



2017

Analyzing Heterogeneity In Neuroimaging With Probabilistic Multivariate Clustering Approaches

Aoyan Dong

University of Pennsylvania, dongaoyan@gmail.com

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Electrical and Electronics Commons](#)

Recommended Citation

Dong, Aoyan, "Analyzing Heterogeneity In Neuroimaging With Probabilistic Multivariate Clustering Approaches" (2017). *Publicly Accessible Penn Dissertations*. 3041.

<https://repository.upenn.edu/edissertations/3041>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/3041>

For more information, please contact repository@pobox.upenn.edu.

Analyzing Heterogeneity In Neuroimaging With Probabilistic Multivariate Clustering Approaches

Abstract

Automated quantitative neuroimaging analysis methods have been crucial in elucidating normal and pathological brain structure and function, and in building in vivo markers of disease and its progression. Commonly used methods can identify and precisely quantify

subtle and spatially complex imaging patterns of brain change associated with brain diseases. However, the overarching premise of these methods is that the disease group is a homogeneous entity resulting from a single, unifying pathophysiological process that has

a single imaging signature. This assumption ignores ample evidence for the heterogeneous nature of neurodegenerative diseases and neuropsychiatric disorders, resulting in incomplete or misleading descriptions. Accurate characterization of heterogeneity is important

for deepening our understanding of neurobiological processes, thus leading to improved disease diagnosis and prognosis.

In this thesis, we leveraged machine learning techniques to develop novel tools that can analyze the heterogeneity in both cross-sectional and longitudinal neuroimaging studies. Specifically, we developed a semi-supervised clustering method for characterizing

heterogeneity in cross-sectional group comparison studies, where normal and patient populations are modeled as high-dimensional point distributions, and heterogeneous disease effects are captured by estimating multiple transformations that align the two distributions, while accounting for the effect of nuisance covariates. Moreover, toward dissecting the heterogeneity in longitudinal cohorts, we proposed a method which simultaneously fits multiple population longitudinal multivariate trajectories and clusters subjects into subgroups. Longitudinal trajectories are modeled using spatiotemporally regularized cubic splines, while clustering is performed by assigning subjects to the subgroup whose population trajectory best fits their data.

The proposed tools were extensively validated using synthetic data. Importantly, they were applied to study the heterogeneity in large clinical neuroimaging cohorts. We identified four disease subtypes with distinct imaging signatures using data from Alzheimer's

Disease Neuroimaging Initiative, and revealed two subgroups with different longitudinal patterns using data from Baltimore Longitudinal Study on Aging. Critically, we were able to further characterize the subgroups in each of the studies by performing statistical analyses

evaluating subgroup differences with additional information such as neurocognitive data. Our results demonstrate the strength of the developed methods, and may pave the road for a broader understanding of the complexity of brain aging and Alzheimer's disease.

Degree Type
Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Electrical & Systems Engineering

First Advisor

Christos Davatzikos

Keywords

CHIMERA, clustering, HELIOS, heterogeneity, neuroimaging

Subject Categories

Electrical and Electronics

ANALYZING HETEROGENEITY IN NEUROIMAGING WITH PROBABILISTIC
MULTIVARIATE CLUSTERING APPROACHES

Aoyan Dong

A DISSERTATION
in
Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

Dr. Christos Davatzikos
Professor of Radiology, University of Pennsylvania

Graduate Group Chairperson

Dr. Alejandro Ribeiro
Professor of Electrical and Systems Engineering, University of Pennsylvania

DISSERTATION COMMITTEE

Victor Preciado, Assistant Professor of Electrical and Systems Engineering
Russell T. Shinohara, Assistant Professor of Biostatistics
Christos Davatzikos, Professor of Radiology

Dedicated to my family.

Acknowledgments

It has been a long journey towards this thesis, six years of Ph.D. study and seven years of life in Section for Biomedical Image Analysis. I am fortunate to arrive at the end of it, with the strong support of many people along the way.

I would like to offer my sincerest gratitude to my advisor, Dr. Christos Davatzikos. It is my honor to work with him and have him as my advisor. I am grateful for his trust and support during my tenure as a graduate student. He is a great advisor, as well as a great researcher, who has carefully and patiently guided me into the beauties in science. I have learned a lot from him: critical thinking, experiment design, scientific writing, presentation style, and more. Nothing in this dissertation would have been possible without his guidance.

I would like to express my deepest appreciation to Dr. Ben Taskar. He inspired me with machine learning and lighted up my confidence in research with his fabulous teaching and advising. It is my life treasure to be his student during his short but wonderful life. His boundless energy and enthusiasm is and will always lead my way in the field of artificial intelligence.

I would like to thank my mentor, Dr. Aristeidis Sotiras, who dedicated a large amount

of time to help me with methodology development, insightful research discussions, and papers writings, and for helping me prepare this manuscript. I would like to thank Dr. Nicolas Honnorat, who ramped me up with fundamentals of research. I am grateful to Dr. Ying Wang, Dr. Yangming Ou, and Dr. Tianhao Zhang, who shared their expertise and experiences with me during my early years of study.

My heartfelt thanks to Dr. Victor Preciado and Dr. Russell T. Shinohara for being on my dissertation committee and for their meaningful feedback on my work. I would also thank Dr. Shinohara for the advising and collaboration in my research.

