to the entire database.

6.3.1 OASIS98

Original studies in this database were performed in collaboration with Maite García-Sebastián while she was doing her PhD. This database has been used as benchmark in several other experiments performed by other members of the group and the database feature vectors are publicly available in the group wiki ². We test classification systems built using the standard SVM, with linear and non-linear (RBF) kernels, and some ANN architectures: Learned Vector Quantization (LVQ), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), Probabilistic Neural Networks (PNN). We have also tested combinations of SVM classifiers trained on independent VBM clusters and an Adaptive Boosting (AdaBoost) strategy tailored to the SVM [129].

6.3.1.1 Experiment 1

Features Feature selection consists in the voxel sites of the VBM clusters found looking for differences between patients and controls. These clusters are used as masks on the MRI voxel intensity data and the GM probability estimation computed during segmentation for the VBM. Two VBM have been performed:

- without any covariate in the GLM (results in Table 6.1), and

²<http://www.ehu.es/ccwintco/index.php/GIC-experimental-databases>

| Feature | size | Accuracy (lk/nlk) | Sensitivity (lk/nlk) | Specificity (lk/nlk) |
|---------------|------|--------------------|----------------------|----------------------|
| GM proportion | 12 | 0.69 / 0.68 | 0.88 / 0.90 | 0.63 / 0.61 |
| MSD | 24 | 0.78 / 0.80 | 0.88 / 0.89 | 0.72 / 0.75 |
| VV | 3611 | 0.73 / 0.76 | 0.75 / 0.76 | 0.72 / 0.77 |

Table 6.1: Classification results with a linear kernel (lk) and a non-linear kernel (nlk). No covariates have been taken into account in the VBM. The values of $\gamma = (2\sigma^2)^{-1}$ for non linear kernel were 0.5, 0.031, 0.0078 for each feature extraction process, respectively.

| Feature | size | Accuracy (lk/nlk) | Sensitivity (lk/nlk) | Specificity (lk/nlk) |
|---------------|------|--------------------|----------------------|----------------------|
| GM proportion | 2 | 0.51 / 0.51 | 1 / 1 | 0.50 / 0.50 |
| MSD | 4 | 0.70 / 0.72 | 0.79 / 0.79 | 0.65 / 0.68 |
| VV | 265 | 0.66 / 0.75 | 0.67 / 0.80 | 0.65 / 0.72 |

Table 6.2: Classification results with a linear kernel (lk) and a non-linear kernel (nlk). The normalized brain volume (nWBV) covariate has been taken into account in the VBM. The values of γ for nlk were 0.5, 2.7, 0.004 for GM proportion, MSD and VV features, respectively.

- taking into account the normalized brain volume (nWBV) as a covariate (results in Table 6.2).

Each VBM process produces different sets of voxel location clusters, and, therefore, different sets of feature vectors. The covariate helps to focus the VBM, giving less and smaller clusters than the VBM without covariates, with smaller dimension feature vectors. Three kind of feature vectors are built:

- GM proportion: is the average GM probability of each cluster
- Mean & StDev (MSD): is the mean and standard deviation of the MRI intensities in each cluster
- Voxel intensities (VV): is the collection of the MRI intensities of the voxels in the clusters.

Baseline Results [83] In the result tables, each entry contains the average cross-validation SVM performance using the linear (lk) kernels and non-linear (nlk) RBF kernels on the corresponding feature vector set. The best accuracy result is reported in Table 6.1: the non-linear SVM reaches 0.80, but this result is not far from the results of the linear SVM (0.78). The classification results of table 6.2, using the covariate nWBV in the design matrix of the GLM applied in the VBM, confirm that the non-linear SVM is more accurate. However, both the size of the feature vectors and performance in table 6.2 are systematically lower than in table 6.1. Due to the low performance, the following experiments do not take into account the VBM features obtained with the nWBV covariate.

Overall the sensitivity results in tables 6.1 and 6.2 are much lower than the specificity. We hypothesize that the source of error is the confusion of mild demented AD patients with control

| Feature | size | γ | Accuracy | Sensitivity | Specificity |
|---------------|------|----------|-------------|-------------|-------------|
| GM proportion | 12 | 0.9 | 0.72 | 0.84 | 0.66 |
| MSD | 24 | 0.6 | 0.87 | 0.89 | 0.86 |
| VV | 3611 | 1.5 | 0.86 | 0.85 | 0.87 |

Table 6.3: Classification results of 40 AD patients vs. 49 control subjects with the SVM and a RBF kernel, 9 possible outliers were taken out from the AD patients subset.

subjects. Mild demented AD patients are subjects with CDR=0.5 (Clinical Dementia Ratio) and a high value for the MMSE (Mini Mental State Examination), i.e. MMSE=30. Therefore we repeat the feature extraction and classification experiment taking out of the population 9 mild demented AD patients. The results for the RBF kernel SVM are given in table 6.3. The classification accuracy of the grows from 80.6% (in the best result of table 6.1) up to 87.5%. Also sensitivity and specificity improve if we compare table 6.1 and table 6.2 against table 6.3.

Results with Neural Networks [174] All classifiers were calculated with a maximum iteration number (epochs) of 100. For the 1-NN classifier, we used the nearest neighbor rule with euclidean distance. In the SVM algorithm, a linear kernel function was used as well as a sequential minimal optimization for the separating hyperplane method. For BPNN, the number of neurons in the hidden layer was 4, the learning rate was set to 0.05, tan-sigmoid transfer function, and training and learning functions were gradient descent with momentum. LVQ2 was trained with 2 hidden neurons, learning rate set to 0.01. The training function used for RBF was according to resilient backpropagation algorithm. In the case of PNN, random order incremental training was used. For the last three algorithms (BPNN, LVQ2 and RBF) zeros were set as initial input and layer delay conditions. These parameters have been selected after a sensitivity analysis.

We tested several cross-validation strategies, because the small database size may have an influence on the results obtained with each of these cross-validation processes. Cross-validation partitions were computed 40 times and we show average accuracy, sensitivity, and specificity for the 10-fold cross-validation procedure.

We tested a variety of Neural Networks based classifiers. We evaluated the performance of the classifiers built with the diverse training and architecture strategies using 10 times the 10-fold cross-validation methodology. The results shown are the mean and standard deviation of the classification performance. We will give results of each different classifiers: Back-propagation (Table 6.3.1.1), RBF (Table 6.3.1.1), PNN (Table 6.3.1.1), LVQ1 (Table 6.3.1.1) and LVQ2 (6.3.1.1).

The best accuracy result (Table 6.3.1.1) is 83% with the LVQ2, but this result is not far from the results of LVQ1 and PNN. Which is a very encouraging result, given that we have not removed critical subjects from the data collection: very mildly demented subjects who could end in a false positive diagnosis. Regarding the usefulness of the features extracted, it is difficult

| Feature | size | #Hidden units | Accuracy | Sensitivity | Specificity |
|---------|------|---------------|-------------|-------------|-------------|
| MSD | 24 | 10 | 0.78 (0.12) | 0.69 (0.14) | 0.88 (0.13) |
| VV | 3611 | 10 | 0.78 (0.11) | 0.72 (0.17) | 0.84 (0.18) |

Table 6.4: Classification results with a BP network with resilient back-propagation. Mean (Standard deviation) of 10 cross-validations.

| Feature | size | Spread | Accuracy | Sensitivity | Specificity |
|---------|------|--------|-------------|-------------|-------------|
| MSD | 24 | 0.02 | 0.66 (0.13) | 0.65 (0.24) | 0.68 (0.14) |
| VV | 3611 | 0.852 | 0.72 (0.10) | 0.65 (0.21) | 0.80 (0.17) |

Table 6.5: Classification results with a RBF network. Mean (Standard deviation) of 10-fold cross-validation.

to make an assessment, because some classifier training algorithms work better with VV than with MSD, and other have the inverse performance. Training and validation on MSD features is obviously more time efficient, and the best result corresponds to this feature extraction process.

Results combining independent SVM per cluster [4] Table 6.9 presents the results of the combination by majority voting of independent SVM classifiers built up over the data extracted from each cluster, searching for the best kernel parameter σ in each classifier independently. The results do not improve over the ones obtained with the whole image feature vector. We note that, contrary to the results in Experiment 1, the results of VV features are a little better than MSD features.

Table 6.10 presents the results of the linear combination of the independent weighted SVM classifier outputs. Each SVM classifier was trained with one VBM cluster feature set and the weights were computed according to its training error. We obtain a significant improvement of the accuracy when considering the voxel intensities as features for the non-linear RBF SVM.

Results with Diverse AdaBoost SVM [4] Table 6.11 shows the results of the Diverse AdaBoost SVM described in Chapter 5. The σ_{min} is set as 0.1, the σ_{ini} is set as 100 and σ_{step} is set as 0.1. The DIV value is set as 0.6.

| Feature | size | Spread | Accuracy | Sensitivity | Specificity |
|---------|------|--------|-------------|-------------|-------------|
| MSD | 24 | 0.02 | 0.78 (0.09) | 0.62 (0.14) | 0.94 (0.1) |
| VV | 3611 | 0.852 | 0.75 (0.14) | 0.68 (0.20) | 0.81 (0.17) |

Table 6.6: Classification results with a PNN network. Mean (Standard deviation) of 10-fold cross-validation.

| Feature | size | #Hidden units | Accuracy | Sensitivity | Specificity |
|---------|------|---------------|-------------|-------------|-------------|
| MSD | 24 | 10 | 0.81 (0.18) | 0.72 (0.27) | 0.90 (0.14) |
| VV | 3611 | 10 | 0.79 (0.13) | 0.76 (0.23) | 0.82 (0.19) |

Table 6.7: Classification results with a LVQ1 network. Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 150 epochs, goal: 0.10 and learning rate: 0.010. Mean (Standard deviation) of 10-fold cross-validation.

| Feature | size | #Hidden units | Accuracy | Sensitivity | Specificity |
|---------|------|---------------|-------------|-------------|-------------|
| MSD | 24 | 10 | 0.83 (0.12) | 0.74 (0.23) | 0.92 (0.1) |
| VV | 3611 | 10 | 0.77 (0.15) | 0.76 (0.23) | 0.78 (0.17) |

Table 6.8: Classification results with a LVQ2 network. Network training parameters: MSD: 200 epochs, goal: 0.01 and learning rate: 0.01; VV: 50 epochs, goal: 0.01 and learning rate: 0.005. Mean (Standard deviation) of 10-fold cross-validation.

| Feature | size | Accuracy (lk/nlk) | Sensitivity (lk/nlk) | Specificity (lk/nlk) |
|---------|------|-------------------|----------------------|----------------------|
| MSD | 24 | 0.74 / 0.75 | 0.51 / 0.56 | 0.97 / 0.95 |
| VV | 3611 | 0.77 / 0.78 | 0.74 / 0.76 | 0.80 / 0.82 |

Table 6.9: Majority voting classification results with linear kernel (lk) and non-linear kernel (nlk) SVM built independently for each VBM cluster.

| | #Features | Accuracy (lk/nlk) | Sensitivity (lk/nlk) | Specificity (lk/nlk) |
|-----|-----------|-------------------|----------------------|----------------------|
| MSD | 24 | 0.71 / 0.79 | 0.54 / 0.78 | 0.88 / 0.80 |
| VV | 3611 | 0.73 / 0.86 | 0.76 / 0.80 | 0.70 / 0.92 |

Table 6.10: Weighted individual SVM per cluster classification results. The value of the RBF kernels for the nonlinear (nlk) classifiers were searched for the best fit to the training set.

| | #Features | Accuracy | Sensitivity | Specificity |
|-----|-----------|-------------|-------------|-------------|
| MSD | 24 | 0.85 | 0.78 | 0.92 |
| VV | 3611 | 0.78 | 0.71 | 0.85 |

Table 6.11: Diverse AdaBoost SVM classification results.

| Size | Accuracy | Sensitivity | Specificity |
|-------|-------------|-------------|-------------|
| 12229 | 0.76 (0.15) | 0.77 (0.28) | 0.75 (0.17) |
| 2000 | 0.79 (0.14) | 0.82 (0.12) | 0.75 (0.20) |
| 1000 | 0.79 (0.10) | 0.87 (0.13) | 0.70 (0.16) |
| 500 | 0.77 (0.13) | 0.87 (0.17) | 0.67 (0.20) |
| 250 | 0.79 (0.10) | 0.90 (0.13) | 0.67 (0.17) |

Table 6.12: Results using linear SVM on DM features obtained from the 0.995 percentile of the Spearman's correlation.

6.3.1.2 Experiment 2

Features The features considered in this experiment are derived from the deformation maps obtained from the non-linear registration of the MRI volumes to the MNI template. A deformation map consists of one displacement vector per voxel. In order to obtain a volume of scalar measures per voxel, we compute the displacement vector magnitude (DM) and Jacobian determinant of the deformation field gradients (JD). The saliency measures for voxel selection are the Pearson's (PC) and the Spearman's (SC) Correlation Coefficient between the voxel scalar deformation measures across subjects and the categorical variable. Voxel sites are selected according to a percentile threshold on the e.c.d.f. of these saliency measures.

Results Linear SVM classifiers performance estimation was measured using a 10-fold cross-validation methodology. The results shown are the mean and standard deviation (stdev) values of the classification results from the cross-validation process. The results obtained on the SC feature vectors are systematically worse than ones obtained from VBM features reported in the Experiment 1[178]. They are shown in tables 6.12, 6.13 and 6.14. Performances from the DM maps with the 0.995 percentile (table 6.12) improve over the 0.999 percentile (table 6.13), reaching values close to the reference values. The application to the JD maps (table 6.14) does not improve the results.

Results of the PC feature vectors, shown in tables 6.15 and 6.16 are better than the results from the SC feature selection. The results on the DM map (table 6.15) are comparable to the reference results in [178]. Note that the use of the larger feature vector gives the same result as the reduced vector of 250 voxel sites, suggesting that a strong feature vector size reduction can be achieved. Results on the JD map (table 6.16) are worse than the DM results. Overall, in terms of accuracy we have only found that the DM features selected with Pearson's correlation coefficient are the only competitive result in this small database.

| Size | Accuracy | Sensitivity | Specificity |
|------|-------------|-------------|-------------|
| 1861 | 0.66 (0.14) | 0.70 (0.20) | 0.62 (0.21) |
| 1000 | 0.66 (0.14) | 0.77 (0.18) | 0.55 (0.20) |
| 500 | 0.71 (0.12) | 0.85 (0.17) | 0.57 (0.17) |
| 250 | 0.72 (0.15) | 0.80 (0.11) | 0.65 (0.29) |

Table 6.13: Results using linear SVM on DM features obtained from the 0.999 percentile of the Spearman's correlation.

| Size | Accuracy | Sensitivity | Specificity |
|-------|-------------|-------------|-------------|
| 17982 | 0.76 (0.14) | 0.77 (0.27) | 0.75 (0.16) |
| 2000 | 0.65 (0.15) | 0.65 (0.21) | 0.65 (0.24) |
| 1000 | 0.60 (0.16) | 0.62 (0.27) | 0.57 (0.31) |
| 500 | 0.58 (0.08) | 0.70 (0.20) | 0.47 (0.18) |
| 250 | 0.61 (0.09) | 0.65 (0.21) | 0.57 (0.20) |

Table 6.14: Results using linear SVM on normalized JD features obtained from the 0.995 percentile of the Spearman's correlation measures.

| Size | Accuracy | Sensitivity | Specificity |
|-------|--------------------|--------------------|--------------------|
| 27474 | 0.84 (0.10) | 0.90 (0.17) | 0.77 (0.14) |
| 2000 | 0.79 (0.12) | 0.85 (0.17) | 0.72 (0.08) |
| 1000 | 0.79 (0.10) | 0.90 (0.13) | 0.67 (0.17) |
| 500 | 0.79 (0.13) | 0.85 (0.21) | 0.72 (0.14) |
| 250 | 0.84 (0.10) | 0.92 (0.12) | 0.75 (0.17) |

Table 6.15: Results using linear SVM on DM features over the 0.995 percentile of the Pearson's correlation saliency measures.

| Size | Accuracy | Sensitivity | Specificity |
|-------|-------------|-------------|-------------|
| 43967 | 0.66 (0.19) | 0.70 (0.20) | 0.62 (0.24) |
| 2000 | 0.75 (0.13) | 0.75 (0.26) | 0.75 (0.17) |
| 1000 | 0.69 (0.15) | 0.77 (0.14) | 0.60 (0.24) |
| 500 | 0.66 (0.16) | 0.72 (0.30) | 0.60 (0.17) |
| 250 | 0.66 (0.17) | 0.70 (0.20) | 0.62 (0.24) |

Table 6.16: Results using linear SVM on JD features over the 0.990 percentile of the Pearson's correlation saliency measures.

6.3.2 OASIS

We have considered the complete OASIS database, described in Appendix A, in order to realize more complete experiments [175].

6.3.2.1 Experiment 1

Features We compute deformation based features as described in Section 4.5 following the hypothesis that the changes in the brain anatomy can serve as image biomarkers for the disease. We perform non-linear registration of the whole OASIS database using FSL FNIRT, computing the following scalar measures of the registration displacement vectors at each voxel, which will be used as features for classification:

- the determinant (denoted *jacs*) and the trace (*trace*) of the Jacobian of the deformation field gradients,
- the modulated grey matter volume (*modgm*),
- the magnitudes of the displacement vectors (*norms*) and
- the geodesic anisotropy of the deformation Green strain tensors (*geodan*).

We use the Pearson's correlation coefficient (PC), the Battacharyya's univariate Gaussian distance measure (BD) and the Welch's t-test (WT) values as significance measures (c.f. Chapter 4) to compute feature selection masks applying percentile thresholds to their e.c.d.f, as discussed in Chapter 4.

Results The results are shown in figures 6.2, 6.3 and 6.4 for PC, BD and WT, respectively. We used a 10-fold stratified cross-validation with a grid-search optimizing the F1-score of a 3x2-fold stratified cross-validation on the training set. In this grid-search the parameters C and σ of a linear and an RBF SVM were adjusted within a given range [175]. Each figure shows the plots of the average accuracy, precision, sensitivity and AUC for varying number of features, depending on the percentile threshold, for each deformation scalar measure. The plots contain also the visualization of the uncertainty as bars giving the span of one standard deviation. The qualitative inspection of these plots shows that in each of them there are no significant differences because all the average values fall inside the rough confidence interval given by the standard deviation bars of the other approaches. Performing a two tail t-test on the data of each plot separately reaches the same conclusion, there are no significant differences between approaches ($p < 0.05$). Overall, the accuracy and AUC results have little variability within each plot, while the precision and sensitivity have more fluctuations. For PC, the best results are obtained with the *modgm* features, while for the other significance measures, best results are given by the *trace* and *geodan* data. It is unexpected that the Jacobian maps produce classification results below other approaches, because they have been used as the reference value for the TBM studies reported in the literature.

Table 6.17 shows the average and standard deviation of cross-validation results obtained for each voxel site significance measure applying the 95% percentile for feature selection. Notice that *modgm* provides the best results for PC, and that results are very close in some instances, in fact there are not statistically significant differences. The results compare well with other approaches to AD prediction from MRI data [173, 52, 161] which report accuracies below those reported here for smaller datasets.

Location of significant voxel sites To have a qualitative validation of the process, we have localized the voxel sites selected by the feature selection process on two probabilistic atlases (MNI structural atlas and the Harvard-Oxford cortical and subcortical atlas). Voxel sites appearing in 90% of the feature selection processes performed independently in each cross-validation fold are considered relevant, thus shown in the figures and localized in the atlas. The regions from the different experiments are all within similar brain areas: frontal and parietal lobes, cerebellum, temporal and occipital lobes, frontal pole, lateral occipital cortex, superior division, precentral gyrus, postcentral gyrus and hippocampus. All these findings agree with the imaging biomarkers reported in the literature [53, 55, 145, 61, 62].

In figures 6.5 and 6.6, we show some slices of the MNI152 template overlaid with the 95% percentile voxel site localizations across the feature selection processes that were performed in the cross-validation folds over the modulated GM data and the Jacobian map data, respectively. Figures show the findings using the three different significance measures.

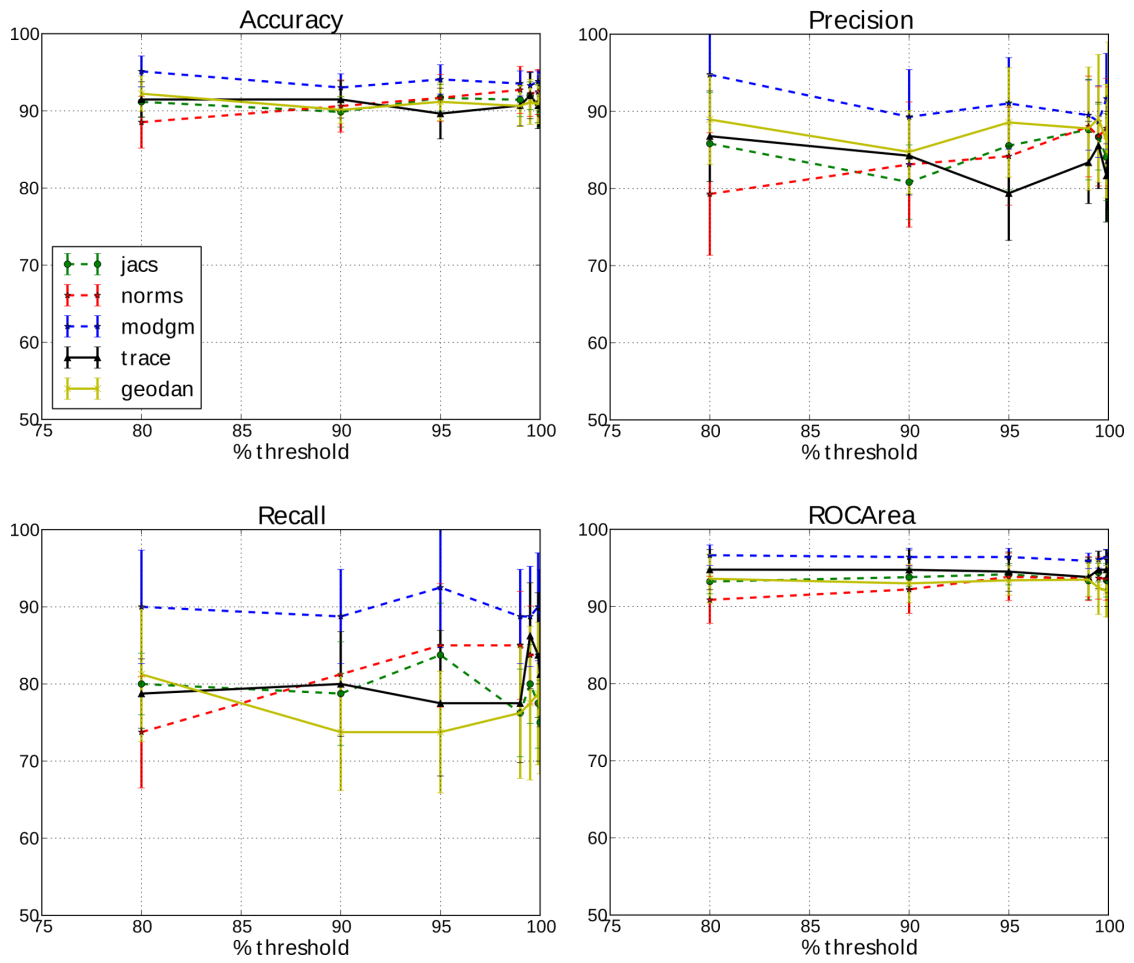


Figure 6.2: Average and standard deviation of Accuracy, Precision, Recall (Sensitivity) and ROC area over the 10-fold cross-tests results of a linear SVM trained to minimize the error rate, features selected using **Pearson's correlation**. Note that the cross-validation population subsets were all the same for all the experiments.

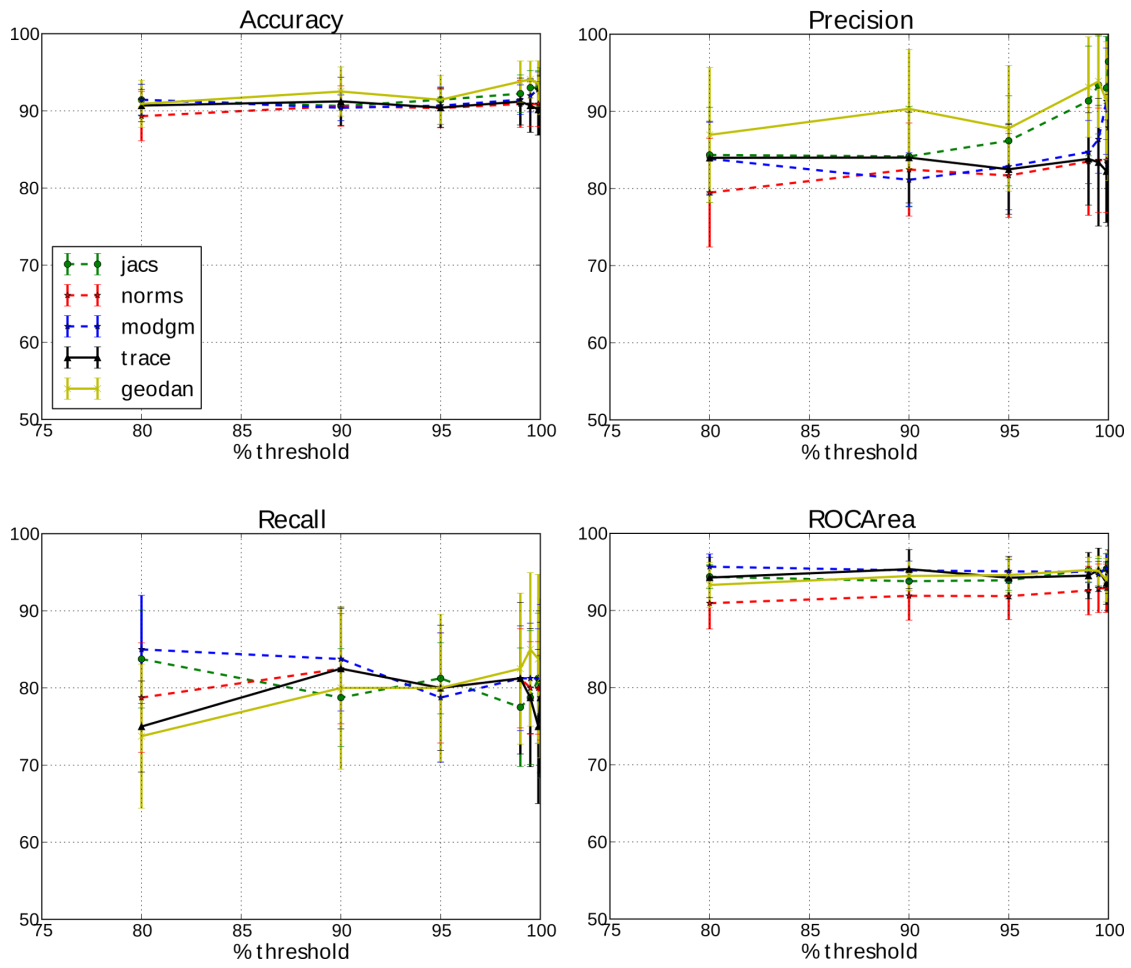


Figure 6.3: Average and standard deviation of Accuracy, Precision, Sensitivity and ROC area over the 10-fold cross-tests results of linear SVM trained to minimize error rate, features selected using **Bhattacharyya's distance**. Note that the cross-validation population subsets were all the same for all the experiments.