The time spent at Penn is colorful with all my friends meet here. I would especially thank Erdem Varol, Harini Eavani and Ke Zeng for their accompany for more than five years as graduate fellows. I would thank Paraskevi Parmpi and Amanda Shacklett who provided me the convenience and support for every administrative issue. I would also like to thank all my colleagues, Aziz Ismail, Bilwaj Gaonkar, Birkan Tunc, Chunming Li, Drew Parker, Guray Erus, Hamed Akbari, Houwei Cao, Jia Wu, Jimit Doshi, Jingjing Gao, Jon Toledo, Kayhan Batmanghelich, Maria Peifer, Mark Bergman, Martin Rozycki, Meng-Kang Hsieh, Mohamad Habes, Sarthak Pati, Shandong Wu, Spyridon Bakas, Weiwei Zhang, and Xiao Da. I appreciate not only the inspirations in research but also the supporting of my life from all of them.

Lastly, my utmost gratitude and appreciation go to my family, to whom this dissertation is dedicated, without whom I could not complete this journey. Their love and support kept me going through all the difficult times in my Ph.D. study.

ABSTRACT

ANALYZING HETEROGENEITY IN NEUROIMAGING WITH PROBABILISTIC MULTIVARIATE CLUSTERING APPROACHES

Aoyan Dong

Christos Davatzikos

Automated quantitative neuroimaging analysis methods have been crucial in elucidating normal and pathological brain structure and function, and in building in vivo markers of disease and its progression. Commonly used methods can identify and precisely quantify subtle and spatially complex imaging patterns of brain change associated with brain diseases. However, the overarching premise of these methods is that the disease group is a homogeneous entity resulting from a single, unifying pathophysiological process that has a single imaging signature. This assumption ignores ample evidence for the heterogeneous nature of neurodegenerative diseases and neuropsychiatric disorders, resulting in incomplete or misleading descriptions. Accurate characterization of heterogeneity is important for deepening our understanding of neurobiological processes, thus leading to improved disease diagnosis and prognosis.

In this thesis, we leveraged machine learning techniques to develop novel tools that can analyze the heterogeneity in both cross-sectional and longitudinal neuroimaging studies. Specifically, we developed a semi-supervised clustering method for characterizing heterogeneity in cross-sectional group comparison studies, where normal and patient populations are modeled as high-dimensional point distributions, and heterogeneous disease effects are captured by estimating multiple transformations that align the two distributions, while accounting for the effect of nuisance covariates. Moreover, toward dissecting

the heterogeneity in longitudinal cohorts, we proposed a method which simultaneously fits multiple population longitudinal multivariate trajectories and clusters subjects into subgroups. Longitudinal trajectories are modeled using spatiotemporally regularized cubic splines, while clustering is performed by assigning subjects to the subgroup whose population trajectory best fits their data.

The proposed tools were extensively validated using synthetic data. Importantly, they were applied to study the heterogeneity in large clinical neuroimaging cohorts. We identified four disease subtypes with distinct imaging signatures using data from Alzheimer’s Disease Neuroimaging Initiative, and revealed two subgroups with different longitudinal patterns using data from Baltimore Longitudinal Study on Aging. Critically, we were able to further characterize the subgroups in each of the studies by performing statistical analyses evaluating subgroup differences with additional information such as neurocognitive data. Our results demonstrate the strength of the developed methods, and may pave the road for a broader understanding of the complexity of brain aging and Alzheimer’s disease.

Contents

Acknowledgments	iii
Abstract	v
Contents	vi
List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Overview	1
1.2 Contributions	5
1.3 Image preprocessing	7
1.3.1 Datasets	7
1.3.2 Feature extraction from structural scans	9
1.4 Organization of this thesis	11
2 Clustering Imaging Patterns of Disease Effect via Distribution Matching	12
2.1 Introduction	12

2.2	Method	14
2.2.1	Log-likelihood term	16
2.2.2	Model regularization	19
2.2.3	Optimization	20
2.2.4	Clustering	25
2.3	Experiments	25
2.3.1	Synthetic data	26
2.3.2	Neurodegenerative disease data	29
2.4	Conclusion and discussion	31
3	Capturing Heterogeneity in Prodromal Alzheimer’s Disease	34
3.1	Introduction	34
3.2	Materials and methods	37
3.2.1	Subjects	37
3.2.2	Cerebrospinal fluid (CSF) collection and measurement	38
3.2.3	MRI acquisition and processing	38
3.2.4	White matter hyperintensities (WMH)	39
3.2.5	Heterogeneity and voxel based morphometry analysis	39
3.2.6	Statistical analysis	40
3.3	Experiments and results	41
3.3.1	Cluster demographic and genetic characteristics	41
3.3.2	Cluster membership confidence	42
3.3.3	Cross-sectional clinical and biomarker associations	42
3.3.4	Group-wise VBM results	46

3.3.5	Longitudinal changes	47
3.3.6	Prevalence of clusters as a function of age	50
3.4	Discussion	52
4	Parsing Heterogeneous Longitudinal Trajectories	64
4.1	Introduction	64
4.2	Method	67
4.2.1	Loss term	69
4.2.2	Regularization term	72
4.2.3	Optimization	74
4.3	Experiments	76
4.3.1	Synthetic data	76
4.3.2	Longitudinal aging data	80
4.4	Discussion and conclusion	87
5	Summary and Future Work	90
5.1	Summary	90
5.2	Future work	92
A	Software	96
B	List of regions of interest	99
C	BLSA cognitive data	102
	Bibliography	103