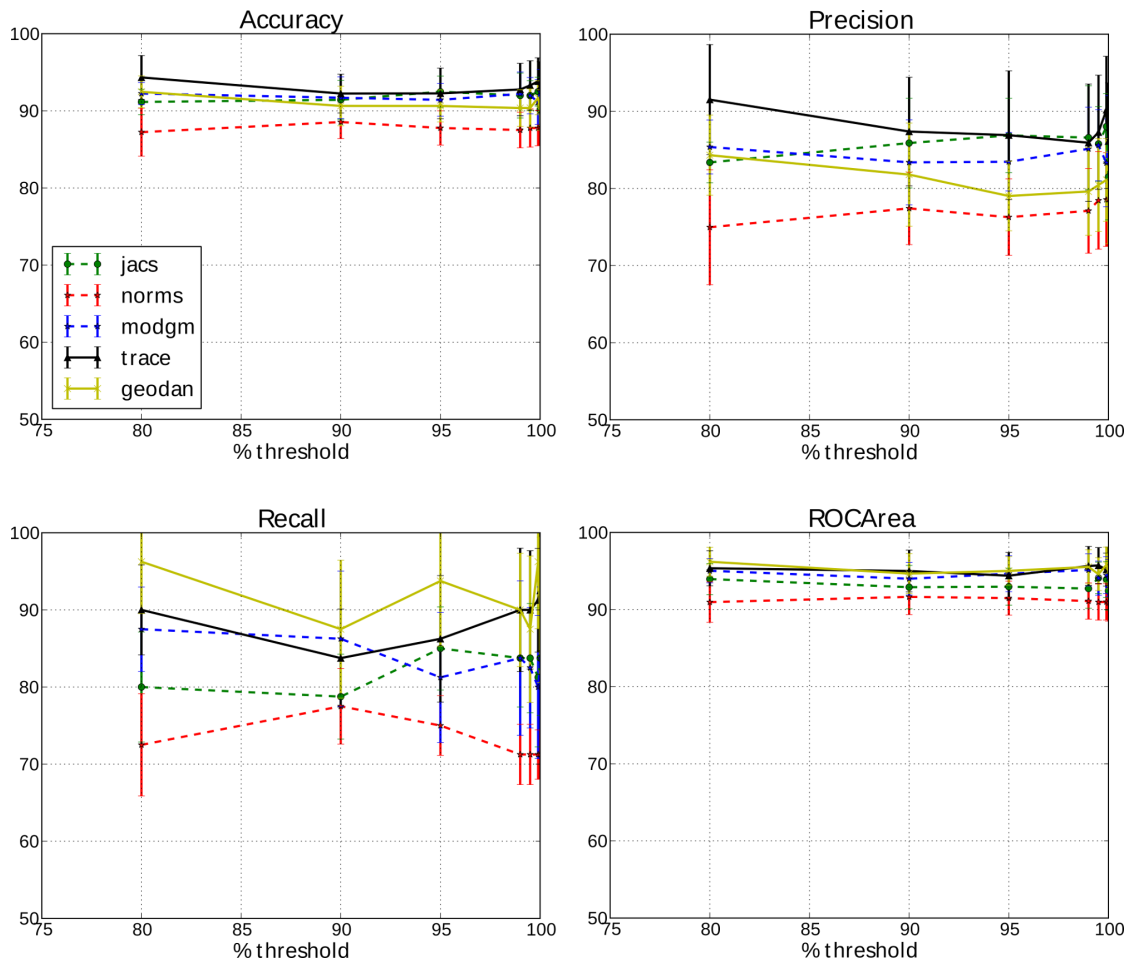


Figure 6.4: Mean and standard deviation of Accuracy, Precision, Recall (Sensitivity) and ROC area over the 10-fold cross-tests results of linear SVM trained minimizing error rate, features selected using **Welch's t-test**. Note that the cross-validation population subsets were all the same for all the experiments.

| | | Accuracy | Precision | Sensitivity | AUC |
|----|--------|---------------------|---------------------|---------------------|---------------------|
| PC | jacs | 88.10 (0.00) | 50.76 (1.70) | 58.33 (3.73) | 92.32 (1.53) |
| | norms | 88.10 (2.75) | 84.14 (6.12) | 58.33 (8.98) | 94.79 (2.53) |
| | modgm | 92.07 (1.12) | 95.83 (5.89) | 86.67 (4.71) | 96.67 (0.44) |
| | trace | 89.43 (2.70) | 79.55 (5.77) | 65.00 (5.00) | 94.27 (0.43) |
| | geodan | 88.15 (2.25) | 87.74 (3.96) | 70.00 (0.00) | 93.49 (1.04) |
| BD | jacs | 91.27 (1.22) | 86.44 (1.89) | 81.67 (7.45) | 95.36 (1.54) |
| | norms | 89.89 (1.77) | 82.62 (6.94) | 79.67 (3.73) | 93.12 (1.48) |
| | modgm | 89.74 (1.77) | 82.94 (4.14) | 78.00 (5.00) | 97.37 (0.29) |
| | trace | 89.52 (1.78) | 82.43 (4.15) | 80.33 (7.45) | 95.67 (1.08) |
| | geodan | 92.09 (2.60) | 88.09 (5.33) | 80.00 (4.00) | 95.37 (0.99) |
| WT | jacs | 92.15 (3.07) | 87.36 (6.67) | 85.373 (6.87) | 93.67 (0.39) |
| | norms | 88.50 (0.89) | 76.40 (3.62) | 75.00 (4.00) | 92.95 (0.87) |
| | modgm | 91.43 (0.89) | 85.98 (1.55) | 83.33 (3.73) | 94.54 (0.50) |
| | trace | 92.83 (0.91) | 85.62 (0.85) | 86.33 (3.73) | 94.32 (0.55) |
| | geodan | 89.92 (1.78) | 79.42 (4.85) | 94.67 (7.45) | 95.00 (0.52) |

Table 6.17: Mean (standard deviation) of Accuracy, Precision, Sensitivity, and ROC area of the full 10-fold cross-validation classification results for features selected with the 95% percentile. Note that the cross-validation population subsets were all the same for all the experiments.

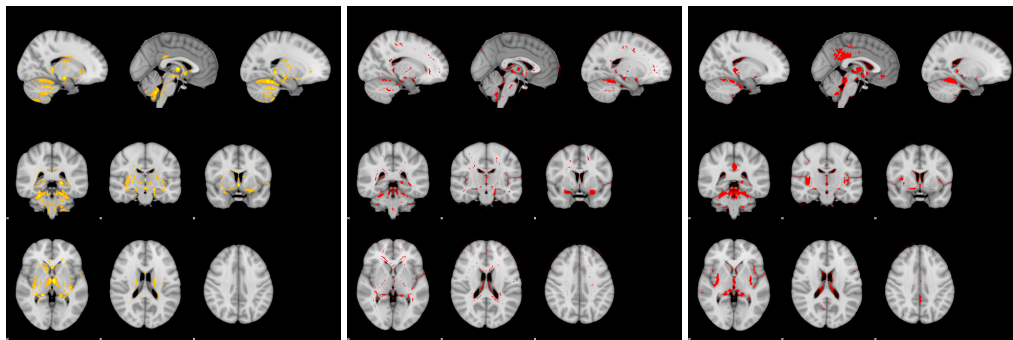


Figure 6.5: Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson's correlation (left), Bhattacharyya's distance (middle) and Welch's t-test (right) computed on the modulated GM (*modgm*).

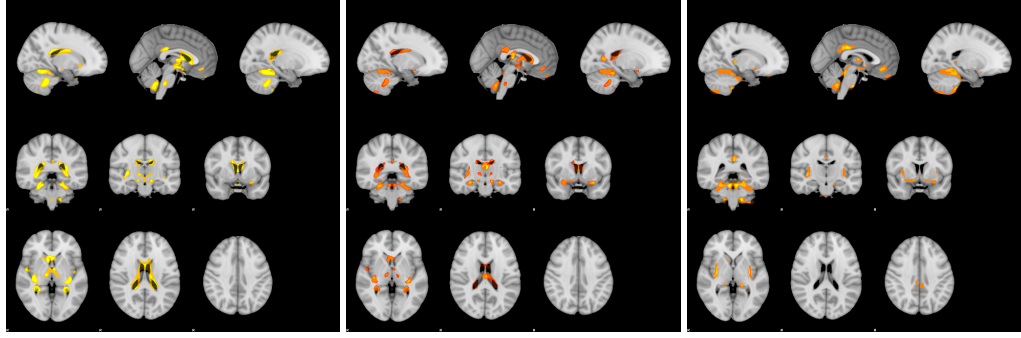


Figure 6.6: Discriminant voxel sites selected applying a 95% percentile on the empirical distribution of the Pearson's correlation (left), Bhattacharyya's distance (middle) and Welch's t-test (right) computed on the Jacobian maps (*jacs*).

| | Accuracy | Sensitivity | Specificity | AUC |
|---------------------------|-------------|-------------|-------------|-------------|
| Majority Voting | 83.6 | 75.0 | 86.4 | 80.7 |
| Best 20 Training AUC | 83.6 | 91.0 | 81.3 | 86.2 |
| Best 20 Training F1-score | 83.2 | 90.0 | 81.0 | 85.5 |
| Left Hem. AD ROIs | 82.4 | 81.0 | 82.9 | 81.9 |
| Mild AD ROIs | 82.0 | 74.0 | 84.5 | 79.9 |
| Normal AD ROIs | 83.2 | 79.0 | 84.5 | 81.7 |

Table 6.18: Accuracy, sensitivity, specificity and ROC area of the leave-one-out cross-validation of SVM classification results.

6.3.2.2 Experiment 2

Features After registration of the anatomical MRI data to the MNI template, the AAL atlas is used to define the ROIs in the brain. The ROI segmentation was used to compute several regional features on the GM segmentation probability map. Seven statistical values computed from each of the AAL ROIs compose the features set. The procedure of feature extraction is better described in Chapter 4.

Results of independent SVM [176] We test ensembles of SVM as performed in Experiment 1 over the OASIS98 data. The independent SVMs are trained on features extracted from disjoint regions of the brain, testing different ensemble decision methods: majority voting between all ROIs, the 20 ROIs with best AUC or F1-score during training, and 3 different sets of ROIs selected on the basis of *a priori* knowledge from AD literature. We report the performance of linear kernel SVMs using leave-one-out cross-validation, instead of the general 10-fold cross-validation applied elsewhere in this thesis. We report accuracy, sensitivity, specificity and area under the ROC curve (AUC) [70] for each ensemble decision. In table 6.18 we show the classification performance of the ensemble with linear SVMs.

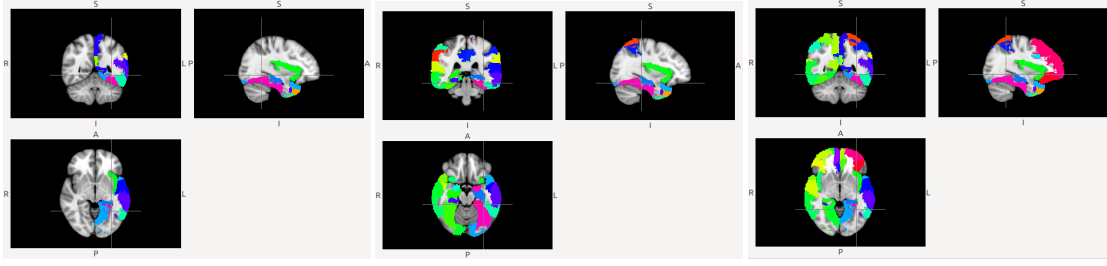


Figure 6.7: Slices of the MNI standard template showing the *a priori* maps of affected regions in the left temporal area (left), corresponding to mild AD (middle) and to normal AD (right) used for ensemble decision.



Figure 6.8: Slices of the MNI standard template where the 20 best training AUC (left) and F1-score (right) ROIs are coloured.

After obtaining the result from all the classifiers in the ensemble we have to make an aggregate decision to have only one class decision. Here we test 6 ensemble majority voting decision criteria using (1) all of the ROI classifiers, the 20 with best average (2) training AUC and (3) training F1-score, (4) an *a priori* set of temporal brain regions and hippocampus only in the left hemisphere known to be affected by AD, (5) another set of brain regions which include temporal and parietal areas of the brain, cingulum, insula and hippocampus from both hemispheres and (6) the same previous set of regions plus frontal lobe regions. In figure 6.7 we show the regions included in each of these *a priori* ROI maps.

Discriminant ROIs In figure 6.8 we show the 20 regions that have obtained best training AUC and F1-score during cross-validation. Most of the ROIs coincide with the manually selected ROIs, mainly with the left hemisphere ROIs and frontal lobes.

Results with ensembles of classifiers [18] This work has been carried out with the collaboration of Borja Ayerdi, in fact proposer of the HERF, and Darya Chyzhyk, in fact inventor of the BDC. We test other ensemble classifier algorithms: Random Forest (RF), Extreme Learning

| | Accuracy | Sensitivity | Specificity |
|-------------|-------------|-------------|-------------|
| RF | 79.0 | 97.5 | 73.1 |
| HERF | 79.3 | 96.3 | 73.9 |
| ELM | 70.8 | 92.5 | 43.7 |
| BDC | 80.8 | 92.5 | 77.1 |

Table 6.19: Ensembles classification performance.

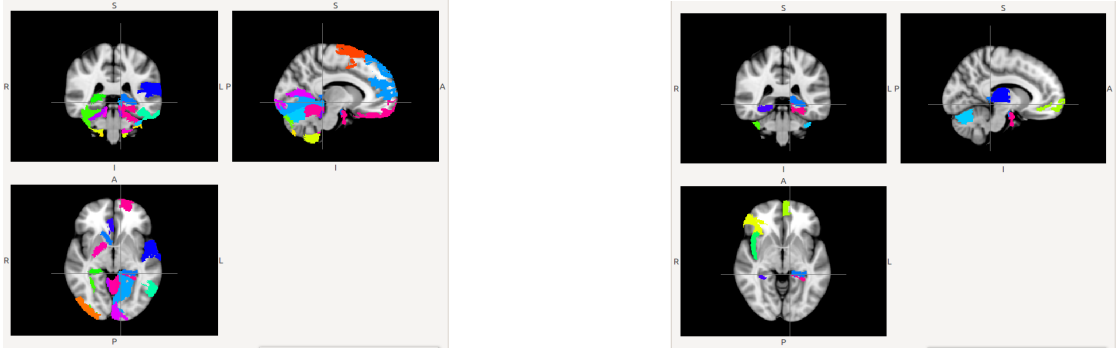


Figure 6.9: Slices of the MNI standard template where the ROIs selected by ELM (left) and BDC (right) are coloured.

Machines (ELM), Hybrid ELM-RF (HERF) and Bootstrapped Dendritic Computing (BDC). In table 6.19 we show the classification performance of the ensembles. Best results correspond to the BDC, but in general results are below other conventional approaches in the literature. These experiments use very few ROIs for the classification so the number of features is very small compared with other methods in the literature. Figures 6.9 and 6.10 show the ROIs selected from the model selection with each of the classifiers. The RF has selected the largest number of ROI (49), while the BDC selected the smallest number (10), meaning that BDC achieves its results with only 70 features. The only ROI which has been selected for all the classifiers is the left parahippocampal gyrus, which is known to show atrophies together with the entorhinal cortex in an early stage of visible anatomical degradation of the brain with Alzheimer's disease.

6.3.3 Data from Hospital de Santiago

This work has been done in collaboration with Maite Termenón, while doing her PhD in the group, and Ariadna Besga from Hospital de Santiago [90]. It is radically different from the previous studies in the sense that deals with diffusion data and the work involved direct communication with the medical doctors.

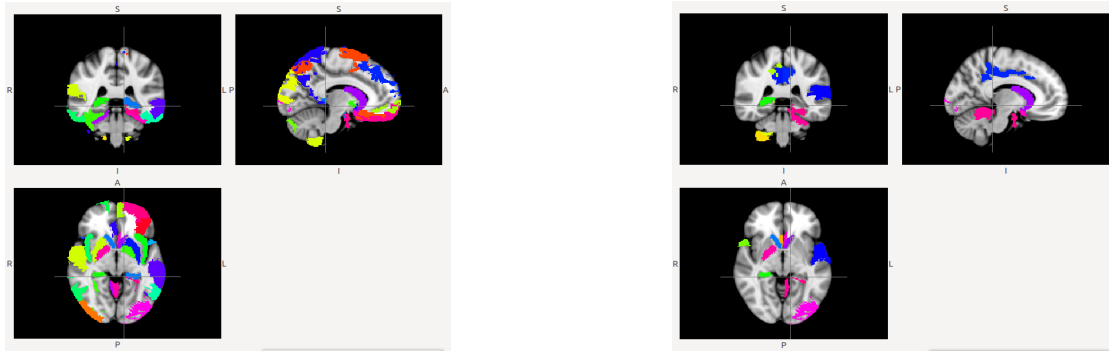


Figure 6.10: Slices of the MNI standard template where the ROIs selected by RF (left) and HERF (right) are coloured.

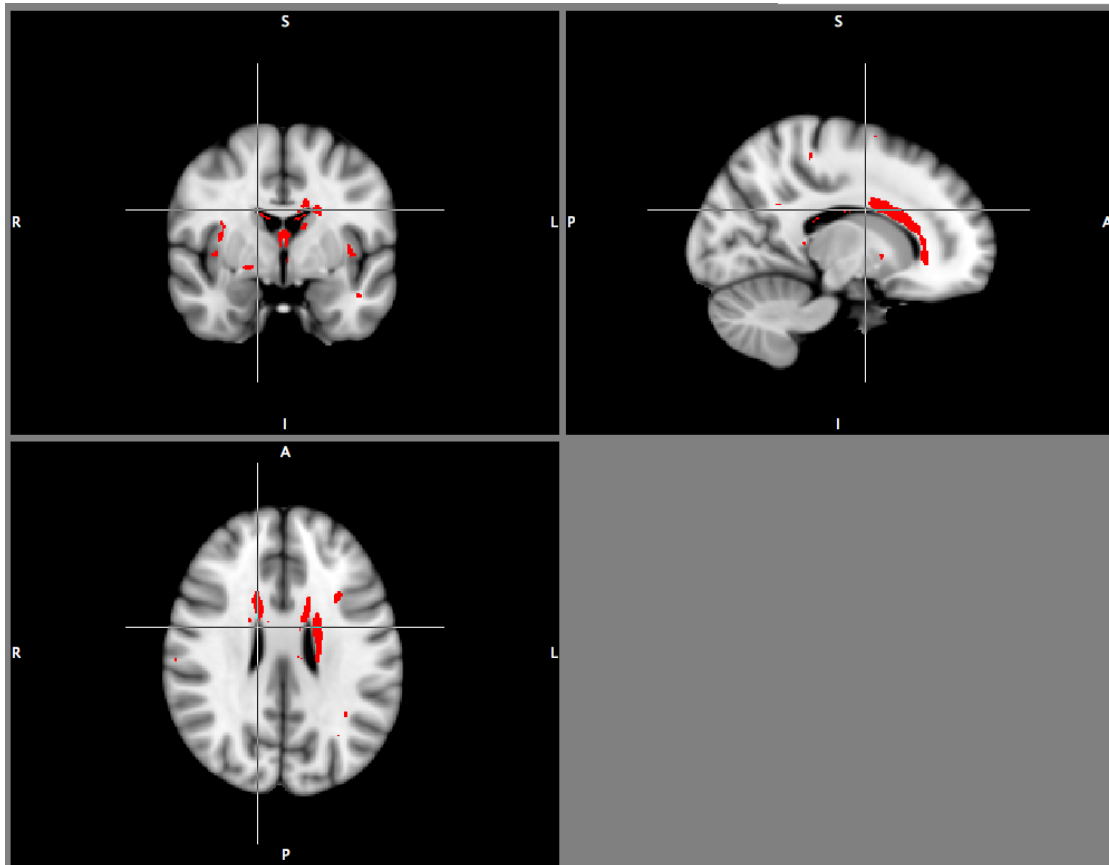


Figure 6.11: Voxel sites (red dots) corresponding to FA features selected setting a 99.5% percentile of the e.c.d.f. of the absolute value of the Pearson's correlation of FA data with the class variable.

Features The data consists in anatomical MRI (T1-weighted) and diffusion (DWI) data. After spatial normalization, the DWI is converted into Diffusion Tensor data (DTI), which are converted into scalar volumetric maps computing the Fractional Anisotropy (FA) and Mean Diffusivity (MD) measures. The Pearson's correlation coefficient is the voxel significance measure for feature selection. The localization of the features selected on the FA map with a high percentile of its e.c.d.f. is shown in Figure 6.11.

The localization of the features selected on the FA map with a 99.5% percentile of its e.c.d.f. overlaid on the MNI template is shown in Figure 6.11. Voxel sites selected to build the feature vectors were localized in the following regions of the brain ³: the caudate [135] and insula (in agreement with [91]) but also in the temporo-parietal lobe (in agreement with [171]), frontal lobe and thalamus for the FA data. In the MD data the most discriminant voxel sites were located in the temporal lobe (in agreement with [41]), insula, temporo-parietal lobe, and frontal lobe. Therefore, voxel site selection for classification is in agreement with previous studies despite no *a priori* information has been used in the process.

Results The average values of the performance measures obtained from the 10-fold cross-validation steps for each percentile (within the given range) of the correlation measures, subsequent combinations of classifiers and feature vectors are presented in figures 6.12, 6.13 and 6.14. Tested ML approaches are the SVM with linear kernels and the nearest neighbour (1-NN). The best results are achieved with linear SVM classifier on the 99.9% correlation percentile of the FA data, resulting in almost perfect accuracy, sensitivity and specificity. The linear SVM reported very high values for all the remaining instances of feature vectors extracted from FA and MD data. The results of the 1-NN are worse in general. Comparing the discriminative power of the FA and MD features, we find that FA features provide significantly better discrimination than MD features: accuracy, sensitivity and specificity are systematically greater for FA than for MD features.

Figure 6.12 shows results of 10-fold cross-validation average accuracy as a function of the number of FA and MD features. The FA features achieve perfect accuracy with the linear SVM, while the MD features achieve much lower accuracy. The classification based on the linear SVM does not suffer from the curse of dimensionality (or Hughes effect) as much as the 1-NN. The 1-NN increases its performance as the number of features increase, until the maximum performance is achieved, after that, increasing the number of features yields worse results. Figures 6.13 and 6.14 plot the sensitivity and specificity, respectively, of the classifiers as a function of the number of features. Results are consistent with the accuracy results. It is remarkable that in the figure 6.14 the decrease in Specificity of the MD features from SVM to 1-NN classifiers is much less than the corresponding decrease observed on the FA features.

³Using "atlasquery" tool from FSL (<http://www.fmrib.ox.ac.uk/fsl/>)

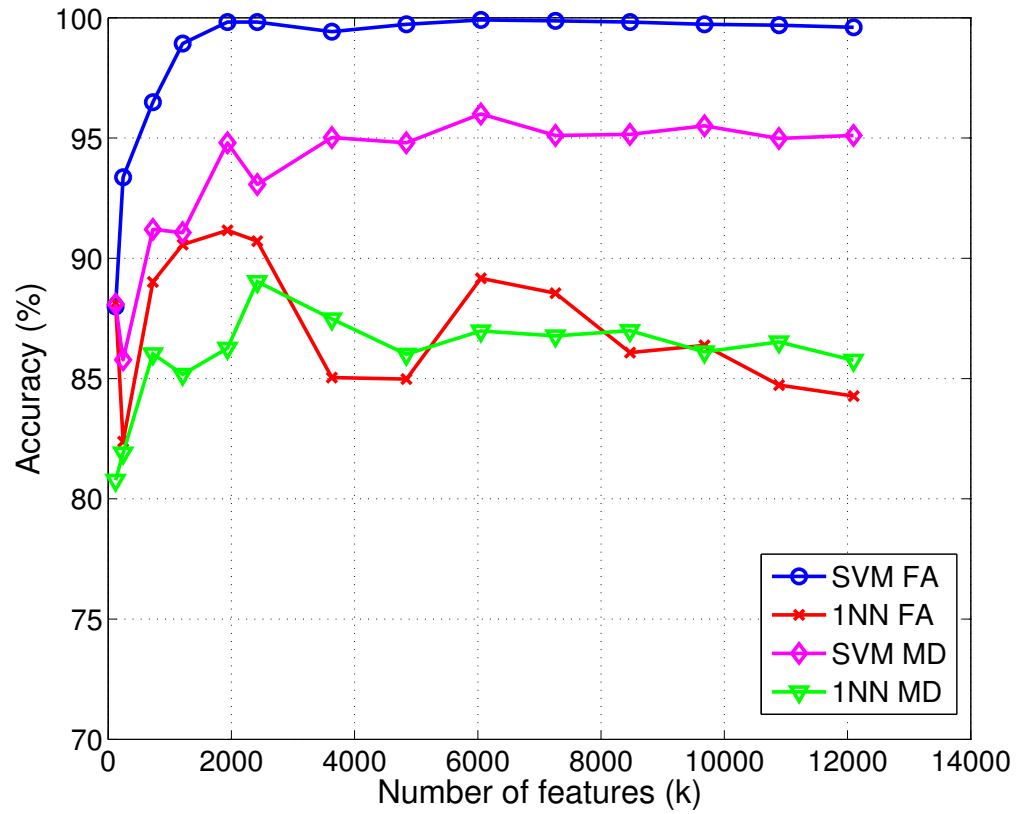


Figure 6.12: Average accuracy of AD patient classification obtained with SVM and 1-NN classifiers varying number of features for FA and MD maps computed from DWI data.

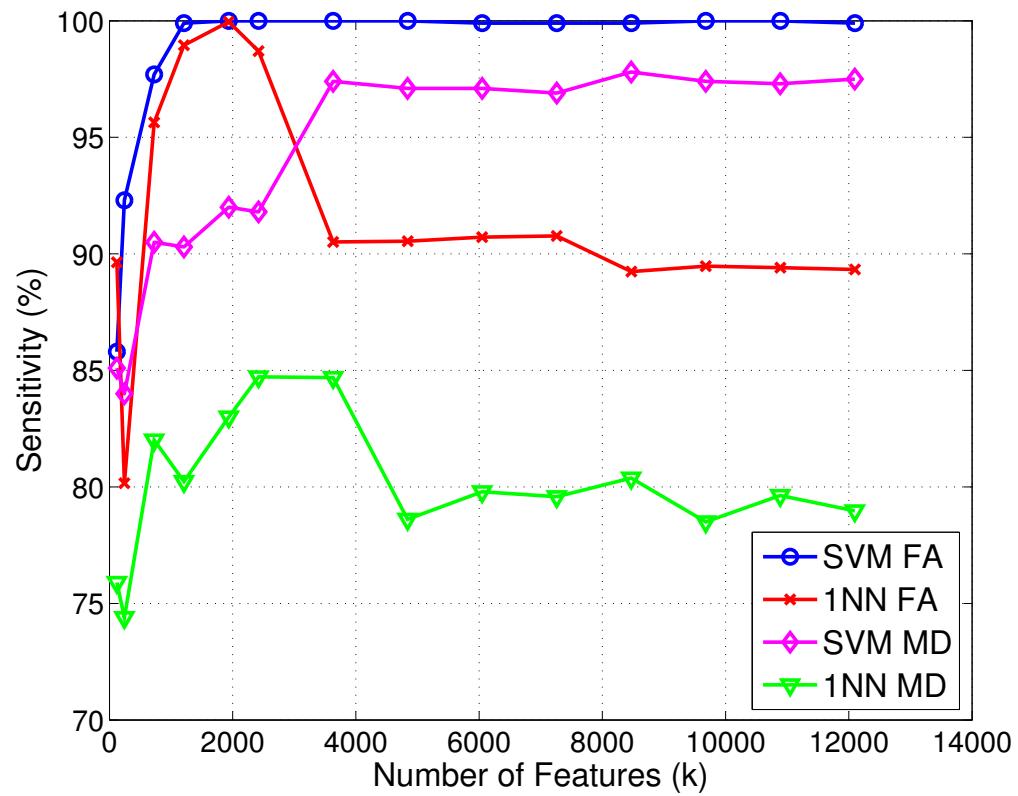


Figure 6.13: Average sensitivity of AD patient classification obtained with SVM and 1-NN classifiers varying number of features for FA and MD maps computed from DWI data.

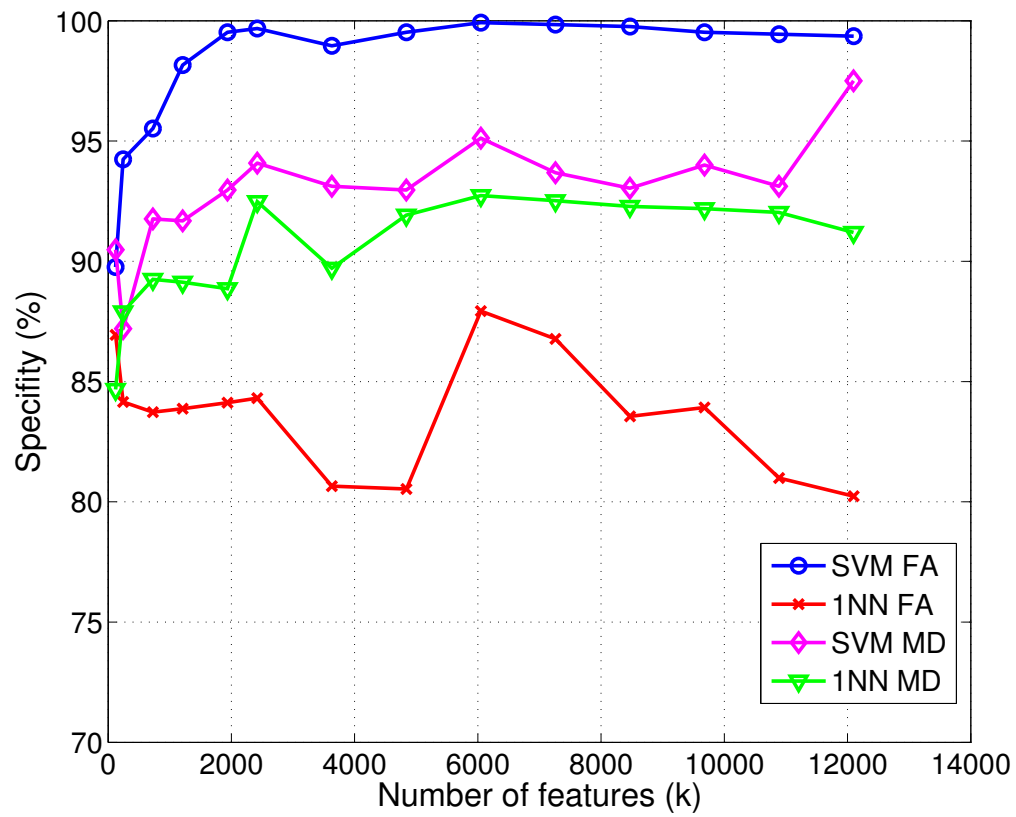


Figure 6.14: Average specificity obtained with SVM and 1-NN classifiers varying number of features for FA and MD data.

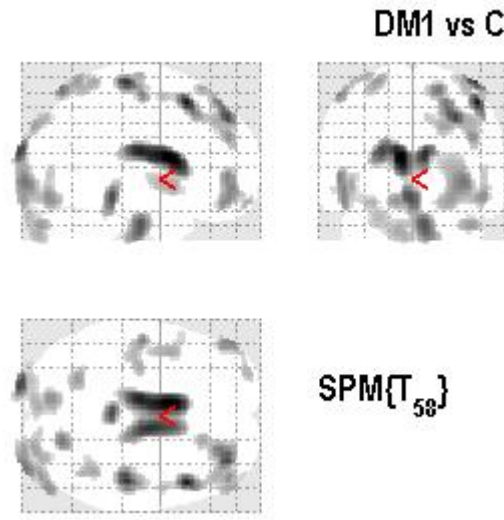


Figure 6.15: SPM results showing GM differences between patients and controls in the VBM of the Myotonic Dystrophy database.

6.4 Myotonic Dystrophy Type 1

The Myotonic Dystrophy data was provided by the Departamento de Neurología of the Hospital Donostia, specifically by Dr. Andone Sistiaga. Work on this data was carried out in collaboration with her and Maite García-Sebastián and reported in [8].

Features As in Experiment 1 on the AD data, we perform VBM on the data using SPM5, selecting features as the voxel sites in the VBM detected clusters. We designed the statistical analysis as a two-sample t-test and as two-sided F-test, in which the first group corresponds to DM subjects. In SPM software terms: the contrast has been set to $[-1 \ 1]$ right-tailed (Controls > AD patients) with $p\text{-value} < 0.05$ Familywise Error rate (FWE) correction.

Feature vectors are extracted from the MRI data at the selected voxel sites. The experiment includes the effect of different Full-Width Half-Maximum (FWHM) sizes and spatial cluster extent threshold (Size-Thr) value settings [8]. In general, VBM identified differences in GM volume both in the cortex and in subcortical areas. For instance, figure 6.15 shows the results of one instance of t-test VBM showing the clusters detected. Figure 6.16 shows the most salient clusters localized over the GM template obtained from the collection of images in the database. VBM finds differences between patients and controls throughout practically the entire cortex, particularly in the frontal lobe and bilateral parietal. A smaller volume of the occipital lobe and bilateral cerebellum in the patients was also seen. The subcortical basal ganglia, specifically the left caudate and bilateral thalamus, had a smaller volume in DM1 patients than in controls.

| FWHM (mm) | thresh | size | accuracy | sensitivity | specificity |
|-----------|--------|------|-------------|-------------|-------------|
| 8 | 0 | 76 | 78.3 | 73.3 | 83.3 |
| | 100 | 8 | 76.7 | 66.7 | 86.7 |
| | 200 | 4 | 76.7 | 66.7 | 86.7 |
| 9 | 0 | 76 | 80.0 | 70.0 | 90.0 |
| | 100 | 16 | 75.0 | 66.7 | 83.3 |
| | 200 | 4 | 76.7 | 66.7 | 86.7 |
| 10 | 0 | 70 | 78.3 | 63.3 | 93.3 |
| | 100 | 22 | 76.7 | 73.3 | 80.0 |
| | 200 | 8 | 78.3 | 70.0 | 86.7 |
| 11 | 0 | 64 | 71.7 | 63.3 | 80.0 |
| | 100 | 24 | 75.0 | 63.3 | 86.7 |
| | 200 | 12 | 75.0 | 63.3 | 86.7 |
| 12 | 0 | 68 | 71.6 | 63.3 | 80.0 |
| | 100 | 36 | 73.3 | 63.3 | 83.3 |
| | 200 | 18 | 75.0 | 70.0 | 80.0 |

Table 6.20: SVM classification results of Myotonic Dystrophy for MSD features extracted from clusters found by t-test VBM, FWE=0.05.

| FWHM (mm) | thresh | size | accuracy | sensitivity | specificity |
|-----------|--------|------|-------------|-------------|-------------|
| 8 | 0 | 2059 | 81.7 | 83.3 | 80.0 |
| | 100 | 1226 | 78.3 | 70.0 | 86.7 |
| | 200 | 958 | 80.0 | 80.0 | 76.7 |
| 9 | 0 | 2826 | 78.3 | 73.3 | 83.0 |
| | 100 | 2044 | 76.7 | 73.3 | 80.0 |
| | 200 | 1182 | 75.0 | 66.7 | 83.3 |
| 10 | 0 | 3710 | 76.7 | 73.3 | 80.0 |
| | 100 | 3103 | 80.0 | 76.7 | 83.3 |
| | 200 | 2131 | 73.3 | 70.0 | 76.7 |
| 11 | 0 | 5022 | 73.3 | 73.3 | 73.3 |
| | 100 | 4278 | 78.3 | 73.3 | 83.3 |
| | 200 | 3434 | 75.0 | 70.0 | 80.0 |
| 12 | 0 | 6542 | 76.6 | 73.3 | 80 |
| | 100 | 6391 | 75.0 | 70.0 | 80 |
| | 200 | 5148 | 73.3 | 70.0 | 76.6 |

Table 6.21: SVM classification results of Myotonic Dystrophy of type I for VV features extracted from clusters found by t-test VBM, FWE=0.05.

| FWHM (mm) | thresh | size | accuracy | sensitivity | specificity |
|-----------|--------|----------|-------------|-------------|-------------|
| 8 | 0 | 52 | 76.7 | 70.0 | 83.3 |
| | 100 | 4 | 78.3 | 70.0 | 86.7 |
| | 200 | 4 | 80.0 | 73.3 | 86.7 |
| 9 | 0 | 48 | 78.3 | 70.0 | 86.7 |
| | 100 | 8 | 75.0 | 63.3 | 86.7 |
| | 200 | 4 | 73.3 | 60.0 | 86.7 |
| 10 | 0 | 52 | 75.0 | 63.3 | 86.7 |
| | 100 | 10 | 73.3 | 63.3 | 83.3 |
| | 200 | 4 | 75.0 | 63.3 | 86.7 |
| 11 | 0 | 58 | 76.7 | 70.0 | 83.3 |
| | 100 | 12 | 80.0 | 66.7 | 93.3 |
| | 200 | 2 | 78.3 | 70.0 | 86.7 |
| 12 | 0 | 56 | 73.3 | 63.3 | 83.3 |
| | 100 | 18 | 76.7 | 66.7 | 86.7 |
| | 200 | 6 | 76.7 | 63.3 | 90.0 |

Table 6.22: SVM classification results of Myotonic Dystrophy of type I for MSD features extracted from clusters found by F-test VBM, FWE=0.05.

| FWHM (mm) | thresh | size | accuracy | sensitivity | specificity |
|-----------|--------|------------|-------------|-------------|--------------|
| 8 | 0 | 1293 | 78.3 | 76.7 | 80.0 |
| | 100 | 769 | 76.7 | 73.3 | 80.0 |
| | 200 | 769 | 76.7 | 70.0 | 83.3 |
| 9 | 0 | 1665 | 80.0 | 80.0 | 80.0 |
| | 100 | 1155 | 78.3 | 70.0 | 86.7 |
| | 200 | 930 | 81.7 | 80.0 | 83.3% |
| 10 | 0 | 2128 | 76.7 | 73.3 | 80.0 |
| | 100 | 1481 | 75.0 | 70.0 | 80.0 |
| | 200 | 1090 | 76.7 | 70.0 | 83.3 |
| 11 | 0 | 2733 | 71.7 | 66.7 | 76.7 |
| | 100 | 2050 | 75.0 | 70.0 | 80.0 |
| | 200 | 1269 | 75.0 | 70.0 | 80.0 |
| 12 | 0 | 3465 | 75.0 | 70.0 | 80.0 |
| | 100 | 2887 | 75.0 | 70.0 | 80.0 |
| | 200 | 2082 | 73.3 | 66.7 | 80.0 |

Table 6.23: SVM classification results of Myotonic Dystrophy of type I for VV features extracted from clusters found by F-test VBM, FWE=0.05.

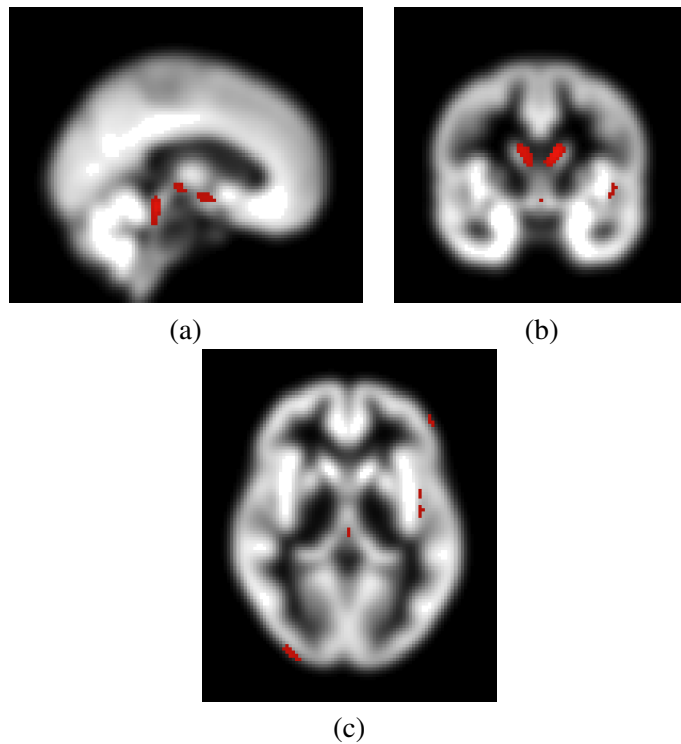


Figure 6.16: SPM results shown as the most significant VBM clusters found in the Myotonic Dystrophy database superimposed on the GM averaged image (obtained from subjects under study, 8 mm FWHM isotropic Gaussian kernel).

Results The classifiers are built using non-linear SVM with RBF kernel. Performance is estimated with a 10-fold cross-validation as explained in Section 6.2, with a nested cross-validation searching exhaustively for the optimal value of the RBF kernel parameter γ . The results of these experiments are displayed in tables 6.20, 6.21, 6.22 and 6.23. The results using the MSD features are given in tables 6.20 and 6.22, while the results related to VV features are given in tables 6.21 and 6.23.

Few instances reach the 80% accuracy. Notice also in both tables that the specificity is systematically greater than the sensitivity. As was said in the AD study above, this trend in the results forces to attribute the lack of sensitivity to the feature extraction process. In order to deal with the milder dementia cases, the process must be improved. It must be also taken into account that until now there have not been findings in MRI associated with the MD1, meaning that the results reported here are the among the first published results in this sense. Another interesting effect that can be appreciated in the tables 6.20 and 6.20 is that growing number of features, obtained with stronger smoothing does not imply improved classification accuracy. This goes contrary to the assessments by human experts, which favor stronger smoothing in order to obtain better and bigger detections of the effect for visualization purposes.

Discussion of feature selection medical relevance The classifiers applied to the data features make decisions based on the whole of the feature vector. However, the locations of the feature sources, the voxel cluster found by VBM, has some clinical interpretations. For MD1 the most discriminant voxels were in the caudate nucleus, fronto-parietal lobe and thalamus. These results involving the fronto-parietal areas, agree with previous studies [11] and with the profile suggested by previous neuropsychological results in a large DM1 sample [187]. In fact, we have found the volume corresponding to the caudate nuclei and the thalamus to be smaller in DM1 patients than in the control subjects. Basal ganglia have traditionally been associated with motor processes, although there is increasing evidence to support their role in cognitive functions [147]. Basal ganglia nuclei relate to one another and to the cortex through different cortico-striatal loops, which emphasizes the functional interrelationship between the neocortex, the striatum and the thalamus. Three of the five loops in which the striatum is involved pass through the caudate and the thalamus on the way to the cortex: the oculomotor, dorsolateral prefrontal and lateral orbitalfrontal loops [89]. The last two target two prefrontal cortical areas that are thought to be involved in various aspects of cognitive behavior. In fact, the dorsolateral cortex is the structure mostly closely associated with executive functions and allows the organization of information to facilitate a response. The orbitofrontal circuit allows the integration of limbic and emotional information with behavioral responses [31].

| Database | Percentile | DT Measure | Number of voxels |
|----------|------------|------------|------------------|
| A | 99.990% | FA | 241 |
| | | MD | 241 |
| B | 99.992% | FA | 193 |
| | | MD | 193 |
| C | 99.995% | FA | 121 |
| | | MD | 121 |
| D | 99.997% | FA | 72 |
| | | MD | 72 |
| E | 99.999% | FA | 24 |
| | | MD | 24 |

Table 6.24: Databases considered, percentile on the correlation distribution and size of the feature vectors.

6.5 Schizophrenia

6.5.1 NAMIC: Brain Multimodality

Our first approach to Schizophrenia diagnostic support systems was using DTI data from the public available NAMIC: Brain Multimodality database which includes anatomical T1-weighted MRI, rs-fMRI and DTI data.

Features We have used the registration information of anatomical data for co-registration of the DTI, in order to obtain spatially normalized DTI data. We compute the scalar FA and MD values [177] and the Pearson’s correlation as the voxel significance measure. In table 6.24 we show the percentiles and the number of voxels selected for each feature vector.

Feature localization in the brain Although the voxel sites selected to build the feature vectors (the feature mask) were localized in many different regions of the subject brains, we found that most were concentrated in regions of characteristic abnormalities found for schizophrenia shown in the literature (see [121] for references). The features voxel locations⁴ were different for FA and MD maps. In the case of FA, the selected voxels were localized mainly in parietal and temporal lobes, but also in the cerebellum and occipital lobe. More specifically, in WM we found discriminant voxel values in the cingulum bundle, superior and inferior longitudinal fasciculus and in the inferior fronto-occipital fasciculus. On the other hand, in the MD maps, the most discriminant voxel values were the ones localized in frontal and parietal lobes, more specifically the cingulum bundle, inferior fronto-occipital and longitudinal fasciculus, and superior longitudinal fasciculus.

⁴This specification of the voxel locations were obtained with the “atlasquery” tool from FMRI’s FSL (<http://www.fmrib.ox.ac.uk/fsl/>) using the “MNI Structural Atlas” and the “JHU White-Matter Tractography Atlas”.

| Database | Classifier | Accuracy | Sensitivity | Specificity |
|----------|-------------|----------|-------------|-------------|
| A | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.75 | 0.67 | 1.00 |
| | RBF | 0.98 | 0.97 | 1.00 |
| | PNN | 1.00 | 1.00 | 1.00 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| B | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.75 | 0.66 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 1.00 | 1.00 | 1.00 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| C | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.77 | 0.68 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 1.00 | 1.00 | 1.00 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| D | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.77 | 0.68 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 0.99 | 0.99 | 1.00 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| E | 1-NN | 0.94 | 0.90 | 0.99 |
| | SVM | 0.95 | 0.90 | 1.00 |
| | BP | 0.76 | 0.67 | 1.00 |
| | RBF | 0.92 | 0.90 | 0.94 |
| | PNN | 0.94 | 0.90 | 0.99 |
| | LVQ2 | 0.97 | 0.94 | 1.00 |

Table 6.25: Average 10-fold cross-validation classification performance results for diverse databases and classifiers on the FA features of the NAMIC database.

Results We tested six classification algorithms with conventional 10-fold cross-validation. The average performance results are presented in table 6.25 and 6.26. The most striking result is that we found optimal performance of almost all classifiers built from the provided feature vectors. The only exceptions were the results of PNN on MD data, which can be explained by the inaccurate selection of the Gaussian kernel variance in the sensitivity analysis. Also BP shows lower performance than the others. The second general result is that MD features seem to perform slightly better than FA features, disregarding the anomaly of PNN classifiers. In the

| Database | Classifier | Accuracy | Sensitivity | Specificity |
|----------|-------------|----------|-------------|-------------|
| A | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.78 | 0.69 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 0.54 | 0.54 | 0.54 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| B | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.78 | 0.70 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 0.52 | 0.52 | 0.52 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| C | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.77 | 0.68 | 1.00 |
| | RBF | 1.00 | 1.00 | 1.00 |
| | PNN | 0.52 | 0.52 | 0.52 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| D | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.77 | 0.68 | 1.00 |
| | RBF | 0.84 | 0.79 | 0.90 |
| | PNN | 0.55 | 0.55 | 0.55 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |
| E | 1-NN | 1.00 | 1.00 | 1.00 |
| | SVM | 1.00 | 1.00 | 1.00 |
| | BP | 0.77 | 0.68 | 1.00 |
| | RBF | 0.89 | 0.91 | 0.88 |
| | PNN | 0.52 | 0.52 | 0.52 |
| | LVQ2 | 1.00 | 1.00 | 1.00 |

Table 6.26: Average 10-fold cross-validation classification performance results for diverse databases and classifiers on the FA features of the NAMIC database.

experimental design we wanted to test if decreasing the size of the feature vectors had an impact on the classifiers performance. We found that performance was not affected down to the smallest feature vector (database E) where decreases in performance can be appreciated in all the classifiers for the FA data, while 1-NN, SVM and LVQ2 maintain their performance for MD data.

6.5.2 COBRE Schizophrenia

Our most recent studies are carried out on a rs-fMRI database released to the public from The Center for Biomedical Research Excellence in Brain Function and Mental Illness (COBRE) (see section A.3) for the study of image biomarkers of Schizophrenia, and the potential development of Computer Aided Diagnostic tools. We have also used the most recent open source and public collection of tools for the preprocessing rs-fMRI multivariate data with C-PAC B.8 .

6.5.2.1 Experiment 1

Features First we preprocess the data in four different ways within C-PAC: with temporal filtering (0.01-0.1Hz) and global signal regression [170] (TPF-GSR), only with temporal filtering (TPF), only with global signal regression (GSR) and without these two steps. From each of these we compute as the baseline data for feature selection scalar measures of brain functional activity described in Chapter 3, namely:

- Amplitude of Low Frequency Fluctuations (ALFF)
- Fractional Amplitude of Low Frequency Fluctuations (fALFF)
- Voxel-Mirrored Homotopic Connectivity (VMHC)
- Regional Homogeneity (ReHo)

The corresponding maps were used in the feature selection and feature vector extraction described in Chapter 4, using as voxel significance measures the Pearson's correlation coefficient, the Bhattacharyya distance, and the Welsh's t-test.

Classification methods A 10-fold cross-validation was carried out to test the classification performance, we stratified training and test set in order to have proportional number of controls and patients in each random disjoint set. Class weights were set proportionally to the number of subjects in each group in the training set. We perform a 3-fold cross-validation on the training set using each possible combination of parameter values in the search grid. This grid for the linear SVM was C in the range $[1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3]$ and for the RBF-SVM we tested C and γ in $[1e-3, 1e-2, 1e-1, 1, 1e1, 1e2, 1e3]$ as well. For Random Forest (RF) we tested for the number of trees in $[3, 5, 10, 30, 50, 100]$.

Classification Results of SVM and RF Tables 6.27,6.28,6.29 and 6.30 show the highest cross-validation mean accuracies and accuracy variances across and feature selection thresholds. The highest value of all is 80% with 0.02 variance using on the TPF-GSR ReHo data and Pearson's correlation. Taking into account all the showed scores, we observe that TPF-GSR and GSR ReHo measures show consistently higher performance in all experiments, mainly using Pearson's correlation and the Bhattacharyya distance.

| | | Pearson | Bhattacharyya | Welch t-test |
|-------|------------|--------------------|---------------|--------------|
| ALFF | Linear SVM | 0.65 (0.02) | 0.61 (0.01) | 0.60 (0.01) |
| | RBF SVM | 0.67 (0.01) | 0.54 (0.01) | 0.58 (0.02) |
| | RF | 0.72 (0.01) | 0.72 (0.02) | 0.62 (0.02) |
| fALFF | Linear SVM | 0.68 (0.01) | 0.63 (0.01) | 0.64 (0.01) |
| | RBF SVM | 0.64 (0.01) | 0.62 (0.01) | 0.64 (0.01) |
| | RF | 0.66 (0.02) | 0.63 (0.01) | 0.58 (0.01) |
| ReHo | Linear SVM | 0.73 (0.01) | 0.76 (0.01) | 0.64 (0.02) |
| | RBF SVM | 0.76 (0.02) | 0.64 (0.02) | 0.67 (0.02) |
| | RF | 0.80 (0.02) | 0.73 (0.01) | 0.67 (0.02) |
| VMHC | Linear SVM | 0.75 (0.01) | 0.74 (0.01) | 0.51 (0.01) |
| | RBF SVM | 0.58 (0.01) | 0.58 (0.01) | 0.64 (0.01) |
| | RF | 0.73 (0.02) | 0.76 (0.02) | 0.60 (0.02) |

Table 6.27: Maximum accuracies and variances across feature selection thresholds for each experiment on the TPF-GSR data.

| | | Pearson | Bhattacharyya | Welch t-test |
|-------|------------|--------------------|---------------|--------------|
| ALFF | Linear SVM | 0.65 (0.02) | 0.61 (0.01) | 0.60 (0.01) |
| | RBF SVM | 0.67 (0.01) | 0.54 (0.01) | 0.58 (0.02) |
| | RF | 0.68 (0.02) | 0.69 (0.02) | 0.60 (0.01) |
| fALFF | Linear SVM | 0.68 (0.01) | 0.63 (0.01) | 0.64 (0.01) |
| | RBF SVM | 0.64 (0.01) | 0.62 (0.01) | 0.64 (0.01) |
| | RF | 0.63 (0.01) | 0.62 (0.02) | 0.61 (0.01) |
| ReHo | Linear SVM | 0.70 (0.02) | 0.74 (0.02) | 0.63 (0.03) |
| | RBF SVM | 0.77 (0.01) | 0.64 (0.02) | 0.67 (0.03) |
| | RF | 0.78 (0.02) | 0.75 (0.01) | 0.66 (0.02) |
| VMHC | Linear SVM | 0.75 (0.01) | 0.74 (0.01) | 0.50 (0.01) |
| | RBF SVM | 0.63 (0.02) | 0.60 (0.01) | 0.62 (0.02) |
| | RF | 0.71 (0.01) | 0.67 (0.02) | 0.60 (0.02) |

Table 6.28: Maximum accuracies and variances across feature selection thresholds for each experiment on the GSR data.