List of Tables

2.1	Notation used in M-step.	22
3.1	Cluster demographics and characteristics of CSF biomarkers and cognitive scores.	45
3.2	Longitudinal neuropsychological associations of the clusters.	50
3.3	Regression coefficients and p-values of studying longitudinal associations of cognitive measures with aHV quartiles (Quartile 1 corresponds to the lowest volume, whereas Quartile 4 is the highest).	52
3.4	Summary of characteristics of clusters.	55
4.1	Longitudinal cognitive score difference between two subgroups	87
B.1	Names of 80 ROIs used in Chapter 3.	100
B.2	Names of 34 ROIs used in Chapter 4.	101

List of Figures

1.1	(A) Schema of analyzing group difference to find disease effects. (B) Underlying heterogeneity of disease effects, mixed with covariates effects.	5
1.2	Multi-atlas region of interest segmentation flowchart.	10
2.1	(A) The problem setting: \mathbf{X} is the reference distribution and \mathbf{Y} is the patient distribution. (B) Our model assumption: \mathbf{X} is transformed into a distribution \mathbf{X}' , covering the distribution \mathbf{Y} , by a set of K different transformations.	15
2.2	Atrophy patterns introduced (in red).	26
2.3	Simulated age effect on the normalized total volume. As age increases, the total volume linearly decreases and the variance of the ROI volumes increases.	27
2.4	Box plot of dice scores on synthetic data between ground truth labels and outputs of clustering methods: (a) K-means, (b) K-means with profile, (c) Hierarchical clustering, (d) Hierarchical clustering with profile, (e) CHIMERA-affine, (f) CHIMERA-duo, and (g) CHIMERA-trans.	29

2.5	Box plot of dice scores on dementia dataset between ground truth labels and outputs of clustering methods: (a) K-means, (b) K-means with profile, (c) Hierarchical clustering, (d) Hierarchical clustering with profile, (e) CHIMERA-affine, (f) CHIMERA-duo, and (g) CHIMERA-trans.	31
3.1	Cluster probabilities for subjects classified within each cluster. The vertical red line indicates a 0.5 probability of belonging to the cluster.	43
3.2	Venn diagram depicting the number of subjects classified tightly or loosely into each cluster. Subjects with a probability > 0.5 were included within a single cluster, whereas subjects with a highest cluster probability < 0.5 are depicted in the interphase of the two top clusters.	44
3.3	VBM between the identified clusters and the CN reference group for the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05.	47
3.4	VBM between clinical groups (MCI and AD) and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05.	48

3.5	VBM between patient clusters, stratified by adjusted hippocampal volumes, and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05. Quartile 1 represents the lowest volume, whereas Quartile 4 is the highest.	49
3.6	(A) Progression from MCI to AD stratified by MRI-defined clusters. (B) Progression from MCI to AD based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.	50
3.7	Longitudinal cognitive changes in ADAS-Cog13, memory and executive composite scores in MCI subjects stratified by MRI-defined clusters.	51
3.8	Standardized yearly MRI changes observed in CN subjects and MCI subjects belonging to the four identified clusters. A total of 124 CN, 57 Cluster 1, 44 Cluster 2, 18 Cluster 3 and 40 Cluster 4 subjects were included in the analysis in ADNI-1 (A). 84 CN, 15 Cluster 1, 17 Cluster 2, 17 Cluster 3 and 13 Cluster 4 subjects were included in the analysis in ADNI-GO/2 (B).	53
3.9	Prevalence of clusters as a function of age. (A) number of subjects with 5-year brackets, (B) relative frequency of clusters, fitted with cubic splines. . .	54
3.10	Cognitive longitudinal changes based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.	62

4.1	HELIOS takes as input a longitudinal dataset (illustrated in (a) by coloring all time points of one subject the same), a set of uniform cubic B-spline bases spanning the entire time range (shown in (b)), and the number of subgroups ($K=2$ here). Given these, HELIOS operates by simultaneously (c) estimating the subgroup-specific global trajectories as an optimal linear combination of the four bases, and (d) assigning each subject to the global trajectory that is most similar to it. The similarity is evaluated after accounting for differences in the vertical direction through the use of offset variable \mathbf{D} . At the end of the algorithm, clustering and fitting for all subjects with respect to the two global trajectories has been performed ((e) and (f), respectively).	68
4.2	Illustration of the coefficient tensor \mathbf{C} . C_{pk} denotes the coefficient vector for a single spline; the $\hat{\mathbf{C}}_p$ cross section contains K sets of coefficients; the $\tilde{\mathbf{C}}_k$ cross section contains P sets of coefficients.	70
4.3	Illustration of the least squares fit: all features of the same subject share basis $\mathbf{B}(t_i)$, but have different coefficients from \mathbf{C} that are chosen by the indicator ζ_i . To ease the illustration, the offset variable \mathbf{D} has not been included here.	72
4.4	(A) The three simulated trajectory patterns. (B) Construction of four subgroups following different multivariate patterns, where each subject has five features. The color of each feature indicates the trajectory it follows.	77
4.5	Comparison between the proposed method and the two k-means variants on simulated data. The performance is quantified by the adjusted rand index, where results are obtained by varying the lengths of individual trajectories.	78

4.6	Comparison between the proposed method and the two k-means variants on simulated data. The performance is quantified by the adjusted rand index, where results are obtained for different signal to noise ration levels, while fixing the individual trajectories length to be equal to 70% of the time range.	79
4.7	Distribution of scans per person by sex across the study age span. Each point denotes a scan; horizontal lines connect scans from the same individual. Red, female; blue, male.	81
4.8	Clustering performance measured by (A) the Adjusted rand index, and (B) the Dunn index, for different sets of hyperparameters K and η	83
4.9	Estimated trajectories for six different ROIs. The two identified subgroups are colored red (N=29) and blue (N=73), respectively. Subgroup trajectories are thick, while the trajectories for each participant are thin.	85
4.10	The brain regions that follow statistically significant (FDR corrected p value < 0.01) different trajectories between the two estimated subgroups found are shown. Red indicates G1 has a faster shrinking rate than G2, while blue indicates G1 has a faster expansion rate than G2.	86
A.1	Command-line interface of CHIMERA.	97
A.2	Command-line interface of HELIOS.	98

Chapter 1

Introduction

1.1 Overview

Over recent years, the advances in neuroimaging have enabled massive quantitative investigations of human brains, under normal and pathological conditions, and across the human lifespan. The in vivo and non-invasive multi-modal brain mapping techniques provide us with a wide array of tools of studying distinct aspects of brain structure and function. To name a few, structural magnetic resonance imaging (sMRI) reveals high-resolution brain anatomy for quantitative analysis of structural changes [85]; functional magnetic resonance imaging (fMRI) helps to measure brain activity by detecting changes associated with blood flow [138, 143]; and diffusion tensor imaging (DTI) makes it possible to understand the properties of the brain's white matter tracts [8].

Advanced neuroimaging methods have the potential to revolutionize our understanding of the brain in order and in disorder, by providing rich multi-parametric information, thus allowing advanced computer-aided diagnosis and prognosis. Towards this end,

studies that aim to characterize the imaging pattern associated with a disease or a normative process are designed. For example, the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort collects patients and elderly control subjects for the sake of learning the pathological process of early Alzheimer’s disease; the Human Connectome Project (HCP) recruits healthy young adults in order to gain knowledge of the neural pathways that underlie brain function and behavior.

With imaging data acquired with different aims, a large amount of research put their emphasis on one common problem: discovering regionally specific effects of brain processes by proposing various analytic tools and conducting experiments on different datasets. The variations of this problem are refined in a wide range of applications: by comparing a group of patients and healthy controls, the pathological effect of brain disease can be delineated; by observing normal subjects across a wide spectrum of age, the normative aging effect can be quantified; by differentiating typical and non-typical developed adolescents, aberrations from normal brain development, potentially leading to neuropsychiatric disorders, can be better understood.

Mass univariate tools are often used for testing hypotheses about the regionally specific effects. Deformation based morphometry (DBM) [20, 55] and tensor based morphometry (TBM) [48, 127] compare deformation fields or their spatial derivatives of different populations using statistical tools to detect specific voxel level anatomical variation within a group. DBM and TBM both rely on highly accurate registration of brain images, that may not always be possible given the large variation of human brains. On the other hand, voxel based morphometry (VBM) analysis [6, 29] conducts voxel wise t-tests across groups of tissue density maps of different populations to investigate focal differences in brain anatomy.

The generation of tissue density maps is robust to small registration errors, which makes VBM perhaps the most popular method in population neuroimaging analysis.

However, the univariate analysis performs statistical tests on a voxel by voxel, or region by region basis. Thus, these methods ignore multivariate relations between brain regions that may best characterize population differences. Instead, multivariate pattern analysis (MVPA) methods [7, 101] take advantage of dependencies among brain regions which leads to increased sensitivity. The MVPA methods can be further grouped into supervised and unsupervised learning approaches. 1) Supervised learning constitutes a set of algorithms that produce hypotheses from instances with known labeling (e.g. diagnosis, group membership), and make predictions about future instances. The supervised learning methods search for multivariate imaging patterns associated with the effect of interest. One of the most widely used methods is support vector machine (SVM) [22], which attempts to maximize the separation margin for different populations, that has been applied to multiple brain disease classifications [84, 87, 156]. 2) Unsupervised learning focuses on uncovering the latent structure of the imaging data. For example, principal component analysis (PCA) [1] and independent component analysis (ICA) [70] extract multivariate imaging signatures that can best explain the data variation, and are often applied to functional imaging [11, 49, 64]; clustering methods find subgroups of individuals with different imaging profiles [110, 115].

A common assumption behind most univariate and multivariate supervised analyses is that the difference between two groups is generated by a single, often unifying (patho)physiological process, that is characterized by a single imaging pattern. For instance, there is a unique disease effect that is found by comparing patients and controls.

However, ample evidence has highlighted the heterogeneity of pathological phenotypes presented by many diseases, such as Alzheimer’s disease [90, 110], Schizophrenia [43, 86, 113], Autism spectrum disorder [78, 145], and Attention-deficit hyperactivity disorder [158]. As a consequence, current approaches miss crucial information when describing diseases effects. By neglecting heterogeneity, these approaches can only find differences in the central tendency, such as a common imaging pattern of difference when comparing two populations, or an average trend of brain changes when modeling longitudinal imaging trajectories. The brain patterns described are therefore incomplete and can be misleading in the worst case.