We present the detailed exploration of the 3 best performing features. We report accuracy, sensitivity, specificity, precision, F1-score and area under the ROC Curve (AUC) for each combination of functional connectivity measure and feature selection method across the selection thresholds (feature vector size) in figures 6.17, 6.18, and 6.19.

Localization of selected features in the brain To provide a qualitative validation of the features, regarding its usefulness as image biomarkers, we localized the voxel sites selected by the feature selection process on the Harvard-Oxford cortical and subcortical atlas. Voxel sites localized are those appearing in 90% of the feature selection processes performed independently

| | | Pearson | Bhattacharyya | Welch t-test |
|-------|------------|--------------------|---------------|--------------|
| ALFF | Linear SVM | 0.68 (0.01) | 0.59 (0.01) | 0.59 (0.01) |
| | RBF SVM | 0.67 (0.01) | 0.56 (0.01) | 0.62 (0.01) |
| | RF | 0.68 (0.02) | 0.65 (0.01) | 0.65 (0.02) |
| fALFF | Linear SVM | 0.69 (0.01) | 0.65 (0.00) | 0.64 (0.02) |
| | RBF SVM | 0.67 (0.01) | 0.62 (0.02) | 0.65 (0.01) |
| | RF | 0.65 (0.02) | 0.66 (0.02) | 0.64 (0.01) |
| ReHo | Linear SVM | 0.72 (0.01) | 0.72 (0.02) | 0.64 (0.01) |
| | RBF SVM | 0.73 (0.02) | 0.63 (0.02) | 0.71 (0.02) |
| | RF | 0.73 (0.02) | 0.67 (0.02) | 0.66 (0.02) |
| VMHC | Linear SVM | 0.71 (0.01) | 0.69 (0.01) | 0.54 (0.02) |
| | RBF SVM | 0.57 (0.01) | 0.58 (0.01) | 0.61 (0.02) |
| | RF | 0.66 (0.02) | 0.70 (0.03) | 0.54 (0.03) |

Table 6.29: Maximum accuracies and variances across feature selection thresholds for each experiment on the TPF data.

| | | Pearson | Bhattacharyya | Welch t-test |
|-------|------------|--------------------|---------------|--------------|
| ALFF | Linear SVM | 0.68 (0.01) | 0.59 (0.01) | 0.59 (0.01) |
| | RBF SVM | 0.67 (0.01) | 0.56 (0.01) | 0.62 (0.01) |
| | RF | 0.70 (0.02) | 0.60 (0.01) | 0.62 (0.01) |
| fALFF | Linear SVM | 0.69 (0.01) | 0.65 (0.00) | 0.64 (0.02) |
| | RBF SVM | 0.67 (0.01) | 0.62 (0.02) | 0.65 (0.01) |
| | RF | 0.67 (0.02) | 0.68 (0.01) | 0.62 (0.01) |
| ReHo | Linear SVM | 0.70 (0.02) | 0.72 (0.02) | 0.60 (0.01) |
| | RBF SVM | 0.75 (0.02) | 0.61 (0.01) | 0.67 (0.02) |
| | RF | 0.73 (0.02) | 0.71 (0.02) | 0.68 (0.01) |
| VMHC | Linear SVM | 0.72 (0.01) | 0.69 (0.01) | 0.51 (0.00) |
| | RBF SVM | 0.58 (0.02) | 0.58 (0.00) | 0.61 (0.02) |
| | RF | 0.73 (0.02) | 0.70 (0.01) | 0.61 (0.02) |

Table 6.30: Maximum accuracies and variances across feature selection thresholds for each experiment on the data without TPF and GSR.

| Brain region | Bhattacharyya | | Pearson |
|--|---------------|----------|--------------|
| | TPF-GSR 95 % | GSR 95 % | TPF-GSR 90 % |
| Angular Gyrus | × | × | × |
| Cingulate Gyrus, anterior division | × | × | × |
| Cingulate Gyrus, posterior division | | × | |
| Cuneal Cortex | | | × |
| Frontal Medial Cortex | × | | |
| Frontal Orbital Cortex | × | × | × |
| Frontal Pole | × | | |
| Heschl's Gyrus (includes H1 and H2) | × | | × |
| Inferior Frontal Gyrus, pars opercularis | × | × | × |
| Inferior Frontal Gyrus, pars triangularis | | | × |
| Inferior Temporal Gyrus, anterior division | × | × | × |
| Inferior Temporal Gyrus, posterior division | × | × | × |
| Inferior Temporal Gyrus, temporooccipital part | × | | |
| Insular Cortex | × | × | × |
| Intracalcarine Cortex | × | × | × |
| Juxtapositional Lobule Cortex | | | × |
| Lateral Occipital Cortex, inferior division | × | × | × |
| Lateral Occipital Cortex, superior division | | | × |
| Lingual Gyrus | × | × | × |
| Middle Frontal Gyrus | | | × |
| Middle Temporal Gyrus, anterior division | × | × | × |
| Middle Temporal Gyrus, posterior division | × | × | × |
| Middle Temporal Gyrus, temporooccipital part | × | × | × |
| Occipital Fusiform Gyrus | × | × | × |
| Occipital Pole | × | × | × |

Table 6.31: Cortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (<http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/>), the Harvard-Oxford Cortical Structural Atlas and a cluster threshold of 95%.

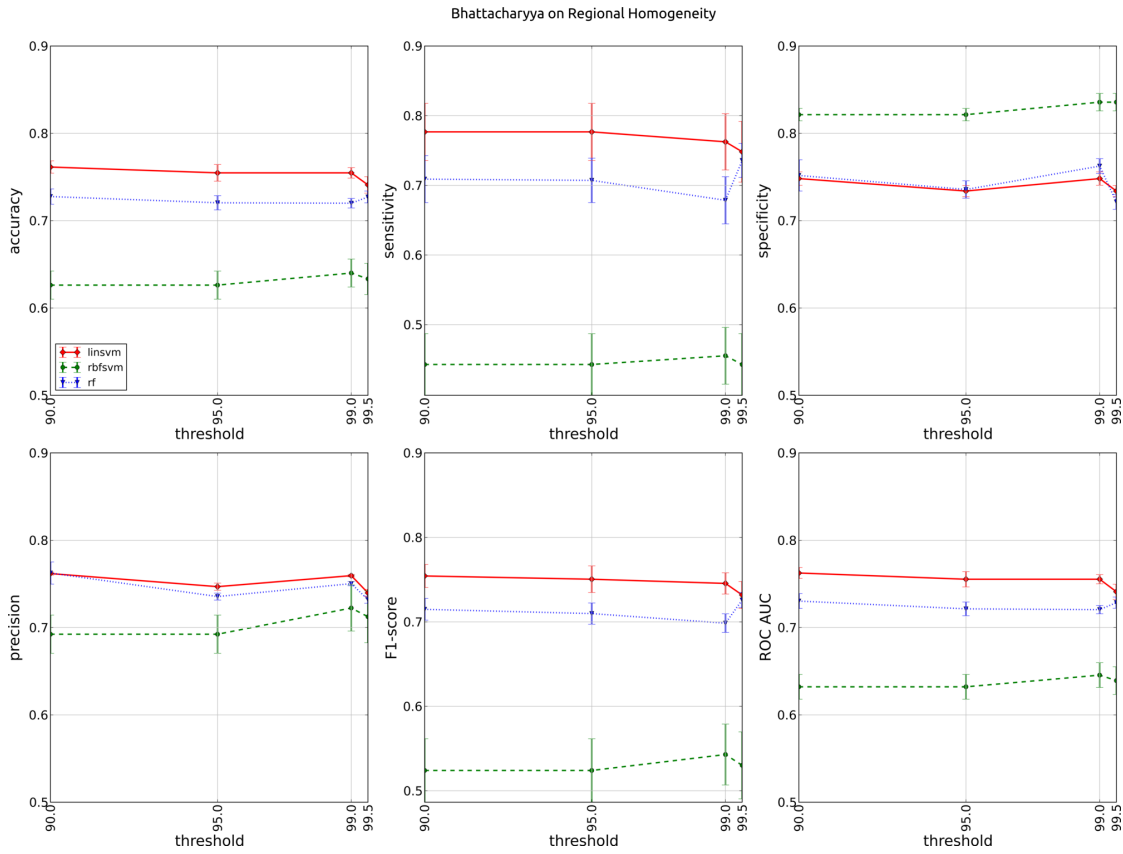


Figure 6.17: Classification performance using TPF-GSR ReHo data and the Bhattacharyya distance.

in each cross-validation fold. Feature locate in the cortex are reported in tables 6.31 and 6.32, while subcortical localizations are reported in table 6.33.

In figures 6.20, 6.21, and 6.22, we show some slices of the MNI152 template with an overlay with the 95% percentile voxel site localizations across the feature selection processes that were performed in the cross-validation folds. Figures show the findings using the three best combinations of functional activity and voxel significance measures.

6.5.2.2 Experiment 2

Features The functional activity maps from the TPF-GSR data considered in the previous experiment were also used restricting the voxel significance to the Pearson's correlation coefficient. In all functional activity measure instances the threshold was given by the 95% percentile of the empirical distribution of the correlation coefficient absolute values.

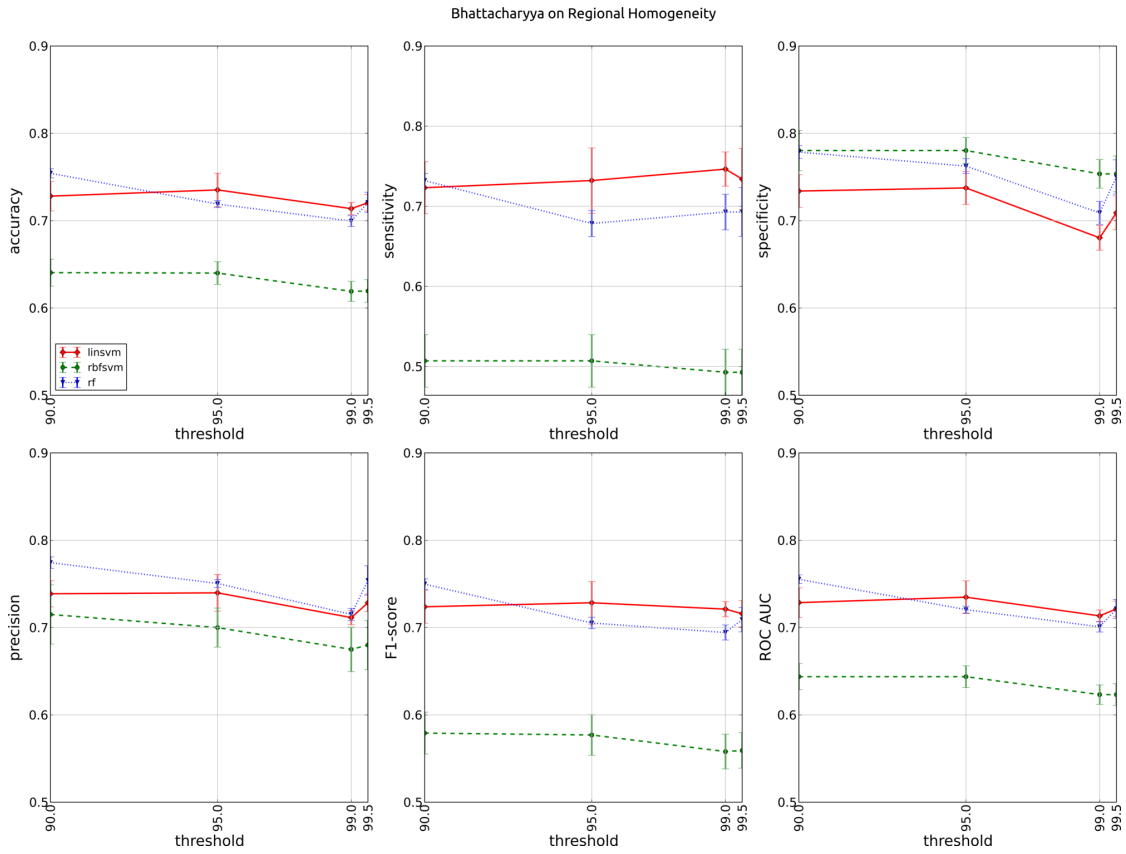


Figure 6.18: Classification performance using GSR ReHo data and the Bhattacharyya distance.

Classification methods The classification has been performed by ensembles of Extreme Learning Machines (V-ELM) combined by majority voting. The ensemble of ELM classifiers has three sensitive parameters: the number of classifiers M , their depth given by the number of hidden neuron and activation function $g(x)$. To get the optimal parameters we set up a first set of experiments. We set the number of ELM to $M = 101$, and the number of hidden units is chosen from the set $N \in \{5, 10, 100, 500, 1000\}$. Features correspond to the values of the functional connectivity maps at voxel sites whose absolute value of the Pearson's correlation coefficient is above the threshold corresponding to the 95% percentile of its empirical distribution. The Accuracy results of this exploration are shown in Tables 6.34, 6.35, 6.36, 6.37 for ReHo, ALFF, fALFF and VMHC functional activity maps, respectively. It can be appreciated that the “hardlim” and “sig” activation functions provide the best results. Regarding the number of hidden units, it seems that 100 hidden units provide the best trade-off between performance and model complexity for all activation functions and functional connectivity features.

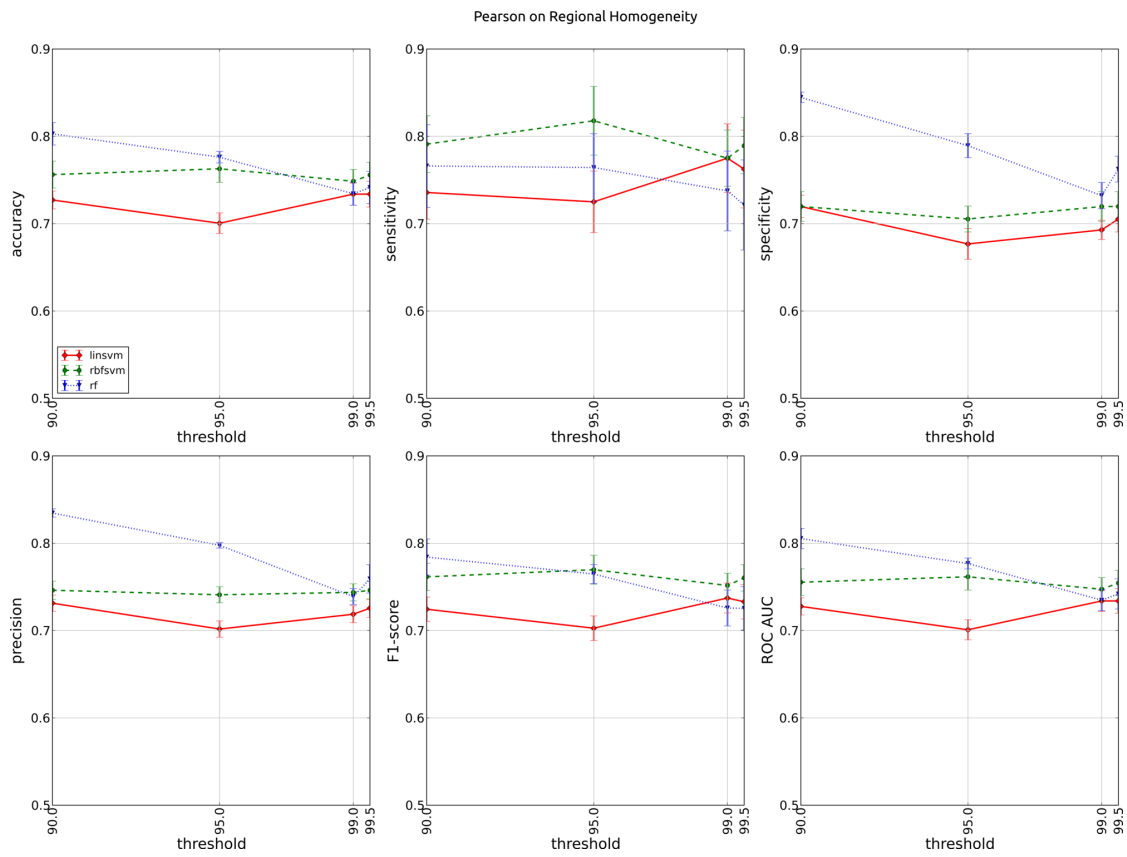


Figure 6.19: Classification performance using TPF-GSR ReHo data and the Pearson correlation.

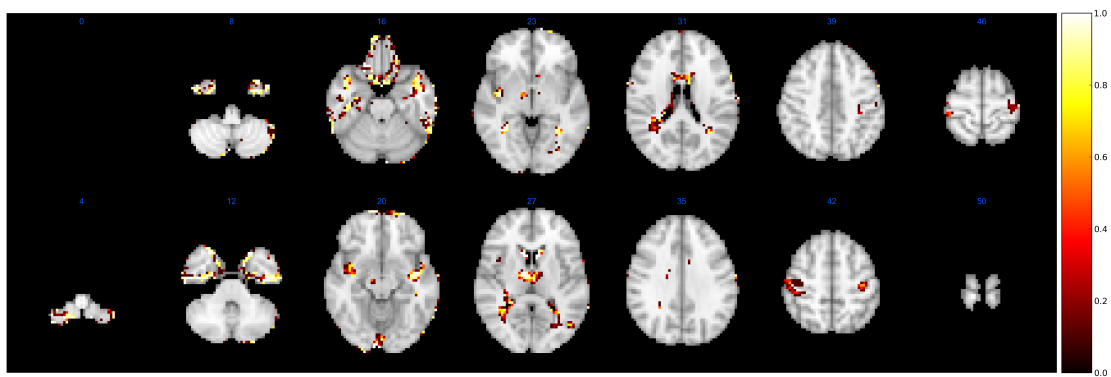


Figure 6.20: Voxels selected from the TPF-GSR ReHo data using Bhattacharyya distance thresholded at 95%. The colormap indicates the voxel selection frequency during cross-validation.

| Brain region | Bhattacharyya | | Pearson |
|--|---------------|---------|-------------|
| | TPF-GSR 95% | GSR 95% | TPF-GSR 90% |
| Parahippocampal Gyrus, anterior division | × | × | × |
| Parahippocampal Gyrus, posterior division | | × | × |
| Parietal Operculum Cortex | | × | × |
| Planum Polare | × | × | × |
| Postcentral Gyrus | × | × | × |
| Precentral Gyrus | × | × | × |
| Precuneous Cortex | × | × | × |
| Subcallosal Cortex | × | | |
| Superior Frontal Gyrus | | | × |
| Superior Parietal Lobule | × | × | × |
| Superior Temporal Gyrus, anterior division | × | × | × |
| Superior Temporal Gyrus, posterior division | × | × | × |
| Supracalcarine Cortex | × | × | × |
| Supramarginal Gyrus, anterior division | × | × | × |
| Supramarginal Gyrus, posterior division | × | × | × |
| Temporal Fusiform Cortex, anterior division | × | × | × |
| Temporal Fusiform Cortex, posterior division | × | × | × |
| Temporal Occipital Fusiform Cortex | × | × | × |
| Temporal Pole | × | × | × |

Table 6.32: (cont.) Cortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (<http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/>), the Harvard-Oxford Cortical Structural Atlas and a cluster threshold of 95%.

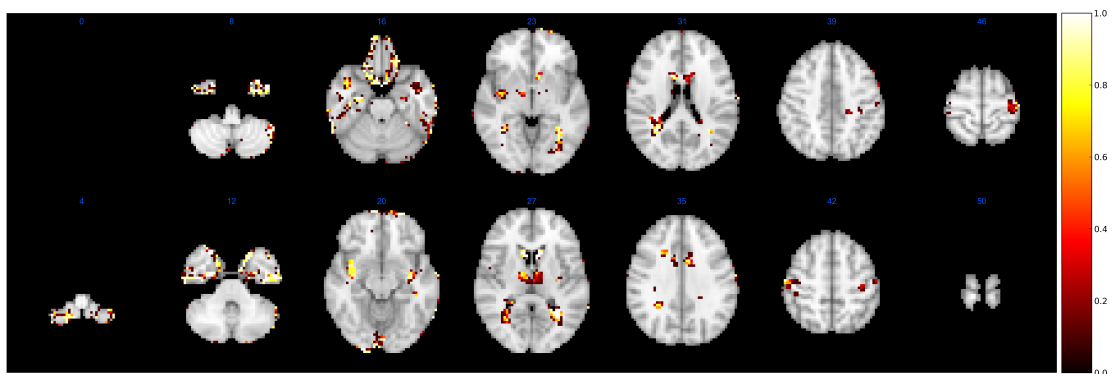


Figure 6.21: Voxels selected from the GSR ReHo data using Bhattacharyya distance thresholded at 95%. The colormap indicates the voxel selection frequency during cross-validation.

| Brain region | Bhattacharyya | | Pearson |
|-----------------------------|---------------|---------|-------------|
| | TPF-GSR 95% | GSR 95% | TPF-GSR 90% |
| Brain-Stem | | | × |
| Left Accumbens | × | × | |
| Left Amygdala | × | × | |
| Left Caudate | × | × | × |
| Left Cerebral Cortex | × | × | × |
| Left Cerebral White Matter | × | × | × |
| Left Hippocampus | × | × | |
| Left Lateral Ventrical | × | × | × |
| Left Pallidum | × | × | × |
| Left Putamen | × | × | |
| Left Thalamus | × | × | × |
| Right Accumbens | | × | |
| Right Amygdala | × | × | × |
| Right Caudate | × | × | × |
| Right Cerebral Cortex | × | × | × |
| Right Cerebral White Matter | × | × | × |
| Right Hippocampus | × | × | × |
| Right Lateral Ventricle | × | × | × |
| Right Pallidum | × | × | × |
| Right Putamen | × | | × |
| Right Thalamus | × | × | × |

Table 6.33: Subcortical regions which are overlapped by voxels selected from the feature selection methods with best performance. These regions were obtained using autoaq (<http://brainder.org/2012/07/30/automatic-atlas-queries-in-fsl/>), the Harvard-Oxford Subcortical Structural Atlas and a cluster threshold of 95%.

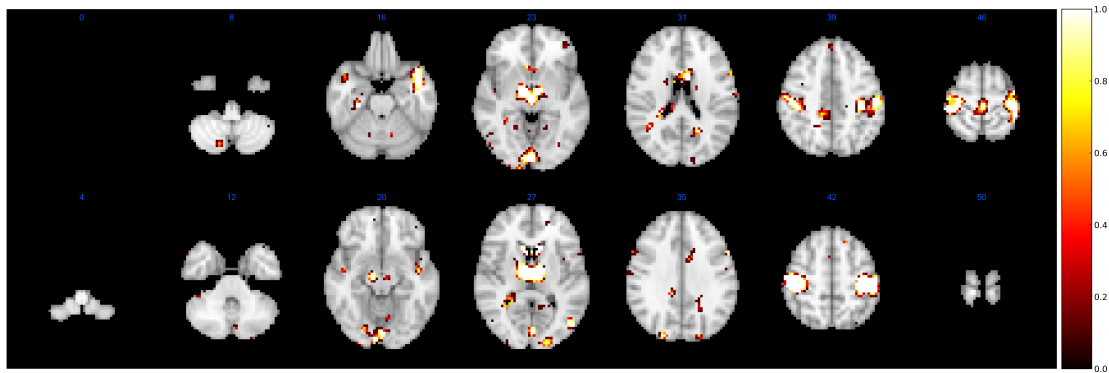


Figure 6.22: Voxels selected from the TPF-GSR ReHo data using Pearson correlation thresholded at 90%. The colormap indicates the selection frequency during cross-validation of each voxel.

| | 5 | 10 | 100 | 500 | 1000 |
|----------------|-------|-------|--------------|--------------|--------------|
| sig | 74.76 | 79.48 | 82.90 | 77.24 | 78.67 |
| sin | 56.52 | 53.43 | 58.05 | 57.33 | 56 |
| hardlim | 75.24 | 77.29 | 82.86 | 84.19 | 83.67 |
| tribas | 49.14 | 49.19 | 63 | 75.57 | 78.67 |
| radbas | 52.14 | 67.86 | 71.24 | 79.43 | 79.19 |

Table 6.34: Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over ReHo functional connectivity map. Bold entries are above 80%.

| | 5 | 10 | 100 | 500 | 1000 |
|----------------|--------------|--------------|--------------|--------------|--------------|
| sig | 77.57 | 81.48 | 87.67 | 81.57 | 82.33 |
| sin | 56 | 54.76 | 54.24 | 52.76 | 53.33 |
| hardlim | 80.05 | 81.71 | 84.86 | 86.43 | 85.76 |
| tribas | 49.19 | 49.14 | 73.33 | 82.33 | 81 |
| radbas | 72.14 | 71.05 | 77.52 | 82.33 | 81.43 |

Table 6.35: Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over ALFF functional connectivity map. Bold entries are above 80%.

| | 5 | 10 | 100 | 500 | 1000 |
|----------------|-------|-------|--------------|--------------|--------------|
| sig | 74.62 | 76.81 | 81.48 | 81.38 | 80.29 |
| sin | 59.52 | 54.76 | 55.48 | 54.71 | 52.29 |
| hardlim | 73.29 | 78.19 | 80.19 | 82.19 | 80.81 |
| tribas | 47.90 | 49.29 | 68.57 | 72.62 | 74.76 |
| radbas | 56.43 | 62.19 | 68.52 | 72.76 | 75.76 |

Table 6.36: Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over fALFF functional connectivity map. Bold entries are above 80%.

| | 5 | 10 | 100 | 500 | 1000 |
|----------------|-------|-------|--------------|--------------|--------------|
| sig | 67.67 | 80.05 | 89.00 | 88.52 | 88.95 |
| sin | 58.10 | 53.43 | 56.24 | 58.10 | 55.57 |
| hardlim | 68.67 | 76.67 | 88.52 | 91.19 | 90.48 |
| tribas | 49.48 | 49.24 | 69.90 | 81.52 | 83.00 |
| radbas | 56.19 | 64.05 | 75.43 | 86.48 | 86.29 |

Table 6.37: Effect of the number of hidden units (columns) and activation function (rows) in the accuracy of the V-ELM over VMHC functional connectivity map. Bold entries are above 80%.