There exist two types of approaches for analyzing the heterogeneity in neuroimaging. 1) The first group of methods uses *a priori* defined neuropathological categories to identify subgroups of subjects [77, 90, 110, 139]. 2) The second group comprises unbiased data-driven approaches to identify different patterns of pathology distribution based on the atrophy patterns inherent to the population [114, 115, 154]. However, in the former approaches, *a priori* definition of disease subtypes may be difficult to obtain (may need autopsies for neuropathological findings), or might be quite noisy and non-specific (e.g., cognitive or clinical evaluations). In the latter approaches, standard unsupervised clustering methods are used to group patients along the direction associated with the largest data variability, which may not be induced by the pathology, and it might conversely reflect effects such as age, gender or disease stage.

A more specific characterization of anatomical heterogeneity of brain processes is likely to lead to grouping of subjects into relatively homogeneous groups with potentially more predictable clinical outcomes and treatment responses. Therefore, the diagnosis of brain

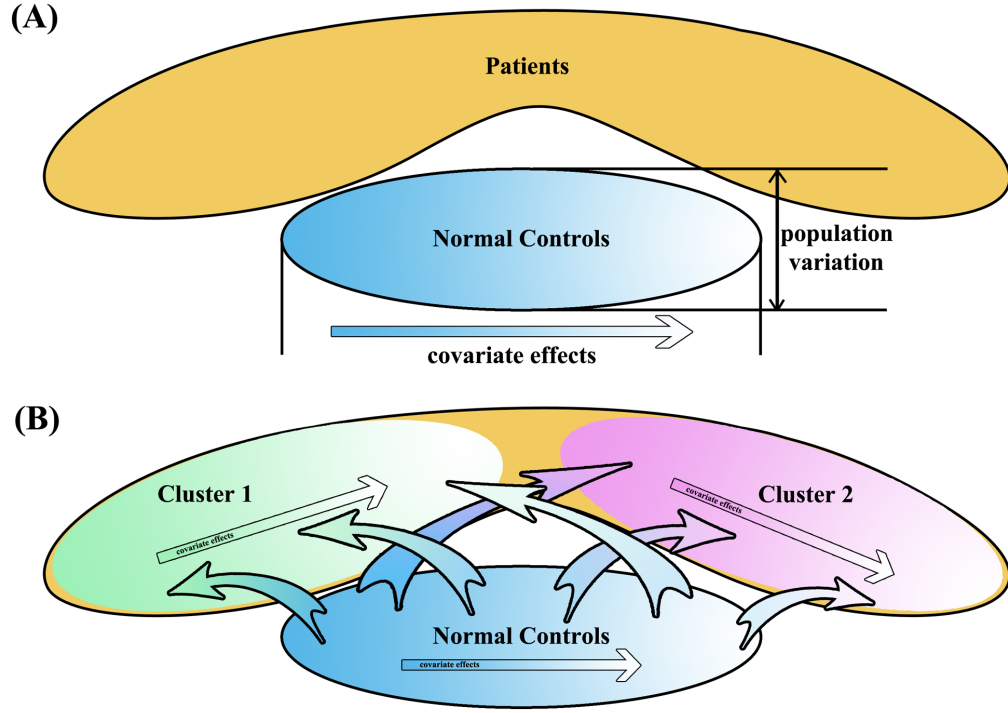


Figure 1.1: (A) Schema of analyzing group difference to find disease effects. (B) Underlying heterogeneity of disease effects, mixed with covariates effects.

diseases can be significantly improved, and the findings can be utilized later in improving disease prognosis, precision medicine and patient recruiting for more targeted clinical trials [83, 119].

1.2 Contributions

Towards tackling the above limitations, we proposed two unbiased data-driven approaches that explicitly take into account heterogeneity in cross-sectional and longitudinal studies, respectively. In summary, this work makes the following major contributions.

- We proposed a semi-supervised clustering algorithm, named CHIMERA (clustering of heterogeneous disease effects via distribution matching of imaging patterns), for

cross-sectional studies as illustrated in Figure 1.1. Instead of directly finding subgroups using standard clustering methods, we utilized information from healthy controls to guide the clustering, by assuming the probability distribution of patients is derived via a number of transformations of the probability distribution of healthy controls. These transformations define imaging signatures of heterogeneous disease processes. Viewed differently, our approach clusters differences between two datasets instead of clustering data itself. This proposed paradigm has two advantages. First, compared to previous work [53], which produces clustering results on the subject level and thus suffers from various uninteresting variations due to the covariates, the employed distribution matching scheme herein generates clustering on the distribution level that helps reduce the influence of population variation significantly. Second, the probabilistic modeling provides an intrinsic kernelized distance metric, which allows measuring the similarity between subjects nonlinearly. Thus, covariate effects that are often removed by an explicit linear regression step [100], can now be taken into account in a generic and nonlinear way.

- The developed semi-supervised clustering algorithm was applied to a large dataset of Alzheimer’s disease. We found four distinct neuroanatomical subtypes in mild cognitive impairment and Alzheimer’s disease dementia relative to cognitive normals, which were replicable and consistent across two different cohorts. We further analyzed in detail the cerebrospinal fluid (CSF) biomarkers, cognitive characteristics and white matter hyperintensity (WMH) volumes of these subtypes. This application opens the way of constructing a global dementia coordinate system, which potentially can lead to a revolution in therapeutic innovation for neurodegenerative

diseases.

- We extended the heterogeneity analysis to tackle longitudinal designs, where we proposed HELIOS (parsing the heterogeneity of longitudinal imaging through integrated clustering and spatiotemporally regularized spline curve fitting). To the best of our knowledge, it is the first study that focuses on heterogeneous longitudinal trajectories of multivariate imaging measures. The proposed method clusters individual trajectories aiming to find multiple global trajectories that can best describe the brain change across the full age range of interest of the population. The trajectories are modeled using spatiotemporal regularized splines, which 1) produces smooth and nonlinear curves in the temporal domain; 2) introduces a biological prior to the modeling. This method can be viewed as an enhanced version of linear-mixed effect models with clustering on top.

1.3 Image preprocessing

Neuroimaging data obtained from the scanner cannot be used for our analysis directly. In this section, we describe the datasets and the image preprocessing steps that were used in all of our experiments.

1.3.1 Datasets

Alzheimer’s Disease Neuroimaging Initiative (ADNI)

The Alzheimer’s Disease Neuroimaging Initiative (ADNI)¹ is an ongoing multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early

¹<http://www.adni-info.org>

detection and tracking of Alzheimer's disease (AD). The initial phase of ADNI (ADNI-1) started in 2004, with \$67 million funding provided by both the public and private sectors. ADNI-1 recruited 400 subjects diagnosed with mild cognitive impairment (MCI), 200 subjects with early AD and 200 elderly cognitive normal subjects. This study was extended with ADNI-GO which added 200 participants identified as having early mild cognitive impairment (EMCI). In 2011, ADNI-2 began with another \$67 million in funding. ADNI-2 assesses participants from the ADNI-1/ADNI-GO cohort in addition to the following new participants: 150 elderly controls, 100 EMCI, 150 LMCI (late "mild cognitive impairment") participants and 150 mild AD patients. A subset of the data from this study is used in Chapter 3 to find Alzheimer's disease subtypes.

Baltimore Longitudinal Study on Aging (BLSA)

The Baltimore Longitudinal Study on Aging (BLSA) [140] is America's longest-running scientific study of human aging at the National Institute on Aging (NIA). Its neuroimaging sub-study began in 1994 and is currently scheduled to continue indefinitely. Approximately 150 healthy elderly adults (age range: 56-85) were initially recruited and have been followed annually with a rich radiologic examination resulting to 740 structural MRI scans. The imaging protocol includes structural (T1-, and T2-weighted MR images) and functional activation scans. More recently, additional subjects were included raising the number of total imaging sessions to over 2500. Imaging data are complemented with detailed neurocognitive evaluations performed under the direction of Dr. Resnick. A subset of the data from this study is used for methodology testing in Chapter 4, where a more extensive analysis of this dataset is described.

1.3.2 Feature extraction from structural scans

Region of interest (ROI) volumetry

The high dimensionality of MR images hinders their analysis and interpretation. Extracting region of interests (ROI) effectively reduces the dimensionality of the data in an interpretable and anatomically meaningful way. We employed a multi-atlas segmentation algorithm [38] which uses a consensus labeling framework to fuse/integrate segmentation hypotheses generated by warping a broad ensemble of labeled atlases to the target space via the use of several warping algorithms, regularization parameters, and atlases. The label fusion integrates two complementary sources of information: a local similarity ranking to select locally optimal atlases and a boundary modulation term to refine the segmentation consistently with the target image's intensity profile. The flowchart of the ROI algorithm is presented in Figure 1.2. In our analyses, we used this algorithm to partition the brain into approximately one hundred disjoint ROIs generated, and obtained the volume of each ROI as a feature representation of the brain.

Tissue density maps

ROIs provide us with data in a dimension that we can easily handle in order to parse disease heterogeneity. However, in order to characterize disease processes in greater spatial detail, we employed tissue density maps for subsequent subgroup-analyses. Towards this end, we employed tissue density maps that allow us to characterize disease processes in greater spatial detail. Specifically, we employed a previously published volumetric approach to generate tissue density map for group comparisons [28, 29], termed RAVENS