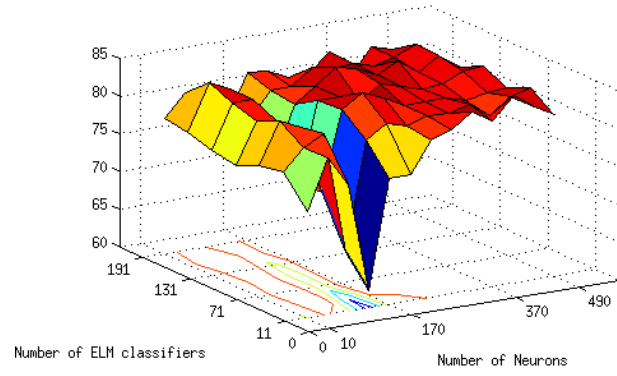
Combined effect of the ensemble size and the number of hidden units. Considering the “hardlim” as the best activation function we proceed to explore the combined effect of the ensemble size and the number of hidden units. We vary the number of classifiers from 10 to 500 in steps of 40, i.e. $M \in \{10, 50, \dots, 500\}$, and the number of hidden units from 11 to 211 in steps of 30, i.e. $N = \{11, 41, \dots, 211\}$. We focus on two of the functional connectivity maps, Figures 6.23 and 6.24 show the 3D response surfaces for the accuracy, sensitivity and specificity of the V-ELM applied to the Pearson’s features extracted from the ReHo and ALFF functional connectivity maps, respectively. The results are averaged accuracy, sensitivity and specificity over 10 trials using 10-fold cross validation method and maximum over 10 repetitions. Notice that the z-axis corresponding to the accuracy, sensitivity or specificity has not the same values for all the plots, so that comparisons can not be immediately extracted by mere visual inspection. There is a common effect of a deep drop in performance for a small ensemble size and a medium number of hidden units (about 170), after that, the system seems to be saturated and does not improve with further addition of ELM or hidden units. The profile of the response surface is quite similar for both, so we may extrapolate that them to the other functional connectivity maps (i.e. VHMC and fALFF).

6.5.2.3 Experiment 3

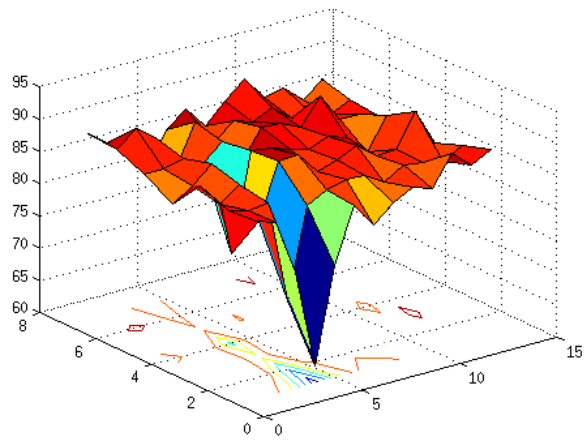
This work has been carried out in collaboration with Darya Chyzyk, aside from her thesis main topics.

Evolutionary wrapper features In this section we look for the application of the evolutionary feature selection wrapper (EW) introduced in Section 4.7 to the selection of a parsimonious and significant set of features, using the V-ELM as the classifier algorithm in the loop. The evolutionary wrapper may depart from two sets of baseline features: the raw functional connectivity map (f.c.m.) and the Pearson’s correlation features. Table 6.38 gives the specific feature vector size for each approach. The most salient features selected by EW over Pearson’s correlation features are selected orderly, the first n features ordered by saliency. On the other hand, the most salient features selected by EW over raw functional connectivity map are selected applying a threshold given by the 95% percentile of the voxel saliency e.c.d.f.

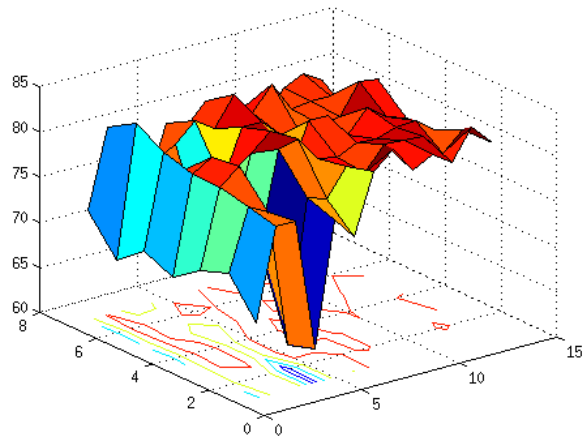
Results Table 6.39 provides the 10-fold cross-validation accuracy over the diverse features. It can be appreciated that the EW selection improves the performance of the classification with a much parsimonious representation, so that the voxel sites have a more pronounced biomarker value. The EW over the raw functional connectivity map does not improve the other approaches, either selecting the same number of salient features as the Pearson’s approach, or selecting the 100 most salient features. Therefore, saliency may be a convenient way to obtain more parsimonious representations for other functional connectivity maps. Table 6.40 shows the plots of the response surface when varying the ensemble size and the number of hidden units. For the



(a)

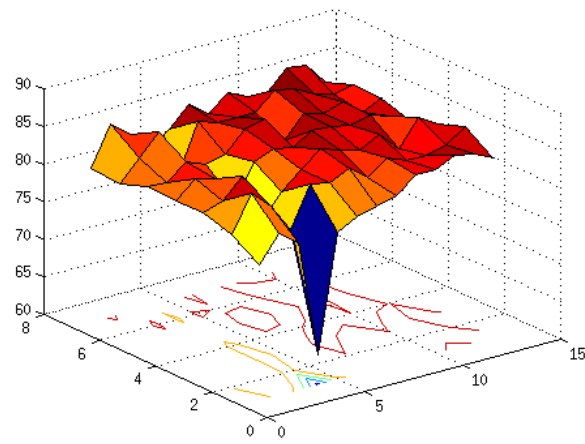


(b)

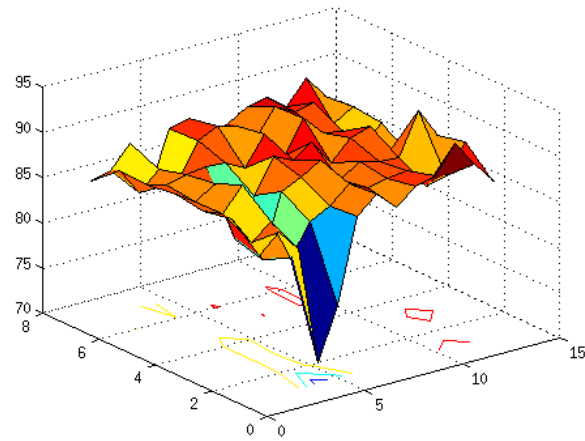


(c)

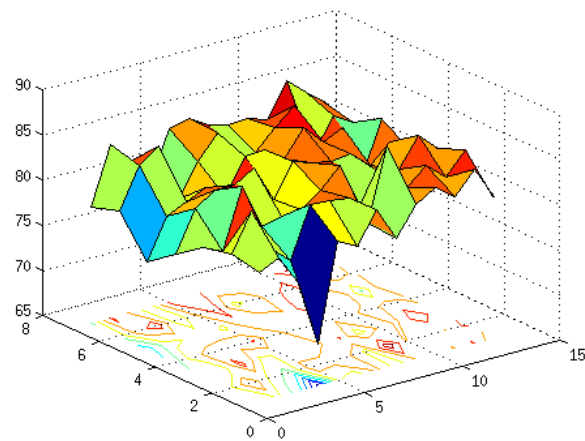
Figure 6.23: Effect of the number of ELM and the number of hidden units in the (a) Accuracy, (b) Sensitivity, (c) Specificity of ELM ensemble for the ReHo functional connectivity map.



(a)



(b)



(c)

Figure 6.24: Effect of the number of ELM and the number of hidden units in the (a) Accuracy, (b) Sensitivity, (c) Specificity of ELM ensemble for the ALFF functional connectivity map.

| | ReHo | ALFF |
|---|---------------|--------|
| Total number of features | 86 559 | 86 559 |
| Removing voxel sites with more than 70% of zeros | 68 666 | 68 592 |
| Removing voxel sites with more than 80% of zeros | 70 573 | 70 497 |
| Pearson's correlation selection | 2 113 | 2 198 |
| Selecting most salient features by EW over raw f.c.m. | 2 113 | 2 198 |
| Selecting most salient features by EW over Pearson | 100, 300, 500 | |

Table 6.38: Feature vector size for the selected functional connectivity maps.

| | EW over raw f.c.m. | | Pearson | EW over Pearson | | |
|--------------------|--------------------|-------|--------------|-----------------|--------------|--------------|
| number of features | 100 | 2113 | 2113 | 100 | 300 | 500 |
| ReHo | 77.3 | 81.52 | 84.9 | 85.71 | 87 | 87.05 |
| number of features | 100 | 2198 | 2198 | 100 | 300 | 500 |
| ALFF | 67.42 | 70.62 | 86.95 | 81.33 | 86.24 | 86.43 |

Table 6.39: 10-fold cross-validation of Accuracy of V-ELM over the different feature selection methods. By columns, evolutionary wrapper over raw functional connectivity map (f.c.m.).

Pearson's features, the deep decrease of performance for 170 hidden units is preserved whatever the EW process applied afterwards. For the EW applied to the raw data, the results are poor, showing that the EW does not extract discriminant voxel sites *per se*.

Feature localization in the brain Finally, we give the localization of the selected features on the brain template. Figure 6.25 provides the localization of the ReHo features for each of the three cases. It can be appreciated that the EW produces quite disperse and unexpressive voxel site localizations. Pearson's correlation produces more compact and biologically plausible localizations, related with the literature findings reported in the introduction. The application of the EW to the Pearson's features produces an sparsification of the localizations, which is pending evaluation by the medical experts. Figure 6.26 shows the corresponding localizations of the ALFF features. Same comments apply to these images. The overlapping of the Pearson's correlations, which are some kind of gold standard in this context, is relatively high between the two visualized functional connectivity map selections, confirming the general value of the analysis.

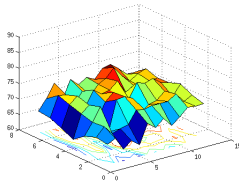
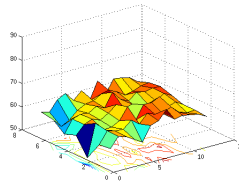
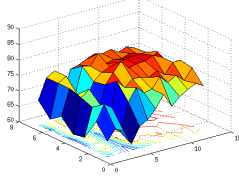
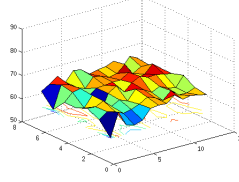
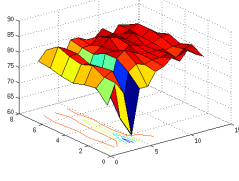
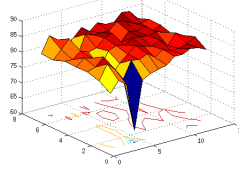
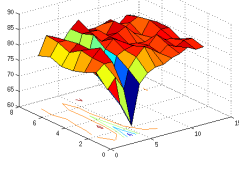
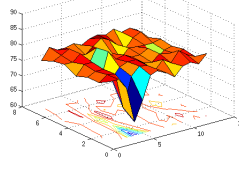
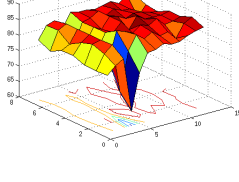
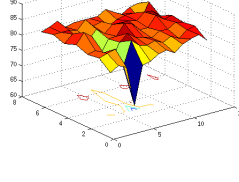
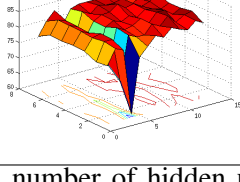
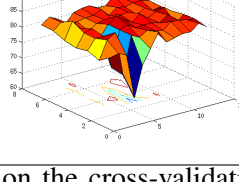
| feature selection | number of features | functional connectivity map | |
|-------------------|--------------------|---|---|
| | | ReHo | ALFF |
| EW f.c.m. | 100 |  |  |
| | 2113/2198 |  |  |
| Pearson | 2113/2198 |  |  |
| EW + Pearson | 100 |  |  |
| | 300 |  |  |
| | 500 |  |  |

Table 6.40: Effect on the ensemble size and number of hidden units on the cross-validation Accuracy for the different feature selection and feature sizes. x/x means the feature vector size per functional connectivity map. Comparison of EW over f.c.m., Pearson selection and EW over Pearson's features

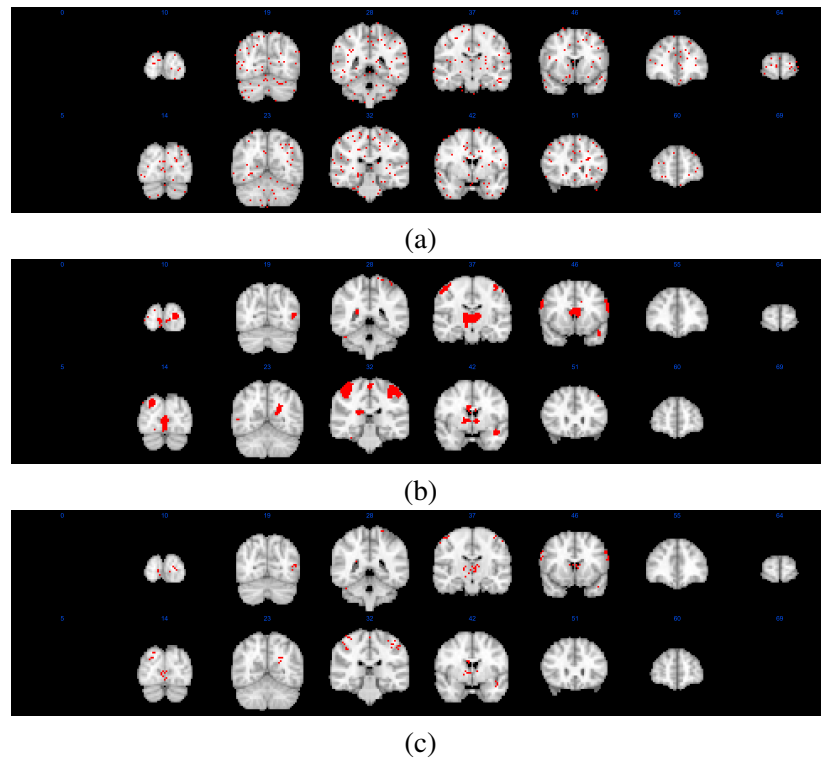


Figure 6.25: ReHo features localization: (a) 2113 features selected by EW over the raw f.c.m, (b) Pearson's 2113 features, (c) 500 selected by EW over Pearson's features.

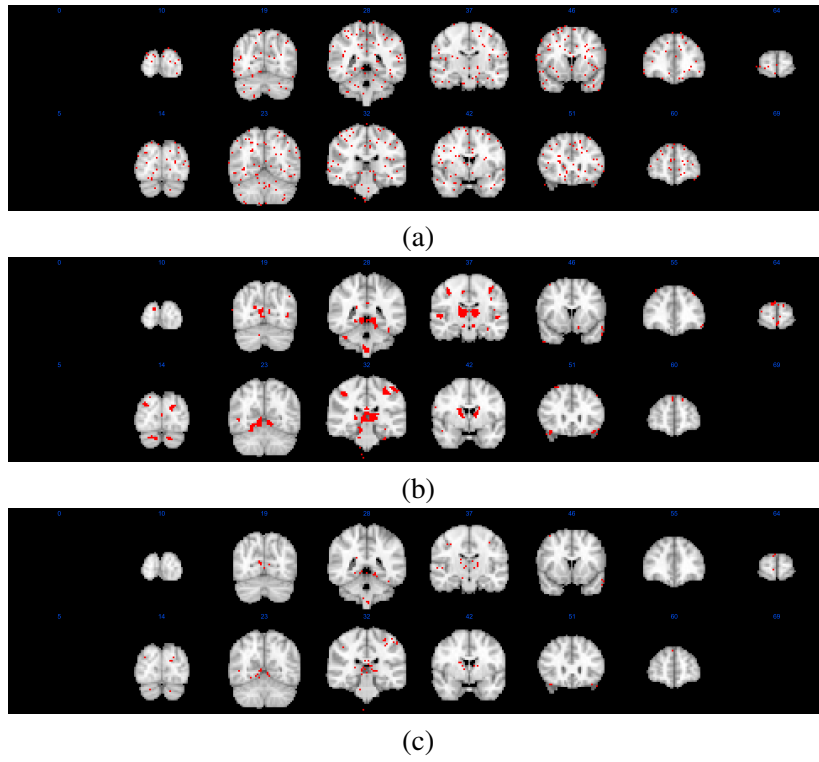


Figure 6.26: ALFF features localization: (a) 2198 features selected with EW over the f.c.m., (b) Pearson's 2198 features, (c) 500 features selected by EW over the Pearson's features.

Chapter 7

Conclusions

In this Chapter we gather the conclusions of the thesis. Section 7.1 provides general methodological conclusions. Section 7.2 details the conclusions related to each neurological disease. Section 7.3 gives the conclusions from the operational point of view.

7.1 Methodological conclusions

We have extensively studied the application of machine learning to different modalities of MRI data. During the evolution of the studies, we found the double dipping or circularity of analysis issue. When dealing with classification processes, preserving these independences can be stated simply as the training of the classifier not contaminated by any information on the test set. We have built and tested a general computational pipeline that can be applied to the various data modalities contemplating the special case of supervised feature selection. This pipeline involves the repetition of the feature selection process at each fold of the cross-validation process. It is free from circularity effects, so that no bias is introduced in the results.

Regarding the general approach we have tested the application of machine learning to several studies on neurological diseases, finding good classification results that encourage the realization of CAD systems for the various cases. Moreover, anatomical feature localizations are in good agreement with the findings reported in the literature, validating the approach from the medical point of view.

7.2 Disease-related conclusions

The research in this thesis has two aspects when considering the application to specific neurological diseases: on one hand, the validation of CAD system proposals for these diseases and on the other hand the identification of localizations in the brain that may be associated with image biomarkers for these diseases.

7.2.1 Alzheimer's Disease (AD)

The study of AD classification from the point of view of anatomical MRI data has been performed initially on a reduced balanced database extracted from the OASIS database. The initial feature selection was the VBM but afterwards we considered DBM based features. The main difference is that VBM provided voxel sites corresponding to clusters found in SPM, while DBM provided the basis for a feature selection process. Studies on DBM features were extended over the entire OASIS database. We think that appropriate pre-processing of the data is of paramount importance and can not be disregarded trusting that ensuing statistical or machine learning processes may cope with the errors introduced by lack of appropriate data normalization.

Anatomical data: VBM related results The first feature selection procedure we have tested consists in considering the clusters findings of the VBM as the voxel sites for features. We have built the feature datasets, publishing them, and applied a host of classification algorithms on them. We have found that sensitivity results are systematically lower than specificity results. Detailed examination of the results reveals that the decrease of sensitivity is due to the bad recognition results on the AD subjects at early stages of the disease. The trend is common to all the classification algorithms tested, so this lack of sensitivity may be either attributed to the feature selection method, or to the inherent difficulty to discriminate prodromal states of AD on anatomical MRI data. Careful design and tuning of classification algorithms, i.e. AdaBoost, for this problem achieves performances close to the 90%, despite the low sensitivity results. Some additional experiments on this data performed by other researchers in the group have achieved performances above 90% on this feature vectors pointing to its use in CAD system for AD.

The classifiers applied to the data features make decisions based on the whole of the feature vector. However, the locations of the feature sources, the voxel cluster found by VBM, has some clinical interpretations. For AD the voxel cluster findings were mostly located in the temporal lobe, though some voxels in the frontal lobe have also been detected. These results confirm previous findings of temporal lobe atrophic changes in AD [22, 42], and suggest that these abnormalities may be confined to specific sites within that lobe, rather than showing a widespread distribution.

Anatomical data: Atlas based features results The experiment involving 7 statistical features extracted from 116 GM brain ROIs defined according to the AAL atlas was performed over the entire OASIS database. The goal of the experiment was to assess the performance of features built using *a priori* ROI maps against those experiments with supervised methods of feature selection. Unsupervised feature selection methods will most probably show worse classification performance than those which are supervised, but lead to more generalized systems and less fitted to the experimental database. We are aware that registration and segmentation errors can lead to biases in the accuracy of the classifiers. The results are in agreement with most of our previous classification experiments [52, 175, 173].

Anatomical data: deformation based results The computational experiments aimed to assess the utility of the features selected from deformation fields, and modulated GM, obtained from non-linear registration processes is one of the most comprehensive experimental designs in the thesis. It covers all combinations of five scalar measures of the deformation and three voxel site significance measures. The classification results over the OASIS database were above 90% accuracy in some instances, promising a good way to develop CAD systems for AD. Differences in classification performance between deformation measures were not statistically significant for two way t-test ($p > 0.1$). These results are unexpected, because DBM approaches assume that the Jacobian maps provide information enough to characterize the anatomical deformations to detect atrophy and other effects of AD and other neurodegenerative diseases.

Post hoc analysis of the voxel sites selected for features, performing an additional selection of those appearing 90% of the cross-validation experiment instances. The regions from the different experiments are all within similar brain areas: frontal and parietal lobes, cerebellum, temporal and occipital lobes, frontal pole, lateral occipital cortex, superior division, precentral gyrus, postcentral gyrus and hippocampus. All these findings agree with the imaging biomarkers reported in the literature [53, 55, 145, 61, 62].

Diffusion data Experiments on diffusion data were carried out over a database provided by the Hospital de Santiago. The features extracted from the data were the FA and MD. Classification experiments showed that this data modality is well suited to build CAD systems for AD, achieving *Accuracy*=99,82%, *Sensitivity*=100% and *Specificity*=99,68% in a 10-fold cross-validation on FA data. Voxel sites selected to build the feature vectors were localized in the following regions of the brain ¹: the caudate [135] and insula (in agreement with [91]) but also in the temporo-parietal lobe (in agreement with [171]), frontal lobe and thalamus for the FA data. In the MD data the most discriminant voxel sites were located in the temporal lobe (in agreement with [41]), insula, temporo-parietal lobe, and frontal lobe. Therefore, voxel site selection for classification is in agreement with previous studies even when no *a priori* information has been used in the feature selection process.

7.2.2 Schizophrenia

Diffusion data Using FA and MD scalar measures of diffusivity as the basic data, we have been able to define feature selection and extraction processes, and perform classification experiments, mostly using Neural Networks. The selected voxels roughly correspond to findings reported in the medical literature. Surprisingly, all the classifiers obtain near perfect results. Despite the simplicity of our feature extraction process, the results compare well with other results found in the literature [44, 206]. The main limitation of this study is that the results come from a small database. Therefore, more extensive testing will be needed to confirm our conclu-

¹Using “atlasquery” tool from FSL (<http://www.fmrib.ox.ac.uk/fsl/>)

sions. Nevertheless, we are making available ² the actual data employed in the computational experiments to allow for independent validation of our results.

The feature voxel locations ³ were different for FA and MD maps. In the case of FA, the selected voxels were localized mainly in parietal and temporal lobes, but also in the cerebellum and occipital lobe. More specifically, in WM we found discriminant voxel values in the cingulum bundle, superior and inferior longitudinal fasciculus and in the inferior fronto-occipital fasciculus. On the other hand, in the MD maps, the most discriminant voxel values were the ones localized in frontal and parietal lobes, more specifically the cingulum bundle, inferior fronto-occipital and longitudinal fasciculus, and superior longitudinal fasciculus.

Functional data The computational experiments on functional MRI have been computed on the COBRE database (c.f. Appendix A). We have computed several functional activity measures as the basis for feature selection and extraction. Classification results of Schizophrenia patients versus controls are very optimistic, reaching values close to 90% in some instances. The best classification results were obtained by the evolutionary wrapper selection of the features, using voting ELM as the classification building algorithm. Feature localization is in agreement with the literature in many places. In this case we have performed an exhaustive image biomarker localization reported in Chapter 7. The most salient result is that the Regional Homogeneity seems to be appropriate for feature selection discriminating Schizophrenia patients. Although the global signal regression is a criticized pre-processing technique [75, 170] for connectivity analysis, it obtained the best classification performance for local measures of brain activity.

7.3 Operational conclusions

From the operational point of view, this thesis is an extensive exercise in the management of large image data of diverse modalities related to specific diseases. As such, we have developed many software pipelines for data management as well as for the realization of the computational experiments. We have followed in the measure of our possibilities the policy of sharing our software and data in order to enhance the reproducibility of our studies, and foster new experiments and results on our datasets. Specifically, we have:

- Published scripts of software usage that specify precisely the realization of our computational experiments. This allows both third party verification of the implementations, application to new data with minimal adjustments, and reproduction of our results on the same data. The summary description is given in the Appendix B, and the actual scripts are published in the author's web page.

²<http://www.ehu.es/ccwintco/index.php/GIC-experimental-databases>

³This specification of the voxel locations were obtained with the “atlasquery” tool from FMRIB's FSL (<http://www.fmrib.ox.ac.uk/fsl/>) using the “MNI Structural Atlas” and the “JHU White-Matter Tractography Atlas”.

- Published the datasets corresponding to some of the feature vectors extracted from the MRI data, allowing both for the repetition of our experiments and the realization of new experiments with innovative machine learning algorithms.
- For the realization of some experiments we have achieved in-house implementation of image processing algorithms, such as DBM, in a small cluster of computers. These implementations are extended to larger clusters, such as the i2Basque clusters available to the research community in the Basque Country <http://www-es.i2basque.es/index.php/Portada>.
- One original classification algorithm has been proposed along the lines of the AdaBoost specifically adapted to the SVM classification of independent features from disjoint clusters or AAL ROIs. This algorithm reported improvements over the state of the art results.
- We proposed, implemented and tested a wrapper feature selection based on voting ELMs. This approach provided the best results on the COBRE database.

Appendix A

Databases

This Appendix is devoted to the description of some of the datasets that have been used in the experimental works along the thesis. From the medical point of view we have treated imaging data from several diseases: Alzheimer’s Disease, Schizophrenia and Myotonic Dystrophy Type 1. Sections A.1, A.2 and A.3 describe three public available databases that have been processed and analyzed in this thesis. Sections A.4 and A.5 show details on databases from research project collaborations with the Neurology Department of the Donostia Hospital (San Sebastián, Spain) and the Hospital de Santiago Apóstol (Vitoria-Gasteiz, Spain).

A.1 Open Access Series of Imaging Studies (OASIS)

In this study we used all the subjects of the first Open Access Series of Imaging Studies (OASIS) [137]. These subjects were selected from a larger database of individuals who had participated in MRI studies at Washington University, they were all right-handed and older adults had a recent clinical evaluation. Older subjects with and without dementia were obtained from the longitudinal pool of the Washington University Alzheimer Disease Research Center (ADRC). This release of OASIS consists of a cross-sectional collection of 416 male (119 controls and 41 patients) and female (197 controls and 59 patients) subjects aged 18 to 96 years (218 aged 18 to 59 years and 198 subjects aged 60 to 96 years). Further demographic and image acquisition details can be found in [137]. The database includes at least 3 raw anatomical MP-RAGE images from each subject as well as post-processed images: (a) corrected for interscan head movement and rigidly aligned to the Talairach and Tournoux space [195], (b) transformed to a template with a 12-parameter affine registration and merged into a 1-mm isotropic image, (c) skull-stripped and corrected for intensity inhomogeneity and (d) segmented by tissue type. To carry out our experiment we used the volumes from (c).

For specific classification experiments we have used a feature dataset which has been computed from a selected collection subjects from the Open Access Series of Imaging Studies (OASIS) database [137](<http://www.oasis-brains.org/>). OASIS data set has a cross-sectional col-

lection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease.

The database includes at least 3 raw anatomical MP-RAGE images *per* subject as well as post-processed images in diverse stages normalization. To carry out our experiments, our point of departure are the MRI volumes which have been corrected for inter-scan head movement, rigidly aligned to the Talairach and Tournoux space [195], spatially registered to the MNI152 template with a 12-parameter affine transformation, resampled into a 1-mm isotropic resolution, skull-stripped, and corrected for intensity inhomogeneity.

Imaging Protocol The OASIS database has been built following a strict imaging protocol, to avoid variations due to imaging protocol which would pose big image normalization problems. Multiple (three or four) high-resolution structural T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) images were acquired [74] on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters: TR=9.7 msec., TE=4.0 msec., Flip angle=10, TI=20 msec., TD=200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256x256 (1x1mm).

A.1.1 OASIS98

Ninety eight right-handed women (aged 65-96 yr) were selected from the OASIS database [137][2]. It has a cross-sectional collection of 416 subjects covering the adult life span aged 18 to 96 including individuals with early-stage Alzheimer's Disease. We have ruled out a set of 200 subjects whose demographic, clinical or derived anatomic volumes information was incomplete. For the present study there are 49 subjects who have been diagnosed with very mild to mild AD and 49 non demented. A summary of subject demographics and dementia status is shown in table A.1.

Many of the classification studies on the detection of AD were done with both men and women. However, it has been demonstrated that brains of women are different from men's to the extent that it is possible to discriminate the gender via MRI analysis [127]. Moreover, it has been shown that VBM is sensitive to the gender differences. For these reasons, we have done this selection of the OASIS database. It must be noted that this is a large number of subjects compared with the other studies referred above.

A.1.2 Whole OASIS

In this study we used all the subjects of the first Open Access Series of Imaging Studies (OASIS) [137]. This release of OASIS consists of a cross-sectional collection of 416 male (119 controls and 41 patients) and female (197 controls and 59 patients) subjects aged 18 to 96 years (218 aged 18 to 59 years and 198 subjects aged 60 to 96 years). One hundred subjects older than 60 have

| | With Dementia | Without Dementia |
|----------------------|---------------|------------------|
| No. of subjects | 49 | 49 |
| Age | 78.08 (66-96) | 77.77 (65-94) |
| Education | 2.63 (1-5) | 2.87 (1-5) |
| Socioeconomic status | 2.94 (1-5) | 2.88 (1-5) |
| CDR (0.5 / 1 / 2) | 31 / 17 / 1 | 0 |
| MMSE | 24 (15-30) | 28.96 (26-30) |

Table A.1: Summary of subject demographics and dementia status. Education codes correspond to the following levels of education: 1 less than high school grad., 2: high school grad., 3: some college, 4: college grad., 5: beyond college. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best).

| Age Group | Total n | Without Dementia | | | With Dementia | | | |
|-----------|---------|------------------|----------|---------|---------------|----------|-------|-------------|
| | | n | Mean age | M/F* | n | Mean age | M/F* | CDR 0.5/1/2 |
| <20 | 19 | 19 | 18.53 | 10/9 | 0 | | 0/0 | 0/0/0 |
| 20s | 119 | 119 | 22.82 | 51/68 | 0 | | 0/0 | 0/0/0 |
| 30s | 16 | 16 | 33.38 | 11/5 | 0 | | 0/0 | 0/0/0 |
| 40s | 31 | 31 | 45.58 | 10/21 | 0 | | 0/0 | 0/0/0 |
| 50s | 33 | 33 | 54.36 | 11/22 | 0 | | 0/0 | 0/0/0 |
| 60s | 40 | 25 | 64.88 | 7/18 | 15 | 66.13 | 6/9 | 12/3/0 |
| 70s | 83 | 35 | 73.37 | 10/25 | 48 | 74.42 | 20/28 | 32/15/1 |
| 80s | 62 | 30 | 84.07 | 8/22 | 32 | 82.88 | 13/19 | 22/9/1 |
| ≥90 | 13 | 8 | 91.00 | 1/7 | 5 | 92.00 | 2/3 | 4/1/0 |
| Total | 416 | 316 | | 119/197 | 100 | | 41/59 | 70/28/2 |

Table A.2: Age and Diagnosis Characteristics of the Data Set. *Male/Female

been clinically diagnosed with very mild to moderate Alzheimer’s disease (70 with CDR=0.5, 28 with CDR=1 and 2 with CDR=2). Table A.2 summarizes the demographics of the dataset.

A.2 NAMIC: Brain Multimodality

Structural MRI and DTI data from twenty men (aged 21-55 yr), ten patients and ten controls, from a publicly available database from the National Alliance for Medical Image Computing (NAMIC) (<http://hdl.handle.net/1926/1687>) were the subjects of this study in this experiment. Table A.3 summarizes the demographics of this dataset.

Imaging protocol For the anatomical MRI volume measurements, images have been acquired using a 3T GE at Brigham and Women’s Hospital (BWH) in Boston, MA. A 8 Channel coil has been used in order to perform parallel imaging using ASSET (Array Spatial Sensitivity Encoding

| | Patients | Controls |
|-----------------------------|-------------------|-------------------|
| Age | 43.20 \pm 10.30 | 42.20 \pm 11.40 |
| Handedness | 0.72 \pm 0.27 | 0.78 \pm 0.18 |
| Education (years) | 13.11 \pm 2.31 | 15.55 \pm 2.24 |
| Socioeconomic status | 3.10 \pm 1.28 | 1.67 \pm 0.50 |
| Parent socioeconomic status | 2.50 \pm 1.27 | 1.89 \pm 1.05 |
| MMSE | 28.80 \pm 1.69 | 29.4 \pm 0.73 |

Table A.3: Summary of subject demographics and dementia status. Categories of socioeconomic status: from 1 (biggest status) to 5 (lowest status). MMSE score ranges from 0 (worst) to 30 (best). Handedness is a continuous measure from 1 (right) to 0 (left).

techniques, GE) with a SENSE-factor (speed-up) of 2. The structural MRI acquisition protocol includes two MRI pulse sequences. The first results in contiguous spoiled gradient-recalled acquisition (fastSPGR) with the following parameters; TR=7.4ms, TE=3ms, TI=600, 10 degree flip angle, 25.6cm² field of view, matrix=256x256. The voxel dimensions are 1x1x1 mm. The second-XETA (eXtended Echo Train Acquisition) produces a series of contiguous T2-weighted images (TR=2500ms, TE=80ms, 25.6cm² field of view, 1 mm slice thickness).

Diffusion Tensor Imaging (DTI) scans have been acquired on a 3T GE system using an echo planar imaging (EPI) DTI Tensor sequence. A double echo option to reduce eddy-current related distortions have been used. To reduce impact of EPI spatial distortion, an 8 Channel coil allowed to perform parallel imaging using ASSET (Array Spatial Sensitivity Encoding Techniques, GE) with a SENSE-factor (speed-up) of 2. Fifty one directions with b=900 and 8 baseline scans with b=0 have been acquired. The following scan parameters are used: TR 17000 ms, TE 78 ms, FOV 24 cm, 144x144 encoding steps, 1.7 mm slice thickness. Eighty five axial slices parallel to the AC-PC line have been acquired, covering the whole brain. In addition, B₀ field inhomogeneity maps are collected.

A.3 COBRE Schizophrenia

The Center for Biomedical Research Excellence in Brain Function and Mental Illness (COBRE) (<http://cobre.mrn.org/>) has published raw anatomical and functional MR data from 72 patients with Schizophrenia and 75 healthy controls (ages ranging from 18 to 65 in each group). Diagnostic information was collected using the Structured Clinical Interview used for DSM Disorders (SCID).

Anatomical MRI, resting-state fMRI and phenotypic data including: gender, age, handedness and diagnostic information, are released for every participant. Table A.4 shows a short summary of the demographic data included in COBRE. Table A.5 and A.6 show subjects diagnoses for each group.

| | Controls | Patients |
|-----------------|--------------|--------------|
| No. of subjects | 74 | 72 |
| Age | 35.8 (18-65) | 38.1 (18-65) |
| Gender (M/F) | 51/23 | 58/24 |

Table A.4: Summary of subject demographics.

| Diagnosis | None | 311 | 296.26 |
|-----------------|------|-----|--------|
| No. of subjects | 72 | 1 | 1 |

Table A.5: Summary of control subjects diagnoses.

Imaging protocol For anatomical imaging a multi-echo MPRAGE (MEMPR) sequence was used with the following parameters: TR/TE/TI=2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900ms, flip angle=7°, FOV=256x256mm, slab thickness=176mm, matrix=256x256x176, voxel size =1x1x1mm, number of echos=5, pixel bandwidth=650Hz, total scan time=6min. With 5 echoes, the TR, TI and time to encode partitions for the MEMPR are similar to that of a conventional MPRAGE, resulting in similar GM/WM/CSF contrast. Resting state functional MRI (rs-fMRI) data was collected with single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction using the intercommissural line (AC-PC) as a reference (TR: 2s, TE: 29ms, matrix size: 64x64, 32 slices, voxel size: 3x3x4mm).

A.4 Myotonic Dystrophy Type 1

The MD1 patients analysed in this work were selected from those attending in the outpatient consultancies at the Neurology Department of the Donostia Hospital (San Sebastián, Spain), a tertiary public hospital which covers a population of 650,000 inhabitants (almost all of Guipuzcoa province). All patients were explored by a neurologist and had previously participated in another study in which they were neuropsychologically [187].

A healthy control matched in terms of age and sex was included for each of the DM1 patients included. This control group consisted of unaffected family members and healthy volunteers with none of these pathologies. All patients were informed of the objectives and details of the study and signed an informed consent. The study was approved by the hospital's ethics committee. The socio-demographic characteristics and main symptoms of the subjects included in this study are listed in Table A.7.

The DM1 patients analysed in this work were selected from those attending the outpatient

| Diagnosis | 290.3 | 295.1 | 295.2 | 295.3 | 295.6 | 295.7 | 295.70 | 295.9 | 295.92 | 296.4 |
|-----------------|-------|-------|-------|-------|-------|-------|--------|-------|--------|-------|
| No. of subjects | 1 | 3 | 1 | 41 | 12 | 3 | 2 | 5 | 1 | 1 |

Table A.6: Summary of patient subjects diagnoses.

| | | MD1 | CS |
|---|-----------|-------------|-------------|
| Socio-demographic characteristics | | | |
| Number of subjects | | 30 | 30 |
| Age | Mean (SD) | 44.0 (11.6) | 44.2 (11.7) |
| | Min-Max | 24-62 | 22-62 |
| Sex n (%) | Male | 14 (47%) | 14 (47%) |
| | Female | 16 (53%) | 16 (53%) |
| Educational level n (%) | Primary | 18 (60%) | 5 (21%) |
| | Secondary | 7 (23%) | 9 (37%) |
| | Higher | 5 (17%) | 10 (41%) |
| Clinical and molecular characteristics | | | |
| Muscle weakness (MIRS ¹) | Mean (SD) | 2.9 (1.2) | – |
| | Min-Max | 1-5 | |
| Molecular defect (CTG) | Mean (SD) | 635 (472) | – |
| | Min-Max | 65-1833 | |
| White matter lesions n (%) | Yes | 16 (53%) | 5 (18%) |
| | No | 14 (47%) | 22 (82%) |

Table A.7: Summary of subject demographics and MD1 status. ¹Muscular Impairment Rating Scale.

consultancies at the Neurology Department of the Donostia Hospital (San Sebastián), a tertiary public hospital which covers a population of 650,000 inhabitants (almost all of Guipuzcoa province). The patient-selection criteria were as follows:

- Inclusion criteria for DM1 patients: Between 18 and 65 years old and molecular confirmation of the clinical diagnosis.
- Exclusion criteria: A history of a major psychiatric or somatic disorder (in accordance with DSM-IV criteria), acquired brain damage or alcohol or drug abuse, and the presence of cerebral anomalies which could affect the volumetric analysis. See figure A.1.

Imaging protocol MR scanning was performed on a 1.5 Tesla scanner (Achieva Nova, Philips). The current results are based on a high-resolution volumetric “turbo field echo” (TFE) series (sagittal 3D T1 weighted acquisition, TR=7.2, TE=3.3, flip angle=8, matrix=256 x 232, slice thickness 1mm, voxel dimensions of 1mm x 1mm x 1mm, NSA=1, n° slices=160, slice thickness=1, gap=0, total scan duration 5’34”). In addition to 3D T1-weighted images, conventional axial dual T2 turbo spin echo images (TR=1800, TE=20, flip angle=90°, FOV=230, matrix=256 x 154, slice thickness=5, gap=1, n° acquisition=2, n° slices=22) and fluid attenuation inversion recovery images (FLAIR) in coronal plane (TR=10000, TI=2800, TE=140, FOV=230, matrix=256x159, slice thickness=5, gap=1, n° acquisition=2, n° slices=25) were acquired to evaluate the presence of white matter lesions (WMLs). All the scans were acquired on the same

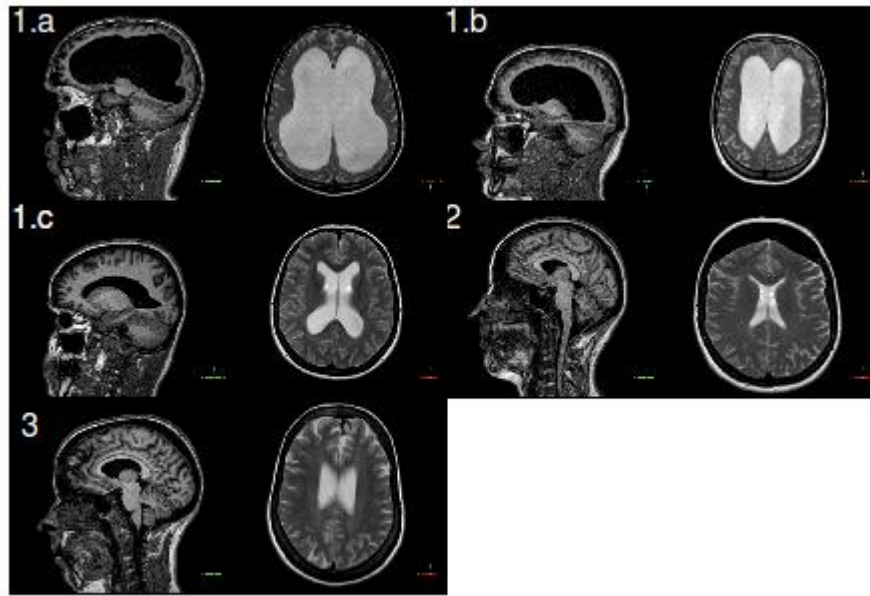


Figure A.1: DM1 patients excluded from the study: sagittal T1 and transversal T2 images. 1.a. and 1.b. Grade 5 ventricular dilation (44-year-old male with 333 CTG and 41-year-old female with 667 CTG respectively), 1.c. grade 3 ventricular dilation (28-year-old female with 500 CTG), 2. Frontal hyperostosis (36-year-old female with 400 CTG) and 3. Frontal calcification (38-year-old female with 833 CTG).

MR scanner and no hardware or software upgrades of the equipment were carried out within the study period.

The DM1 patients showed significant differences with respect to the control group in terms of educational level and they had significantly more white matter lesions than the controls. An analysis of the presence of white matter lesions (WMLs) showed that 53% of DM1 patients had focal lesions ranging from subcortical and deep to confluent, whereas such lesions were only identified in 18% of the controls. Lesions can be appreciate in the example T2 images shown in figure A.2.

A.5 Hospital de Santiago

Forty five men and women (aged 60-89), twenty five healthy controls and twenty AD patients, from Hospital de Santiago Apóstol (Vitoria-Gasteiz, Spain), were recruited for this study.

Imaging protocol Structural MRI and DTI data were used to validate the proposed CAD approach. MR scanning was performed on a 1.5 Tesla scanner (Magnetom Avanto, Siemens). Study protocol consists of 3D T1-weighted acquisition (isotropic 1x1x1mm, 176 slices, TR=1900ms,

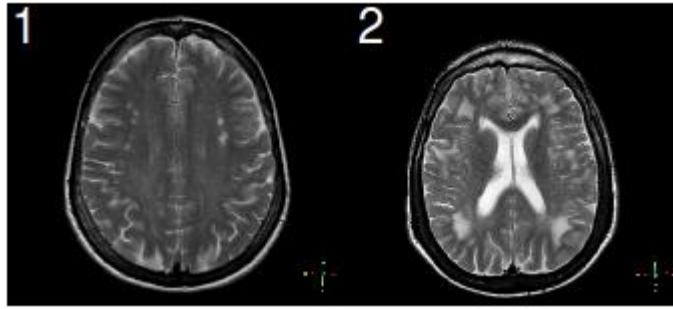


Figure A.2: Patients with white matter lesions; transversal T2 images: 1) 52-year-old female with 333 CTG and both subcortical and deep focal lesions. 2) 44-year-old male with 667 CTG and confluent subcortical lesions.

TE=337ms and FOV=256/75%), a 3D Flair sequence (isotropic 1x1x1mm, 176 slices, TR=5000ms, TE=333ms and FOV=260/87.5%) and diffusion tensor imaging (isotropic 2.5x2.5x2.5mm, 40 slices, 30% gap, TR=5400ms, TE=88ms, resolution FOV 320/100% , $b=1000 \text{ s} \cdot \text{mm}^{-2}$ and 30 gradient directions).

Appendix B

Software

“Software is a central part of modern scientific discovery and the public availability of code is essential of the scientific method, as it is a requirement to reproducing scientific results” [163].

B.1 NeuroDebian Repository[163]

The NeuroDebian repository is a software repository that provides a large collection of popular neuroscience research software for the Debian operating system (OS) as well as Ubuntu and other derivatives [94] (<http://neuro.debian.net/>). Popular packages include FSL (Section B.3), AFNI (Section B.4), PyMVPA and many others. It also offers a virtual machine with Debian configured to use this repository, which can be an excellent way to maintain an analysis environment that remains identical throughout the lifetime of a study.

This repository makes of these free (libre) OS distributions very attractive for neuroscience research by simplifying installation and update procedures, controlling deployed software versions, also avoiding the need of compilation and software testing before deployment in some cases.

B.2 SPM

The Statistical Parametric Mapping (SPM) [79](<http://www.fil.ion.ucl.ac.uk/spm/>) is a freely available suite of Matlab (<http://www.mathworks.com/>) functions and subroutines with a comprehensive Graphical User Interface (GUI), which has made it very popular between the less technical researchers, i.e. psychologists and clinicians. It has been implemented for the analysis of brain imaging data sequences. The current release is designed for the analysis of fMRI, positron emission tomography (PET), single-photon emission computerized tomography (SPECT), electroencephalography (EEG) and magnetoencephalography (MEG).

It has been used in this thesis for VBM experiments on the OASIS98 database (see section A.1.1). Although SPM is distributed under the General Public License (GPL) with its source

code, it needs Matlab for execution which is a privative software. In consequence we discarded SPM from our research in the last publications.

B.3 FSL

The FMRIB Software Library (FSL) [108, 188] (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>) is a comprehensive library of analysis tools for FMRI, MRI and DTI brain imaging data. It runs on Apple and PCs (both Linux, and Windows via a Virtual Machine), and is very easy to install. Some of the tools can be run both from the command line and as GUIs ("point-and-click" graphical user interfaces). To quote the relevant references for FSL tools you should look in the individual tools' manual pages, and also please reference the FSL overview papers:

- M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith. FSL. *NeuroImage*, 62:782-90, 2012
- M.W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, S.M. Smith. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45:S173-86, 2009
- S.M. Smith, M. Jenkinson, M.W. Woolrich, C.F. Beckmann, T.E.J. Behrens, H. Johansen-Berg, P.R. Bannister, M. De Luca, I. Drobnjak, D.E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J.M. Brady, and P.M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208-19, 2004

Comments on usage Although there are GUIs for the most used tools, they are not very useful when you have to repeat the same process for many subjects in one experiment. In addition, there are many options in the command line tools that haven't been added to the GUI. If you run a GUI application from the console, e.g. Flirt, it will show what command line execution it performs once you press Run. In this manner, if you plan to use it more than once you can use the GUI the first time to copy the command line and know which arguments does the GUI uses by default. Reading the command line help of the commands is very important when starting for understanding all the processing options you have before starting an experiment. In order to see the command help open a console, write the command name and press ENTER, it will show information on how to use the command, its arguments and their meaning. Using FSL through the command line, as an alternative to a GUI, gives more options but also more computational power, i.e., the possibility of running the commands in batch or in parallel, conditional and looping constructs and all the operation controls a programming language as Bash (<http://www.gnu.org/software/bash/manual/bashref.html>) can give.

Brain Extraction Tool (BET) is a widely used skull-stripping software tool for T1 and T2-weighted images. Often the resulting segmentation is not good enough with the default parameter values, so that you need to perform some manual parameter tuning. Before taking that long pursuit, please take a careful look to the image you want to skull-strip. Often, too much neck, face and noise can lead BET to wrong segmentations. Manually editing the image by eliminating all you can of the neck and face, without eliminating any chunk of brain, can improve the segmentation without changing the parameters.

B.4 AFNI

AFNI [56] (<http://afni.nimh.nih.gov/>) is a free software set of C programs for processing, analyzing, and displaying fMRI data. It runs on Unix systems, including SGI, Solaris, Linux, and Mac OS X. It is available free (in C source code format, and some precompiled binaries) for research purposes.

In this thesis there weren't many experiments on fMRI, the only AFNI tool directly used for this thesis was 3dresample which offers the easiest way for resampling a volume using the `-dxyz` argument. AFNI has mostly been indirectly used through C-PAC for the experiment on the Schizophrenia COBRE dataset.

In GNU/Linux, I recommend installing AFNI by downloading the binaries from their website, unpacking it in a folder and adding it to the execution path variable by editing the `~/.bashrc` file or `/etc/bash.bashrc`.

B.5 FreeSurfer

FreeSurfer [72] is a set of automated tools for reconstruction of the brain's cortical surface from structural MRI data, and overlay of functional MRI data onto the reconstructed surface. We have not made any use of it.

B.6 LibSVM

LIBSVM [?] (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) is an integrated software for support vector classification, (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM). It supports multi-class classification.

Since version 2.8, it implements an SMO-type algorithm proposed in this paper: R.-E. Fan, P.-H. Chen, and C.-J. Lin. Working set selection using second order information for training SVM. *Journal of Machine Learning Research* 6, 1889-1918, 2005. You can also find a pseudo code there. (how to cite LIBSVM)

Our goal is to help users from other fields to easily use SVM as a tool. LIBSVM provides a simple interface where users can easily link it with their own programs. Main features of

LIBSVM include

Different SVM formulations Efficient multi-class classification Cross validation for model selection Probability estimates Various kernels (including precomputed kernel matrix) Weighted SVM for unbalanced data Both C++ and Java sources GUI demonstrating SVM classification and regression Python, R, MATLAB, Perl, Ruby, Weka, Common LISP, CLISP, Haskell, OCaml, LabVIEW, and PHP interfaces. C# .NET code and CUDA extension is available. It's also included in some data mining environments: RapidMiner, PCP, and LIONSolver. Automatic model selection which can generate contour of cross validation accuracy.

B.7 SVM-Perf

SVMperf [110] http://www.cs.cornell.edu/people/tj/svm_light/svm_perf.html is an implementation of the Support Vector Machine (SVM) formulation for optimizing multivariate performance measures. Furthermore, SVMperf implements the alternative structural formulation of the SVM optimization problem for conventional binary classification with error rate and ordinal regression. So, there are three reasons for why you might want to use SVMperf instead of SVMlight:

Optimize binary SVM classification rules directly to ROC-Area, F1-Score, and Precision/Recall Break-Even Point. Train conventional linear classification SVMs optimizing error rate in time that is linear in the size of the training data through an alternative, but equivalent formulation of the training problem. This can be much faster than SVMlight for large training sets. Train conventional linear ordinal regression SVMs optimizing the number of misordered pairs in time that is $O(n \log(n))$ in the size of the training data through an alternative, but equivalent formulation of the training problem. Currently this is implemented only for rankings with two ranks. Training sparse kernel SVMs via the Cutting-Plane Subspace Pursuit (CPSP) algorithm. The algorithm enables large-scale training of non-linear kernel SVMs and fast predictions on test examples. This is achieved by limiting the number of support vectors and by allowing support vectors (or, more precisely, basis functions) that are not necessarily training vectors. The new features of version V3.00 are described here. This implementation is an instance of SVMstruct. More information on SVMstruct is available here.

B.8 C-PAC

The Configurable Pipeline for the Analysis of Connectomes (C-PAC) [1] (<http://fcp-indi.github.io/>) is an open-source software pipeline for automated pre-processing and analysis of resting-state fMRI data. C-PAC builds upon a robust set of existing software packages including AFNI, FSL, and FreeSurfer, and makes it easy for both novice users and experts to explore their data using a wide array of analytic tools. Users define analysis pipelines by specifying a combination of

preprocessing options and analyses to be run on an arbitrary number of subjects. Results can then be compared across groups using the integrated group statistics feature.

C-PAC has been designed to reliably pre-process and analyze data for hundreds of subjects in a single run, either on a single machine or on a compute cluster using Sun Grid Engine, HTCondor (thanks to the author of this thesis), or Portable Batch System.

Different analysis pipelines can produce significantly different results, raising questions about the reproducibility and reliability of brain imaging findings. C-PAC makes it easy to explore the impact of particular processing decisions by allowing users to run a factorial number of analysis pipelines, each with a different set of preprocessing and analysis options. An integrated Quality Control interface facilitates rapid manual examination of pipeline outputs and selection of subjects to be included in group comparisons.