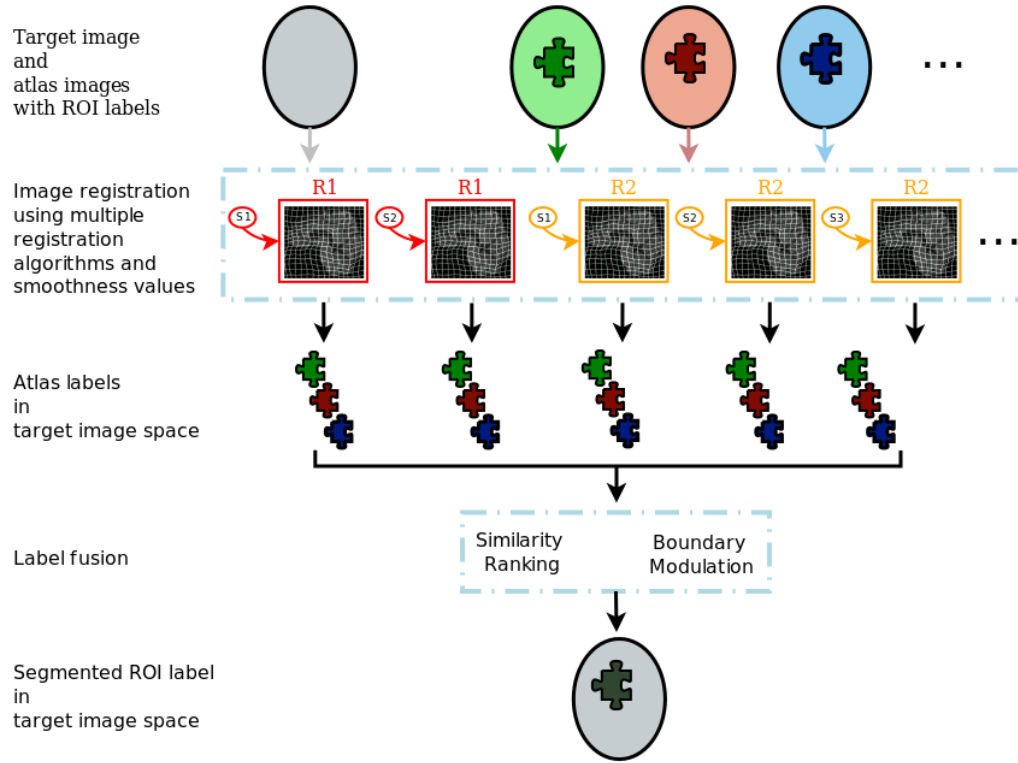


Figure 1.2: Multi-atlas region of interest segmentation flowchart.

(regional analysis of volumes examined in normalized space) map. The RAVENS maps are obtained with the following procedures. An established deformable registration method [118] is used for warping individual images to a single subject brain template. The brain image scans are segmented into three tissue types: gray matter, white matter and cerebrospinal fluid [95]. RAVENS maps encode, locally and separately for each tissue type, the volumetric changes (local expansion or shrinkage) observed during the registration. They hence have the advantage of accounting for imperfect registration by taking the residual (error) of the imperfect registration into account.

1.4 Organization of this thesis

The two main methodological contributions of this thesis are described in Chapters 2 and 4. Chapter 2 details the CHIMERA method, its performance on simulated data, and validation on real data. Chapter 3 presents the main application of CHIMERA to Alzheimer's disease dataset (ADNI), with a detailed analysis of the four subtypes found, including the imaging signatures, demographics characteristics, the level of CSF biomarkers and the cognitive performances. In Chapter 4, we describe the longitudinal approach for heterogeneity trajectories, validate its performance using synthetic data, and apply it to a real dataset (BLSA). Chapter 5 summarizes all the contributions of this thesis and discusses future work.

Chapter 2

Clustering Imaging Patterns of Disease Effect via Distribution Matching

2.1 Introduction

Group analyses are ubiquitous in neuroimaging, which are commonly applied to study the differences between populations. Typical applications include, but are not limited to, analyses for describing disease effects by comparing patients and controls [44, 79, 167], studies for characterizing aging effects by comparing old and young subjects [17, 59], as well as efforts to characterize brain development by comparing subjects of different ages [57, 143]. Statistical group analyses are carried out throughout studies using diverse types of images, including functional MRI [148, 162], structural MRI [29, 50, 67], and diffusion tensor imaging [60, 144].

Most of the group analyses assume that the members of a group share a common imaging pattern that differentiates them from the other group. For example, they assume that there is a unique disease effect that is found by comparing patients and controls. Such an approach can only find changes in the central tendency, i.e., a “common denominator”. However, various clinical studies have highlighted the heterogeneity of pathological phenotypes presented by many diseases, such as Alzheimer’s disease [110, 114], schizophrenia [43, 105], autism spectrum disorder [145], attention-deficit hyperactivity disorder [158] and cancer [54, 98]. In such cases, where we can assume that two groups differ by one pattern in one sub-population, and a different pattern in another sub-population, a “common denominator” is at best incomplete, and at worst misleading. As a consequence, current approaches are limited in the presence of heterogeneity as they miss crucial information when modeling disease effects.

Towards addressing the challenges posed by heterogeneity, the use of clustering methods to partition the population of patients has been proposed [80, 97, 115, 160]. However, direct clustering of patient images puts emphasis on the similarities/distances between individuals, rather than on the heterogeneity of the disease effect itself. Hence, they produce clusterings which reflect the largest contributors of data variability, such as brain size, participant sex, and scanner/protocol discrepancies, and may fail to cluster the individuals according to their pathology subtypes. In order to mitigate this problem, general linear regression [100, 116] is commonly applied to residualize imaging features with respect to covariates. Nevertheless, the effectiveness of this approach suffers from two limitations: 1) the linear assumption may not hold thus the regression might introduce more noise to the analysis; 2) covariates may correlate to the disease such that removing the covariate effect

may undermine the strength of disease effect analysis.

We propose to address the aforementioned limitations by proposing a novel regularized clustering method based on establishing a mapping between two statistical distributions. The first statistical distribution corresponds to the reference population, e.g., healthy controls, cognitive stable participants, or normally developing adolescents. The second distribution corresponds to the patient population that has been deviated from the reference population under the influence of a number of effects that we would like to describe. These effects may include heterogeneous disease processes, pathophysiological processes leading to cognitive decline, or aberrations from normal brain development. As shown in Figure 2.1, we model the heterogeneous effects as a set of transformations from the reference to the patient distribution, where each transformation corresponds to one pathology subtype. The transformations are found by matching patient and reference distributions, while taking covariates such as age, sex, scanner, etc. into account (which exactly covariates are to be used depends highly on the specific application/study). In other words, given that a 70-year-old male Alzheimer’s disease patient would have been a 70-year-old male control had he been spared from the disease, the transition between these two states is considered to be the disease effect. This covariate-informed matching reduces the confounding influence of the covariates, which leads to a better description of the disease effects.