- Preprocessing Options
 - Skull Stripping
 - Template-Based Registration
 - Automatic Tissue Segmentation
 - Anatomical / Functional Coregistration
 - Volume Realignment
 - Slice Timing Correction
 - Intensity Normalization
 - Temporal Filtering
 - Nuisance Signal Correction
 - Median Angle Correction
 - Spatial Smoothing
 - Motion Scrubbing
- Analytic Tools
 - Seed-based Correlation Analysis
 - Amplitude of Low Frequency Fluctuations (ALFF) and fALFF
 - Regional Homogeneity
 - Voxel-Mirrored Homotopic Connectivity
 - Timeseries Extraction
 - Network Centrality
 - Connectome-Wide Association Studies

- Dual Regression
- Bootstrap Analysis of Stable Clusters
- Group Comparisons

B.9 Python Libraries

In the last few years of this PhD the author has been able to shift from Matlab to Python. In this way towards free software, the author improved the readability and reusability of his code, and also contributed to a few other software projects (C-PAC and scikit-learn). In addition, a set of tools for performing similar experiments

B.9.1 Numpy and Scipy

Numpy (<http://www.numpy.org/>) along with Scipy [65] (<http://www.scipy.org/>) are the fundamental packages for scientific computing in Python. Numpy provides: a powerful N-dimensional array object, sophisticated (broadcasting) functions, tools for integrating Fortran and C/C++ code and also useful linear algebra, Fourier transform and random number capabilities. On the other hand, Scipy provides many user-friendly and efficient numerical routines such as routines for numerical integration and optimization.

For a smoother transition from Matlab to Numpy, there is a webpage about the differences, mostly for the purpose of helping proficient MATLAB users become proficient NumPy and SciPy users (http://wiki.scipy.org/NumPy_for_Matlab_Users).

B.9.2 NiBabel

NiBabel (<http://nipy.sourceforge.net/nibabel/>) is a Python package that provides read and write access to some common medical and neuroimaging file formats, including: ANALYZE (plain, SPM99, SPM2), GIFTI, NIfTI1, MINC, MGH and ECAT as well as PAR/REC. It can read and write Freesurfer geometry, and read Freesurfer morphometry and annotation files. There is some very limited support for DICOM.

The various image format classes give full or selective access to header (meta) information and access to the image data is made available via NumPy arrays.

B.9.3 Scikit-Learn

Scikit-Learn [158] (<http://scikit-learn.org/>) is a recent Python package integrating classique machine learning algorithms which aims to provide simple and efficient solutions to learning problems that are accessible to everybody and reusable in various contexts: machine-learning as a versatile tool for science and engineering. It is currently a very active community-oriented software project.

The simplicity of the code and object-oriented design allows to benefit very easily of software patterns in order to create experiments with a vast set of parameters to tune in an elegant piece of source code. It includes:

- Clustering algorithms
- Covariance estimators
- Cross-validation, Grid search and pipelining helpers
- Matrix decomposition algorithms
- Ensemble methods, e.g., RF and AdaBoost
- Feature selection helpers
- Hidden Markov Models
- Gaussian processes
- Linear Discriminant Analysis
- Generalized Linear Models
- Gaussian Mixture Models
- Multiclass and multilabel classification helpers
- Performance metrics
- Nearest Neighbors
- Support Vector Machines
- Decision Trees

B.9.4 Author's software releases

B.9.4.1 Aizkolari

Aizkolari (<https://github.com/alexsavio/aizkolari>) is a set of tools in Python created by Alexandre Savio to measure mathematical distances from different groups of Nifti images, e.g., brain MRI to do a supervised feature selection for supervised classification with SVMPerf or scikit-learn. Special care has been taken to avoid circularity effects in the validation process. It makes use of Numpy, Scipy, NiBabel, scikit-learn, Matplotlib, SVMPerf and FSL 5.0. One of the reasons to build this tool set was to be able to measure group differences in very big datasets when available RAM memory wasn't enough to load the whole dataset. In the feature extraction

process it performs a slice-by-slice group comparison to finally obtain a volume with group difference measures. New additions to this toolset has made easier to change the group difference measures, the classification algorithm and also adding other feature selection processes available in scikit-learn. It has been only tested on Ubuntu 12.10 and 13.04.

B.9.4.2 Other Python tool-sets

In the group website (<http://www.ehu.es/ccwintco/index.php/Portada>) and in the author's GitHub account (<https://github.com/alexsavio>) the author has released a few Python tool-sets and Bash shell scripts which have been used for the computational experiments performed in this thesis. Specifically:

visualize_volume This is a small Python library which contains some brain volume visualization functions taking advantage of Matplotlib and Mayavi2. The slices images made in experiments with COBRE database were done with these. The functions using Mayavi2 also export 3D geometrical models of brain volumes with volume overlapping, although they show impressive results they are not useful for publication.

oasis_feets This is a set of Python scripts to extract data from brain MRI and apply different classifiers. It also contains a feature extraction method which calculate deformation 3D histograms, not used yet. This library makes use of Aizkolari (c.f. B.9.4.1).

oasis_aal This is the source code for the experiment performed with the AAL atlas and the deformation field features. It has been also programmed in Python and uses Aizkolari. Here, the classification and feature extraction algorithms setup allows to easily perform a wide variety of pipelines, including most of the supervised algorithms of scikit-learn (c.f. B.9.3).

Bibliography

- [1] C-PAC by FCP-INDI. <http://fcp-indi.github.io/>.
- [2] Open Access Series of Imaging Studies (OASIS) available at <http://www.oasis-brains.org/>.
- [3] Statistical Parametric Mapping available at <http://www.fil.ion.ucl.ac.uk/spm/>.
- [4] M. Graña J. Villanúa A. Savio, M. García-Sebastián. Results of an Adaboost approach on Alzheimer's disease detection on MRI. *Bioinspired applications in Artificial and Natural Computation*. J. Mira, J. M. Ferrández, J.R. Alvarez, F. dela Paz, F.J. Toledo (Eds.) LNCS 5602, pages 114–123, 2009.
- [5] K. Abe, H. Fujimura, K. Toyooka, S. Yorifuji, Y. Nishikawa, T. Hazama, and T. Yanagihara. Involvement of the central nervous system in myotonic dystrophy. *Journal of the Neurological Sciences*, 127(2):179–185, December 1994. PMID: 7707077.
- [6] W. J. Adie and J. G. Greenfield. Dystrophia myotonica (myotonia atrophica). *Brain*, 46:73–127, 1923.
- [7] H Seznecothers. Agbulut, N. Sergeant, C. Savouret, A. Ghestem, et al. Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Human Molecular Genetics*, 10(23):2717–2726, November 2001.
- [8] Andone Sistiaga Darya Chyzyk Esther Fernandez Fermin Moreno Elsa Fernandez Manuel Grana Jorge Villanua Alexandre Savio, Maite Garcia-Sebastian and Adolfo Lopez de Munain. Machine learning approach for myotonic dystrophy diagnostic support from MRI. In Juan Manuel Górriz, Elmar W. Lang, and Javier Ramírez, editors, *Recent Advances in Biomedical Signal Processing*, pages 141–148. BENTHAM SCIENCE PUBLISHERS, March 2012.
- [9] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural computation*, 9(7):1545–1588, 1997.

- [10] N. Andreone, M. Tansella, R. Cerini, A. Versace, G. Rambaldelli, C. Perlini, N. Dusi, L. Pelizza, M. Balestrieri, C. Barbui, M. Nosè, A. Gasparini, and P. Brambilla. Cortical white-matter microstructure in schizophrenia. diffusion imaging study. *The British Journal of Psychiatry: The Journal of Mental Science*, 191:113–119, August 2007. PMID: 17666494.
- [11] G. Antonini, C. Mainero, A. Romano, F. Giubilei, V. Ceschin, F. Gragnani, S. Morino, M. Fiorelli, F. Soscia, A. Di Pasquale, and F. Caramia. Cerebral atrophy in myotonic dystrophy: a voxel based morphometric study. *Journal of Neurology, Neurosurgery & Psychiatry*, 75(11):1611–1613, November 2004.
- [12] John Ashburner and Karl J. Friston. Voxel-Based Morphometry–The methods. *NeuroImage*, 11(6):805–821, June 2000.
- [13] F. Gregory Ashby. *Statistical Analysis of fMRI Data*. The MIT Press, 1 edition, March 2011.
- [14] Alzheimer’s Association. 2011 Alzheimer’s disease facts and figures. *Alzheimer’s and Dementia*, 7(2):208–244, 2011.
- [15] American Psychiatric Association. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders (Diagnostic & Statistical Manual of Mental Disorders)*. American Psychiatric Press Inc., 4th text revision edition, July 2000.
- [16] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with Cross-Correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, February 2008. PMID: 17659998 PMCID: 2276735.
- [17] Borja Ayerdi and Manuel Graña. Hybrid extreme rotation forest. *Neural Networks*, Submitted, 2013.
- [18] Borja Ayerdi, Alexandre Savio, and Manuel Graña. Meta-ensembles of classifiers for Alzheimer’s disease detection using independent roi features. In José Manuel Ferrández Vicente, José Ramón Álvarez Sánchez, Félix Paz López, and Fco.Javier Toledo Moreo, editors, *Natural and Artificial Computation in Engineering and Medical Applications*, volume 7931 of *Lecture Notes in Computer Science*, pages 122–130. Springer Berlin Heidelberg, 2013.
- [19] Daniella B. Parente, Emerson L. Gasparetto, Luiz Celso Hygino da Cruz, et al. Potential role of diffusion tensor MRI in the differential diagnosis of mild cognitive impairment and alzheimer’s disease. *American Journal of Roentgenology*, 190(5):1369–1374, May 2008.

- [20] G. Bachmann, M. S. Damian, M. Koch, G. Schilling, B. Fach, and S. Stöppler. The clinical and genetic correlates of MRI findings in myotonic dystrophy. *Neuroradiology*, 38(7):629–635, October 1996.
- [21] Iñigo Barandiaran, Celine Paloc, and Manuel Graña. Real-time optical markerless tracking for augmented reality applications. *Journal of Real-Time Image Processing*, 5:129–138, 2010.
- [22] J. C. Baron, G. Chetelat, 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. *Neuroimage*, 14(2):298–309, 2001.
- [23] P.E. Barta, G.D. Pearlson, L.B. Brill, R. Royall, I.K. McGilchrist, A.E. Pulver, R.E. Powers, M.F. Casanova, A.Y. Tien, S. Frangou, and R.G. Petty. Planum temporale asymmetry reversal in schizophrenia: replication and relationship to gray matter abnormalities. *The American Journal of Psychiatry*, 154(5):661–667, May 1997. PMID: 9137122.
- [24] P. J. Basser, J. Mattiello, and D. LeBihan. Estimation of the effective Self-Diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B*, 103(3):247–254, March 1994.
- [25] P.J. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, January 1994. PMID: 8130344.
- [26] P. G. Batchelor, M. Moakher, D. Atkinson, F. Calamante, and A. Connelly. A rigorous framework for diffusion tensor calculus. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 53(1):221–225, January 2005. PMID: 15690523.
- [27] T.E.J. Behrens, M.W. Woolrich, M. Jenkinson, H. Johansen-Berg, R.G. Nunes, S. Clare, P.M. Matthews, J.M. Brady, and S.M. Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 50(5):1077–1088, November 2003. PMID: 14587019.
- [28] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T. Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101, August 2007. PMID: 17560126 PMCID: PMC2214855.
- [29] A. Bhattacharyya. On a measure of divergence between two multinomial populations. *Sankhya: The Indian Journal of Statistics (1933-1960)*, 7(4):401–406, July 1946. Copyright 1946 Indian Statistical Institute.

- [30] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407, November 1986. PMID: 3763909.
- [31] R.M. Bonelli and J.L. Cummings. Frontal-subcortical circuitry and behavior. *Dialogues Clin. Neurosci.*, 9:141–151, 2007.
- [32] Matias Bossa, Ernesto Zacur, and Salvador Olmos. Tensor-based morphometry with stationary velocity field diffeomorphic registration: Application to ADNI. *NeuroImage*, 51(3):956–969, July 2010.
- [33] G. Bouchard, R. Roy, M. Declos, J. Mathieu, and K. Kouladjian. Origin and diffusion of the myotonic dystrophy gene in the saguenay region (Quebec). *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, 16(1):119–122, 1989.
- [34] A. Breier, R.W. Buchanan, A. Elkashef, R.C. Munson, B. Kirkpatrick, and F. Gellad. Brain morphology and schizophrenia: A magnetic resonance imaging study of limbic, prefrontal cortex, and caudate structures. *Archives of General Psychiatry*, 49(12):921–926, December 1992.
- [35] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.
- [36] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [37] J.D. Brook, M.E. McCurrach, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J.P. Thirion, and T. Hudson. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, 68(4):799–808, February 1992.
- [38] D.S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.
- [39] C. Bungener, R. Jouvent, and C. Delaporte. Psychopathological and emotional deficits in myotonic dystrophy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 65(3):353–356, 1998.
- [40] Christopher Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):167–121, 1998.
- [41] G. F. Busatto, G. E. J. Garrido, O. P. Almeida, C. C. Castro, C. H. P. Camargo, C. G. Cid, C. A. Buchpiguel, S. Furuie, and C. M. Bottino. A voxel-based morphometry study of temporal lobe gray matter reductions in Alzheimer's disease. *Neurobiology of Aging*, 24(2):221–231, 2003.

- [42] Geraldo Busatto. Expert review of neurotherapeutics; Voxel-Based Morphometry in Alzheimers disease. 8(11):1691–1702, 2008.
- [43] Jiuwen Cao, Zhiping Lin, Guang-Bin Huang, and Nan Liu. Voting based extreme learning machine. *Information Sciences*, 185(1):66 – 77, 2012.
- [44] A. Caprihan, G.D. Pearlson, and V.D. Calhoun. Application of principal component analysis to distinguish patients with Schizophrenia from healthy controls based on fractional anisotropy measurements. *NeuroImage*, 42(2):675–682, August 2008.
- [45] B. Censori, L. Provinciali, M. Danni, L. Chiaramoni, M. Maricotti, N. Foschi, M. Del Pesce, and U. Salvolini. Brain involvement in myotonic dystrophy: MRI features and their relationship to clinical and cognitive conditions. *Acta Neurologica Scandinavica*, 90(3):211–217, September 1994. PMID: 7847063.
- [46] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [47] R. Chaves, J. Ramírez, J.M. Górriz, M. López, D. Salas-Gonzalez, I. Álvarez, and F. Segovia. SVM-based computer-aided diagnosis of the Alzheimer’s disease using t-test NMSE feature selection with feature correlation weighting. *Neuroscience Letters*, 461(3):293–297, September 2009.
- [48] R. Chaves, J. Ramírez, J.M. Górriz, and C.G. Puntonet. Association rule-based feature selection method for Alzheimer’s disease diagnosis. *Expert Systems with Applications*, 39(14):11766–11774, October 2012.
- [49] S. Chen, C.F.N. Cowan, and P.M. Grant. Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2):302–309, 1991.
- [50] T. Chen, A. Rangarajan, et al. CAVIAR: Classification via aggregated regression and its application in classifying OASIS brain database. *Proceedings / IEEE International Symposium on Biomedical Imaging: from nano to macro.*, 2010:1337–1340, April 2010.
- [51] Darya Chyzhyk. Bootstrapped dendritic classifiers for Alzheimer’s disease classification on MRI features. In Manuel Graña, Carlos Toro, Jorge Posada, Robert J. Howlett, and Lakhmi C. Jain, editors, *KES*, volume 243 of *Frontiers in Artificial Intelligence and Applications*, pages 2251–2258. IOS Press, 2012.
- [52] Darya Chyzhyk, Alexandre Savio, and Manuel Graña. Hybrid dendritic computing with kernel-LICA applied to Alzheimer’s disease detection in MRI. *Neurocomputing*, 75(1):72–77, January 2012.

- [53] G. Chételat, B. Desgranges, et al. Mapping gray matter loss with voxel-based morphometry in mild cognitive impairment. *Neuroreport*, 13(15):1939–1943, October 2002. PMID: 12395096.
- [54] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge Academic, 2 edition, January 1988.
- [55] A. Convit, J. de Asis, et al. Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer’s disease. *Neurobiology of Aging*, 21(1):19–26, February 2000. PMID: 10794844.
- [56] Robert W. Cox. AFNI: what a long strange trip it’s been. *NeuroImage*, 62(2):743–747, August 2012.
- [57] R. Cuingnet, E. Gerardin, et al. Automatic classification of patients with Alzheimer’s disease from structural MRI: a comparison of ten methods using the ADNI database. *NeuroImage*, 56(2):766–781, May 2011.
- [58] M. S. Damian, G. Schilling, G. Bachmann, C. Simon, S. Stöppler, and W. Dorndorf. White matter lesions and cognitive deficits: relevance of lesion pattern? *Acta Neurologica Scandinavica*, 90(6):430–436, 1994.
- [59] C. Davatzikos, Y. Fan, et al. Detection of prodromal Alzheimer’s disease via pattern classification of MRI. *Neurobiology of aging*, 29(4):514–523, April 2008. PMID: 17174012 PMCID: 2323584.
- [60] A. Lopez de Munain, A. Blanco, J. I. Emparanza, J. J. Poza, J. F. Marti Masso, A. Cobo, L. Martorell, M. Baiget, and J. M. Martinez Lage. Prevalence of myotonic dystrophy in Guipuzcoa (Basque Country, Spain). *Neurology*, 43(8):1573–1576, August 1993.
- [61] B.C. Dickerson, I. Goncharova, M.P. Sullivan, C. Forchetti, R.S. Wilson, D.A. Bennett, L.A. Beckett, and L. deToledo Morrell. MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer’s disease. *Neurobiology of Aging*, 22(5):747 – 754, 2001.
- [62] Bradford C. Dickerson and Reisa A. Sperling. Neuroimaging biomarkers for clinical trials of Disease-Modifying therapies in Alzheimer’s disease. *NeuroRx*, 2(2):348–360, April 2005. PMID: 15897955 PMCID: 1064996.
- [63] G. Douaud, S. Jbabdi, T. E. J. Behrens, R. A. Menke, A. Gass, A. U. Monsch, A. Rao, B. Whitcher, G. Kindlmann, P. M. Matthews, and S. Smith. DTI measures in crossing-fibre areas: Increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer’s disease. *NeuroImage*, 55(3):880–890, April 2011.

- [64] A.E. Emery. Population frequencies of inherited neuromuscular diseases - a world survey. *Neuromuscular Disorders: NMD*, 1(1):19–29, 1991.
- [65] Eric Jones and Travis Oliphant and Pearu Peterson and others. SciPy: Open source scientific tools for Python.
- [66] Y. Fan, , S.M. Resnick, S. Wu, and C. Davatzikos. Structural and functional biomarkers of prodromal Alzheimer’s disease: a high-dimensional pattern classification study. *NeuroImage*, 41(2):277–285, 2008.
- [67] Y. Fan, D. Shen, et al. Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 1):1–8, 2005.
- [68] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos. COMPARE: classification of morphological patterns using adaptive regional elements. *IEEE Transactions on Medical Imaging*, 26(1):93–105, 2007.
- [69] Yong Fan, Dinggang Shen, and Christos Davatzikos. Classification of structural images via High-Dimensional image warping, robust feature extraction, and SVM. In *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2005*, pages 1–8. 2005.
- [70] D. Faraggi and B. Reiser. Estimation of the area under the ROC curve. *Statistics in Medicine*, 21(20):3093–3106, October 2002.
- [71] Tomer Fekete, Meytal Wilf, Denis Rubin, Shimon Edelman, Rafael Malach, and Lilianne R. Mujica-Parodi. Combining classification with fMRI-Derived complex network measures for potential neurodiagnostics. *PLoS ONE*, 8(5):e62867, May 2013.
- [72] Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, August 2012.
- [73] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327, January 2011.
- [74] A.F. Fotenos, A.Z. Snyder, L.E. Girton, J.C. Morris, and R.L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64(6):1032–1039, March 2005.
- [75] Michael D. Fox, Dongyang Zhang, Abraham Z. Snyder, and Marcus E. Raichle. The global signal and observed anticorrelated resting state brain networks. *Journal of Neurophysiology*, 101(6):3270–3283, June 2009. PMID: 19339462.

- [76] M. Frederikse, A. Lu, E. Aylward, P. Barta, T. Sharma, and G. Pearlson. Sex differences in inferior parietal lobule volume in schizophrenia. *The American Journal of Psychiatry*, 157(3):422–427, March 2000.
- [77] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, pages 37, 23, 1995.
- [78] G.B. Frisoni, C. Testa, A. Zorzan, F. Sabattoli, A. Beltramello, H. Soininen, and M.P. Laakso. Detection of grey matter loss in mild Alzheimer’s disease with Voxel Based Morphometry. *J. Neurol. Neurosurg. Psychiatry*, 73(6):657–664, December 2002.
- [79] K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny. *Statistical Parametric Mapping The Analysis of Functional Brain Images*. Elsevier, 2007.
- [80] K.J. Friston, S. Williams, R. Howard, R.S. Frackowiak, and R. Turner. Movement-related effects in fMRI time-series. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 35(3):346–355, March 1996. PMID: 8699946.
- [81] G. Fung and J. Stoeckel. SVM feature selection for classification of SPECT images of Alzheimer’s disease using spatial information. *Knowl. Inf. Syst.*, 11(2):243–258, 2007.
- [82] G. G. Antonini, F. Soscia, F. Giubilei, A. De Carolis, F. Gragnani, S. Morino, A. Ruberto, and R. Tatarelli. Health-related quality of life in myotonic dystrophy type 1 and its relationship with cognitive and emotional functioning. *Journal of Rehabilitation Medicine*, 38(3):181–185, 2006.
- [83] M.T. García-Sebastián, A. Savio, M. Graña, and J. Villanúa. On the use of morphometry based features for Alzheimer’s disease detection on MRI. In Alberto Prieto Juan M. Corchado (Editors) Joan Cabestany, Francisco Sandoval, editor, *Bio-Inspired Systems: Computational and Ambient Intelligence. / IWANN 2009 (Part I)*, volume 5517 of *LNCS*, pages 957–964. Springer-Verlag, 2009.
- [84] E. Gerardin, G. Chételat, et al. Multidimensional classification of hippocampal shape features discriminates alzheimer’s disease and mild cognitive impairment from normal aging. *NeuroImage*, 47(4):1476–1486, October 2009.
- [85] Antonio Giorgio, Maria T. Dotti, Marco Battaglini, Silvia Marino, et al. Cortical damage in brains of patients with adult-form of myotonic dystrophy type 1 and no or minimal MRI abnormalities. *Journal of Neurology*, 253(11):1471–1477, November 2006.
- [86] D.E. Goldberg. *Genetic Algorithms in Search Optimization and Machine Learning*. Addison Wesley, 1989.

- [87] J.M. Goldstein, J.M. Goodman, L.J. Seidman, D.N. Kennedy, N. Makris, H. Lee, J. Tourville, V.S. Caviness, S.V. Faraone, and M.T. Tsuang. Cortical abnormalities in schizophrenia identified by structural magnetic resonance imaging. *Archives of General Psychiatry*, 56(6):537–547, June 1999.
- [88] 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. *Neuroscience Letters*, 460(2):108–111, August 2009.
- [89] J.A. Grahn, J.A. Parkinson, and A.M. Owen. The cognitive functions of the caudate nucleus. *Prog Neurobiol.*, 86:141–155, 2008.
- [90] M. Graña, M. Termenon, A. Savio, A. Gonzalez-Pinto, J. Echeveste, J.M. Pérez, and A. Besga. Computer aided diagnosis system for alzheimer’s disease using brain diffusion tensor imaging features selected by pearson’s correlation. *Neuroscience Letters*, 502(3):225–229, September 2011.
- [91] X. Guo, Z. Wang, K. Li, Z. Li, Z. Qi, Z. Jin, L. Yao, and K. Chen. Voxel-based assessment of gray and white matter volumes in alzheimer’s disease. *Neuroscience Letters*, 468(2):146–150, January 2010.
- [92] R.E. Gur, P.E. Cowell, A. Latshaw, B.I. Turetsky, R.I. Grossman, S.E. Arnold, W.B. Bilker, and R.C. Gur. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. *Archives of General Psychiatry*, 57(8):761–768, August 2000.
- [93] M.T. Hagan, H.B. Demuth, and M.H. Beale. *Neural Network Design*. PWS Pub. Co., Har/Dsk edition, December 1995.
- [94] Yaroslav O. Halchenko and Michael Hanke. Open is not enough. let’s take the next step: an integrated, community-driven computing platform for neuroscience. *Frontiers in Neuroinformatics*, page 22, 2012.
- [95] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, April 1982.
- [96] H. G. Harley, Rundle S. A., J. C. Myring, and J. Brook. Size of the unstable ctg repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am.J.Hum.Genet.*, 52:1164–1174, 1993.
- [97] T. Hashimoto, M. Tayama, M. Miyazaki, K. Murakawa, H. Kawai, H. Nishitani, and Y. Kuroda. Neuroimaging study of myotonic dystrophy. II. MRI measurements of the brain. *Brain and Development*, 17(1):28–32, 1995.
- [98] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition, July 1998.

- [99] T.K. Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8):832–844, 1998.
- [100] D.P. Holinger, M.E. Shenton, C.G. Wible, R. Donnino, R. Kikinis, F.A. Jolesz, and R.W. McCarley. Superior temporal gyrus volume abnormalities and thought disorder in Left-Handed schizophrenic men. *The American Journal of Psychiatry*, 156(11):1730–1735, November 1999.
- [101] D. Horínek, A. Varjassyová, and J. Hort. Magnetic resonance analysis of amygdalar volume in alzheimer’s disease. *Current Opinion in Psychiatry*, 20(3):273–277, 2007.
- [102] Chengzhong Huang, Bin Yan, Hua Jiang, and Dahui Wang. Combining voxel-based morphometry with artificial neural network theory in the application research of diagnosing Alzheimer’s disease. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 1, pages 250–254, 2008.
- [103] G.-B. Huang, L. Chen, and C.-K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks*, 17(4):879–892, 2006.
- [104] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3):489 – 501, 2006.
- [105] A. Hunter, C. Tsilfidis, G. Mettler, P. Jacob, M. Mahadevan, L. Surh, and R. Korneluk. The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *Journal of Medical Genetics*, 29(11):774–779, 1992.
- [106] I.A. Illan, J.M. Gorriz, J. Ramirez, D. Salas-Gonzalez, M.M. Lopez, F. Segovia, R. Chaves, M. Gomez-Rio, and C.G. Puntonet. 18f-fdg pet imaging analysis for computer aided Alzheimer’s diagnosis. *Information Sciences*, 181(4):903–916, 2011.
- [107] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, October 2002.
- [108] Mark Jenkinson, Christian F. Beckmann, Timothy E.J. Behrens, Mark W. Woolrich, and Stephen M. Smith. FSL. *NeuroImage*, 62(2):782–790, August 2012.
- [109] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, June 2001.
- [110] T. Joachims. A support vector method for multivariate performance measures. *Proceedings of the 22nd International Conference on Machine Learning*, pages 377—384, 2005.

- [111] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, February 1967.
- [112] J. A. Kaye, T. Swihart, et al. Volume loss of the hippocampus and temporal lobe in healthy elderly persons destined to develop dementia. *Neurology*, 48(5):1297–1304, May 1997. PMID: 9153461.
- [113] the late Sir Maurice Kendall and Jean Dickinson Gibbons. *Rank Correlation Methods*. Oxford University Press, USA, 5 edition, September 1990.
- [114] R. J. Killiany, T. Gomez-Isla, et al. Use of structural magnetic resonance imaging to predict who will get Alzheimer’s disease. *Annals of Neurology*, 47(4):430–439, April 2000. PMID: 10762153.
- [115] Arno Klein, Jesper Andersson, Babak A. Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E. Christensen, D. Louis Collins, James Gee, Pierre Hellier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P. Woods, J. John Mann, and Ramin V. Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802, 2009.
- [116] S. Kloppel. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131(3):681–689, 2008.
- [117] S. Kloppel, C. M. Stonnington, et al. Automatic classification of MR scans in Alzheimer’s disease. *Brain*, 131(3):681–689, March 2008.
- [118] T. Kohonen. *Self-organization and associative memory: 3rd edition*. Springer-Verlag New York, Inc., 1989.
- [119] T. Kohonen. Learning vector quantization. In *The handbook of brain theory and neural networks*, pages 537–540. MIT Press, 1998.
- [120] N. Kriegeskorte, W. K. Simmons, et al. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci*, 12(5):535–540, May 2009.
- [121] M. Kubicki, R. McCarley, C.-F. Westin, H-J Park, S. Maier, R. Kikinis, F.A. Jolesz, and M.E. Shenton. A review of diffusion tensor imaging studies in schizophrenia. *Journal of psychiatric research*, 41(1-2):15–30, 2007. PMID: 16023676 PMCID: 2768134.
- [122] M. Kubicki, H. Park, C.-F. Westin, P.G. Nestor, R.V. Mulkern, S.E. Maier, M. Niznikiewicz, E.E. Connor, J.J. Levitt, M. Frumin, R. Kikinis, F.A. Jolesz, R.W. McCarley, and M.E. Shenton. DTI and MTR abnormalities in schizophrenia: Analysis of white matter integrity. *NeuroImage*, 26(4):1109–1118, July 2005. PMID: 15878290 PMCID: 2768051.

- [123] M. Kubicki, C.F. Westin, R.W. McCarley, and M.E. Shenton. The application of DTI to investigate white matter abnormalities in schizophrenia. *Annals of the New York Academy of Sciences*, 1064(1):134–148, 2005.
- [124] M. Kyriakopoulos, T. Bargiotas, G.J. Barker, and S. Frangou. Diffusion tensor imaging in schizophrenia. *European Psychiatry*, 23(4):255–273, June 2008.
- [125] L. Laberge, P. Bégin, J. Montplaisir, and J. Mathieu. Sleep complaints in patients with myotonic dystrophy. *Journal of Sleep Research*, 13(1):95–100, 2004.
- [126] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, January 2004.
- [127] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, January 2004.
- [128] N. Lepore, C. Brun, et al. Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. *IEEE Transactions on Medical Imaging*, 27(1):129–141, January 2008. PMID: 18270068.
- [129] Xuchun Li, Lei Wang, and E. Sung. A study of AdaBoost with SVM based weak learners. In *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, volume 1, pages 196–201 vol. 1, 2005.
- [130] Dongqiang Liu, Chaogan Yan, Juejing Ren, Li Yao, Vesa J. Kiviniemi, and Yufeng Zang. Using coherence to measure regional homogeneity of resting-state fMRI signal. *Frontiers in Systems Neuroscience*, 4:24, 2010.
- [131] M. Liu, D. Zhang, et al. Ensemble sparse classification of alzheimer’s disease. *NeuroImage*, 60(2):1106–1116, April 2012. PMID: 22270352.
- [132] Y. Liu, L. Teverovskiy, et al. Discriminative mr image feature analysis for automatic schizophrenia and alzheimer’s disease classification. *Med Image Comput Comput Assist Interv. MICCAI 2004*, 7(Pt 1):393–400, September 2004.
- [133] M.M. López, J. Ramírez, J.M. Górriz, I. Álvarez, D. Salas-Gonzalez, F. Segovia, and R. Chaves. SVM-based CAD system for early detection of the Alzheimer’s disease using kernel PCA and LDA. *Neuroscience Letters*, 464(3):233–238, October 2009.
- [134] O. Luaces, F. Taboada, et al. Predicting the probability of survival in intensive care unit patients from a small number of variables and training examples. *Artificial Intelligence in Medicine*, 45(1):63–76, January 2009. PMID: 19185475.

- [135] S.K. Madsen, A.J. Ho, X. Hua, P.S. Saharan, A.W. Toga, C.R. Jack Jr, M.W. Weiner, and P.M. Thompson. 3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects. *Neurobiology of Aging*, 31(8):1312–1325, August 2010.
- [136] P. Malloy, S. K. Mishra, and S. H. Adler. Neuropsychological deficits in myotonic muscular dystrophy. *Journal of Neurology, Neurosurgery, and Psychiatry*, 53(11):1011–1013, 1990.
- [137] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, September 2007. PMID: 17714011.
- [138] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neuroscience*, 19(9):1498–1507, 2007.
- [139] J.S. Maritz. *Distribution-Free Statistical Methods, Second Edition*. Chapman and Hall/CRC, 2 edition, April 1995.
- [140] I. Marques and M. Graña. Face recognition with lattice independent component analysis and extreme learning machines. *Soft Computing*, on line, February 2012.
- [141] F. Martinello, A. Piazza, E. Pastorello, C. Angelini, and C.P. Trevisan. Clinical and neuroimaging study of central nervous system in congenital myotonic dystrophy. *Journal of Neurology*, 246(3):186–192, March 1999. PMID: 10323316.
- [142] L. Martorell, G Monckton D., and Gamez J. Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Human Molecular Genetics*, 7:307–312, 1998.
- [143] S. J. Mason and N. E. Graham. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society*, 128(584):2145–2166, 2002.
- [144] P.K. McGuire and C.D. Frith. Disordered functional connectivity in schizophrenia. *Psychological Medicine*, 26(4):663–667, July 1996. PMID: 8817700.
- [145] D. Medina, L. DeToledo-Morrell, et al. White matter changes in mild cognitive impairment and AD: a diffusion tensor imaging study. *Neurobiology of Aging*, 27(5):663–672, May 2006. PMID: 16005548.

- [146] G. Meola, V. Sansone, D. Perani, S. Scarone, S. Cappa, C. Dragoni, E. Cattaneo, M. Cotelli, C. Gobbo, F. Fazio, et al. Executive dysfunction and avoidant personality trait in myotonic dystrophy type 1 (DM-1) and in proximal myotonic myopathy (PROMM/DM-2). *Neuromuscular Disorders*, 13(10):813–821, 2003.
- [147] F.A. Middleton and P.L. Strick. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain and Cognition*, 42(2):183–200, March 2000.
- [148] M. L. Mostacciuolo, G. Barbujani, M. Armani, G. A. Danieli, C. Angelini, and D. C. Rao. Genetic epidemiology of myotonic dystrophy. *Genetic Epidemiology*, 4(4):289–298, 1987.
- [149] Susanne G. Mueller, Michael W. Weiner, et al. Ways toward an early diagnosis in Alzheimer’s disease: the Alzheimer’s disease neuroimaging initiative (ADNI). *Alzheimer’s & dementia: the journal of the Alzheimer’s Association*, 1(1):55–66, July 2005. PMID: 17476317.
- [150] P. T. Nelson, H. Braak, and W. R. Markesbery. Neuropathology and cognitive impairment in Alzheimer’s disease: a complex but coherent relationship. *Journal of Neuropathology and Experimental Neurology*, 68(1):1–14, January 2009. PMID: 19104448.
- [151] J. NEYMAN and E. S. PEARSON. ON THE USE AND INTERPRETATION OF CERTAIN TEST CRITERIA FOR PURPOSES OF STATISTICAL INFERENCE PART i. *Biometrika*, 20A(1-2):175–240, December 1928.
- [152] F. Nielsen and S. Boltz. The burbea-rao and bhattacharyya centroids. *IEEE Transactions on Information Theory*, 57(8):5455–5466, August 2011.
- [153] M. Niznikiewicz, R. Donnino, R.W. McCarley, P.G. Nestor, D.V. Iosifescu, B.O’Donnell, J. Levitt, and M.E. Shenton. Abnormal angular gyrus asymmetry in schizophrenia. *The American Journal of Psychiatry*, 157(3):428–437, March 2000.
- [154] S. Ono, K. Inoue, T. Mannen, F. Kanda, K. Jinnai, and K. Takahashi. Neuropathological changes of the brain in myotonic dystrophy - some new observations. *Journal of the Neurological Sciences*, 81(2-3):301–320, November 1987.
- [155] M. Ota, N. Sato, Y. Ohya, Y. Aoki, K. Mizukami, T. Mori, and T. Asada. Relationship between diffusion tensor imaging and brain morphology in patients with myotonic dystrophy. *Neuroscience Letters*, 407(3):234–239, October 2006.
- [156] Reiko Oyamada, Masaharu Hayashi, Yuji Katoh, Kuniaki Tsuchiya, Toshio Mizutani, Itaru Tominaga, and Haruo Kashima. Neurofibrillary tangles and deposition of oxidative

- products in the brain in cases of myotonic dystrophy. *Neuropathology*, 26:107–114, April 2006.
- [157] P. Padilla, J.M. Górriz, J. Ramírez, E.W. Lang, R. Chaves, F. Segovia, M. López, D. Salas-González, and I. Álvarez. Analysis of SPECT brain images for the diagnosis of Alzheimer’s disease based on NMF for feature extraction. *Neuroscience Letters*, 479(3):192–196, August 2010.
 - [158] F. Pedregosa, G. Varoquaux, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011.
 - [159] M. F. Phillips, H. M. Steer, J. R. Soldan, C. M. Wiles, and P. S. Harper. Daytime somnolence in myotonic dystrophy. *Journal of neurology*, 246(4):275–282, 1999.
 - [160] C. Pierpaoli, P.Jezzard, P.J. Basser, A. Barnett, and G. Di Chiro. Diffusion tensor MR imaging of the human brain. *Radiology*, 201(3):637–648, December 1996. PMID: 8939209.
 - [161] C. Plant, S. J. Teipel, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of disease. *NeuroImage*, 50(1):162–174, March 2010.
 - [162] Jonathan D. Power, Kelly A. Barnes, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3):2142–2154, February 2012.
 - [163] Christophe Pradal, Gaël Varoquaux, and Hans Peter Langtangen. Publishing scientific software matters. *Journal of Computational Science*, 4(5):311–312, September 2013.
 - [164] J. Ramirez, J.M. Gorriz, M. Lopez, D. Salas-Gonzalez, I. Alvarez, F. Segovia, and C.G. Puntonet. Early detection of the Alzheimer’s disease combining feature selection and kernel machines. In *15th International Conference on Neural Information Processing of the Asia-Pacific Neural Network Assembly (ICONIP 2008)*, 2008.
 - [165] J. Ramirez, J.M. Gorriz, D. Salas-Gonzalez, A. Romero, M. Lopez, I. Alvarez, and M. Gomez-Rio. Computer-aided diagnosis of Alzheimer’s type dementia combining support vector machines and discriminant set of features. *Information Sciences*, In Press, Corrected Proof:–, 2009.
 - [166] J. Ramírez, J.M. Górriz, F. Segovia, R. Chaves, D. Salas-Gonzalez, M. López, I. Álvarez, and P. Padilla. Computer aided diagnosis system for the Alzheimer’s disease based on partial least squares and random forest SPECT image classification. *Neuroscience Letters*, 472(2):99–103, March 2010.

- [167] J. S. Rubinsztein, D. C. Rubinsztein, S. Goodburn, and A. J. Holland. Apathy and hypersomnia are common features of myotonic dystrophy. *Journal of neurology, neurosurgery, and psychiatry*, 64(4):510–515, 1998.
- [168] J. S. Rubinsztein, D. C. Rubinsztein, P. J. McKenna, S. Goodburn, and A. J. Holland. Mild myotonic dystrophy is associated with memory impairment in the context of normal general intelligence. *Journal of medical genetics*, 34(3):229–233, 1997.
- [169] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning internal representations by error propagation*, pages 318–362. MIT Press, 1986.
- [170] Ziad S Saad, Stephen J Gotts, Kevin Murphy, Gang Chen, Hang Joon Jo, Alex Martin, and Robert W Cox. Trouble at rest: how correlation patterns and group differences become distorted after global signal regression. *Brain connectivity*, 2(1):25–32, 2012. PMID: 22432927.
- [171] D. Salas-Gonzalez, J. M. Górriz, J. Ramírez, M. López, I. A. Illan, F. Segovia, C. G. Puntonet, and M. Gómez-Río. Analysis of SPECT brain images for the diagnosis of Alzheimer’s disease using moments and support vector machines. *Neuroscience Letters*, 461(1):60–64, September 2009.
- [172] Theodore D. Satterthwaite, Daniel H. Wolf, James Loughhead, Kosha Ruparel, Mark A. Elliott, Hakon Hakonarson, Ruben C. Gur, and Raquel E. Gur. Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*, 60(1):623–632, March 2012.
- [173] A. Savio, M.T. García-Sebastián, D. Chyzyk, C. Hernández, M. Graña, A. Sistiaga, A. López de Munain, and J. Villanúa. Neurocognitive disorder detection based on feature vectors extracted from VBM analysis of structural MRI. *Computers in Biology and Medicine*, 41(8):600–610, August 2011. PMID: 21621760.
- [174] A. Savio, M.T. García-Sebastián, C. Hernandez, M. Graña, and J. Villanúa. Classification results of artificial neural networks for Alzheimer’s disease detection. In Hujun Yin (eds) Emilio Corchado, editor, *Intelligent Data Engineering and Automated Learning- IDEAL 2009*, volume 5788 of *LNCS*, pages 641–648. Springer-Verlag, 2009.
- [175] A. Savio and M. Graña. Deformation based feature selection for computer aided diagnosis of Alzheimer’s disease. *Expert Systems with Applications*, 40(5):1619–1628, April 2013.
- [176] A. Savio and M. Graña. An ensemble of classifiers guided by the AAL brain atlas for Alzheimer’s disease detection. In Ignacio Rojas, Gonzalo Joya, and Joan Cabestany, editors, *Advances in Computational Intelligence*, number 7903 in *Lecture Notes in Computer Science*, pages 107–114. Springer Berlin Heidelberg, January 2013.

- [177] Alexandre Savio, Juliette Charpentier, Maite Termenón, Ann K. Shinn, and Manuel Graña. Neural classifiers for schizophrenia diagnostic support on diffusion imaging data. *Neural Network World*, 20:935–949, January 2010.
- [178] Alexandre Savio, Manuel Graña, and Jorge Villanúa. Deformation based features for Alzheimer’s disease detection with linear SVM. In Emilio Corchado, Marek Kurzynski, and Michal Wozniak, editors, *Hybrid Artificial Intelligent Systems*, number 6679 in Lecture Notes in Computer Science, pages 336–343. Springer Berlin Heidelberg, January 2011.
- [179] R. I. Scahill, J. M. Schott, J. M. Stevens, M. N. Rossor, and N. C. Fox. Mapping the evolution of regional atrophy in Alzheimer’s disease: Unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences*, 99(7):4703, 2002.
- [180] R.E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37(3):297–336, December 1999.
- [181] T.E. Schlaepfer, G.J. Harris, A.Y Tien, L.W. Peng, S. Lee, E.B. Federman, G.A. Chase, P.E. Barta, and G.D. Pearlson. Decreased regional cortical gray matter volume in schizophrenia. *The American Journal of Psychiatry*, 151(6):842–848, June 1994.
- [182] F. Segovia, J.M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, M. López, R. Chaves, and P. Padilla. Classification of functional brain images using a GMM-based multi-variate approach. *Neuroscience Letters*, 474(1):58–62, April 2010.
- [183] N. Sergeant, B. Sablonnière, S. Schraen-Maschke, A. Ghestem, et al. Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Human Molecular Genetics*, 10(19):2143–2155, September 2001.
- [184] Laura Serra, Mara Cercignani, Delia Lenzi, Roberta Perri, Lucia Fadda, Carlo Caltagirone, Emiliano Macaluso, and Marco Bozzali. Grey and white matter changes at different stages of Alzheimer’s disease. *Journal of Alzheimer’s Disease: JAD*, September 2009. PMID: 19749411.
- [185] M.E. Shenton, C.C. Dickey, M. Frumin, and R.W. McCarley. A review of MRI findings in schizophrenia. *Schizophrenia research*, 49(1-2):1–52, April 2001. PMID: 11343862 PMCID: 2812015.
- [186] M.E. Shenton, R. Kikinis, F.A. Jolesz, S.D. Pollak, M. LeMay, C.G. Wible, H. Hokama, J. Martin, D. Metcalf, and M. Coleman. Abnormalities of the left temporal lobe and thought disorder in schizophrenia. a quantitative magnetic resonance imaging study. *The New England journal of medicine*, 327(9):604–612, August 1992. PMID: 1640954.

- [187] A. Sistiaga, I. Urreta, M. Jodar, A. M. Cobo, J. Emparanza, D. Otaegui, J. J. Poza, J. J. Merino, H. Imaz, J. F. Marti-Masso, and A. Lopez de Munain. Cognitive/Personality pattern and triplet expansion size in adult myotonic dystrophy type 1 (DM1): CTG repeats, cognition and personality in DM1. *Psychological Medicine*, First View:1–9, 2009.
- [188] Stephen M. Smith, Mark Jenkinson, Mark W. Woolrich, Christian F. Beckmann, Timothy E. J. Behrens, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1:S208–219, 2004. PMID: 15501092.
- [189] P. Somervuo and T. Kohonen. Self-Organizing maps and learning vector quantization for feature sequences. *Neural Process. Lett.*, 10(2):151–159, 1999.
- [190] Donald F. Specht. Probabilistic neural networks. *Neural Netw.*, 3(1):109–118, 1990.
- [191] Robert Stahl, Olaf Dietrich, Stefan J. Teipel, Harald Hampel, Maximilian F. Reiser, and Stefan O. Schoenberg. White matter damage in Alzheimer’s disease and mild cognitive impairment: Assessment with Diffusion-Tensor MR imaging and parallel imaging techniques. *Radiology*, 243(2):483–492, May 2007.
- [192] David E. Stark, Daniel S. Margulies, Zarrar E. Shehzad, Philip Reiss, A. M. Clare Kelly, Lucina Q. Uddin, Dylan G. Gee, Amy K. Roy, Marie T. Banich, F. Xavier Castellanos, and Michael P. Milham. Regional variation in interhemispheric coordination of intrinsic hemodynamic fluctuations. *The Journal of Neuroscience*, 28(51):13754–13764, December 2008. PMID: 19091966.
- [193] P.R. Szeszko, R.M. Bilder, T. Lencz, S. Pollack, J.M. Alvir, M. Ashtari, H. Wu, and J.A. Lieberman. Investigation of frontal lobe subregions in first-episode schizophrenia. *Psychiatry Research*, 90(1):1–15, February 1999. PMID: 10320207.
- [194] Yoichiro Takayanagi, Tsutomu Takahashi, Lina Orikabe, Yuriko Mozue, Yasuhiro Kawasaki, Kazue Nakamura, Yoko Sato, Masanari Itokawa, Hidenori Yamasue, Kiyoto Kasai, Masayoshi Kurachi, Yuji Okazaki, and Michio Suzuki. Classification of first-episode schizophrenia patients and healthy subjects by automated MRI measures of regional brain volume and cortical thickness. *PLoS ONE*, 6(6):e21047, June 2011.
- [195] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. Thieme, January 1988.
- [196] Yan Tang, Lifeng Wang, Fang Cao, and Liwen Tan. Identify Schizophrenia using resting-state functional connectivity: an exploratory research and analysis. *BioMedical Engineering OnLine*, 11(1):50, August 2012. PMID: 22898249.
- [197] D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for Support Vector Machines-based relevance feedback in image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(7):1088–1099, 2006.

- [198] S. J. Teipel, C. Born, et al. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *NeuroImage*, 38(1):13–24, October 2007.
- [199] L. T. Timchenko, J. W. Miller, N. A. Timchenko, D. R. DeVore, K. V. Datar, L. Lin, R. Roberts, C. T. Caskey, and M. S. Swanson. Identification of a (CUG)_n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic acids research*, 24(22):4407–4414, 1996.
- [200] N. Tzourio-Mazoyer, B. Landeau, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, January 2002. PMID: 11771995.
- [201] Koene R.A. Van Dijk, Mert R. Sabuncu, and Randy L. Buckner. The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 59(1):431–438, January 2012.
- [202] V. Vapnik. *Statistical learning theory*. Wiley-Interscience, 1998.
- [203] V. N. Vapnik. *Statistical Learning Theory*. Wiley-Interscience, September 1998.
- [204] E. Varol, B. Gaonkar, et al. Feature ranking based nested support vector machine ensemble for medical image classification. In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 146–149, May 2012.
- [205] P. Vemuri, J.L. Gunter, M.L. Senjem, J.L. Whitwell, K. Kantarci, D.S. Knopman, B.F. Boeve, R.C. Petersen, and C.R. Jack. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–1197, February 2008.
- [206] P. Wang and R. Verma. On classifying Disease-Induced patterns in the brain using diffusion tensor images. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, pages 908–916. 2008.
- [207] B. L. Welch. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, January 1947. Copyright 1947 Biometrika Trust.
- [208] C. Wernicke. *Grundriss der Psychiatrie in klinischen Vorlesungen / von Carl Wernicke*. VDM Verlag Dr. Müller, Saarbrücken, 2007.
- [209] E. Westman, A. Simmons, et al. Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment and healthy controls. *NeuroImage*, 54(2):1178–1187, January 2011.

- [210] Eric Westman, Andrew Simmons, J-Sebastian Muehlboeck, et al. AddNeuroMed and ADNI: similar patterns of Alzheimer’s atrophy and automated MRI classification accuracy in Europe and North America. *NeuroImage*, 58(3):818–828, October 2011.
- [211] C.G. Wible, M.E. Shenton, H. Hokama, R. Kikinis, F.A. Jolesz, D. Metcalf, and R.W. McCarley. Prefrontal cortex and schizophrenia: A quantitative magnetic resonance imaging study. *Archives of General Psychiatry*, 52(4):279–288, April 1995.
- [212] B Wilson, H Balleny, K Patterson, and J Hodges. Myotonic dystrophy and progressive cognitive decline: A common condition or two separate problems? *Cortex*, 35(1):113–121, 1999.
- [213] S. Winblad, P. Hellstrom, C. Lindberg, and S. Hansen. Facial emotion recognition in myotonic dystrophy type 1 correlates with CTG repeat expansion. *Journal of neurology, neurosurgery, and psychiatry*, 77(2):219–223, 2006.
- [214] S. Winblad, C. Lindberg, and S. Hansen. Temperament and character in patients with classical myotonic dystrophy type 1 (DM-1). *Neuromuscular Disorders*, 15(4):287–292, 2005.
- [215] Wang Xi-zhao, Shao Qing-yan, Miao Qing, and Zhai Jun-hai. Architecture selection for networks trained with extreme learning machine using localized generalization error model. *Neurocomputing*, (0):–, 2012.
- [216] U. Yoon, J.-M. Lee, K. Im, Y.-W. Shin, B.H. Cho, I.Y. Kim, J.S. Kwon, and S.I. Kim. Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. *NeuroImage*, 34(4):1405–1415, February 2007.
- [217] Yang Yu, Hui Shen, Huiran Zhang, Ling-Li Zeng, Zhimin Xue, and Dewen Hu. Functional connectivity-based signatures of Schizophrenia revealed by multiclass pattern analysis of resting-state fMRI from schizophrenic patients and their healthy siblings. *BioMedical Engineering OnLine*, 12(1):10, February 2013. PMID: 23390976.
- [218] Yu-Feng Zang, Yong He, Chao-Zhe Zhu, Qing-Jiu Cao, Man-Qiu Sui, Meng Liang, Li-Xia Tian, Tian-Zi Jiang, and Yu-Feng Wang. Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain & development*, 29(2):83–91, March 2007. PMID: 16919409.
- [219] Yufeng Zang, Tianzi Jiang, Yingli Lu, Yong He, and Lixia Tian. Regional homogeneity approach to fMRI data analysis. *NeuroImage*, 22(1):394–400, May 2004.
- [220] J. Zhang, B. Yan, X. Huang, P. Yang, and C. Huang. The diagnosis of Alzheimer’s disease based on Voxel-Based Morphometry and Support Vector Machine. In *Proceedings of the*

- 2008 *Fourth International Conference on Natural Computation (ICNC)*, pages 197–201, Washington, DC, USA, 2008. IEEE Computer Society.
- [221] Yongyue Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.
- [222] Qi-Hong Zou, Chao-Zhe Zhu, Yihong Yang, Xi-Nian Zuo, Xiang-Yu Long, Qing-Jiu Cao, Yu-Feng Wang, and Yu-Feng Zang. An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF. *Journal of Neuroscience Methods*, 172(1):137–141, July 2008.
- [223] Xi-Nian Zuo, Adriana Di Martino, Clare Kelly, Zarrar E. Shehzad, Dylan G. Gee, Donald F. Klein, F. Xavier Castellanos, Bharat B. Biswal, and Michael P. Milham. The oscillating brain: Complex and reliable. *NeuroImage*, 49(2):1432–1445, January 2010. PMID: 19782143 PMCID: PMC2856476.
- [224] Xi-Nian Zuo, Clare Kelly, Adriana Di Martino, Maarten Mennes, Daniel S. Margulies, Saroja Bangaru, Rebecca Grzadzinski, Alan C. Evans, Yu-Feng Zang, F. Xavier Castellanos, and Michael P. Milham. Growing together and growing apart: Regional and sex differences in the lifespan developmental trajectories of functional homotopy. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(45):15034–15043, November 2010. PMID: 21068309 PMCID: PMC2997358.

Abstract

The application of Machine Learning algorithms to Neuroscience data has two main goals in this thesis. First, the construction of Computer Aided Diagnostic (CAD) systems to help alleviate the burden of increasing amounts of data for diagnosis. Second, the identification of image biomarkers corresponding to anatomical locations of the features selected for classification. This thesis is an empirical exploration of these ideas in the case of three neurological diseases. We have developed sound methodological frameworks, avoiding circularity effects in the validation process. For feature selection we have been working with supervised methods which produce good classification results and the ability to determine the spatial location of the features in the brain, among them an evolutionary wrapper selection method based on Extreme Learning Machines. The thesis covers experiments with a wide spectrum of classifiers for comparison. Finally, the approach has been tested on different MRI modalities showing its general applicability. A critical issue for multivariate modalities is the definition of appropriate scalar measures that may be useful for feature selection and extraction. The thesis tests the most appropriate for each modality: FA and MD for diffusion data, functional activity measures such as ReHo for functional data and measures from deformation maps resulting from non-linear registration of anatomical data. The results obtained show that the approach is useful for CAD systems in a variety of neurological diseases.

Alexandre Manhães Savio

Keywords: Medical Image Processing, Computer Aided Diagnostics, Image Biomarkers, Feature Selection, Alzheimer's Disease, Schizophrenia.