2.2 Method

Let us assume that the dataset contains M normal control (NC) samples $\mathbf{X} = \{x_1, \dots, x_M\}$ and N patient samples $\mathbf{Y} = \{y_1, \dots, y_N\}$. Let us assume that the samples are described by

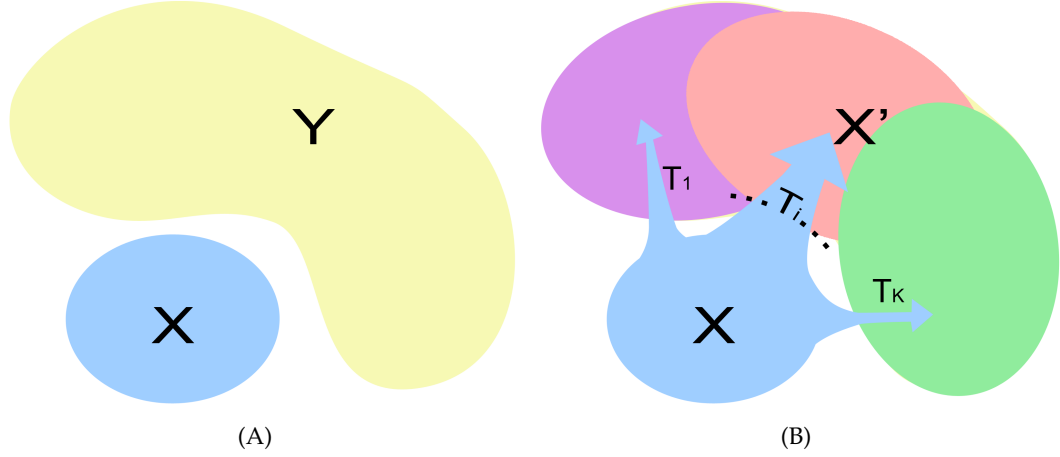


Figure 2.1: (A) The problem setting: \mathbf{X} is the reference distribution and \mathbf{Y} is the patient distribution. (B) Our model assumption: \mathbf{X} is transformed into a distribution \mathbf{X}' , covering the distribution \mathbf{Y} , by a set of K different transformations.

two sets of features: a set of D_1 -dimensional imaging features: $x_m^v, y_n^v \in \mathbb{R}^{D_1}$; and a set of D_2 -dimensional covariate features (these are known variables, such as age, sex, tumor type, treatment type): $x_m^c, y_n^c \in \mathbb{R}^{D_2}$. For the sake of simplicity, we will denote the samples in the compact vector forms: $x_m = (x_m^v, x_m^c)$ and $y_n = (y_n^v, y_n^c)$.

Without loss of generality, let us consider the samples as points in the imaging space (Figure 2.1). In this setting, the pathology can be viewed as the difference between \mathbf{Y} (patient) and \mathbf{X} (NC) point distributions. In order to model this difference, we adopt a probabilistic framework where the pathologic transition between the two groups is modeled as a transformation \mathbf{T} , mapping the NC distribution to the patient distribution. One key assumption of the proposed framework is that we have collected enough data to allow us to adequately describe the NC population, and that the estimated anatomy of the patients would have been covered by the NC distribution, had they been spared of the disease. These assumptions imply that all patients can be associated with NCs, and conversely, the transformed NC points cover the entire set of patients. In this setting, we can

estimate the transformation \mathbf{T} by matching the patient and NC distributions. The distribution matching paradigm consists of distance measures of imaging features, thus it is convenient to introduce covariates into the matching criteria by combining imaging and covariate-specific distances in a multi-kernel way [91].

The distribution matching problem is formulated as a maximum *a posteriori* (MAP) optimization problem. Thus, the optimal transformation is estimated by minimizing the following energy:

$$\mathcal{E}(\mathbf{X}, \mathbf{Y}, \Theta) = -\mathcal{L}(\mathbf{X}, \mathbf{Y}, \Theta) + \mathcal{R}(\Theta), \quad (2.1)$$

where Θ denotes the parameters of our model, such as transformations that are applied to \mathbf{X} for generating \mathbf{Y} , \mathcal{L} is the log-likelihood of the distributions \mathbf{X} and \mathbf{Y} given the parameters, while a regularization/penalty \mathcal{R} improves the stability/reliability of the estimation. These two parts are presented in detail in the next two sections.

2.2.1 Log-likelihood term

Due to the heterogeneity of the effects of a given disease, the pathological transition might take several directions. Therefore, \mathbf{T} is modeled using multiple possible transformations, where each of them represents a pathological direction of imaging change. The transformed NC samples are denoted as $\mathbf{X}' = [x'_1, \dots, x'_M]$, where the imaging feature x_m^v is transformed to $\mathbf{T}(x_m^v)$, while the covariate feature x_m^c remains the same:

$$x'_m = \mathbf{T}(x_m) = (\mathbf{T}(x_m^v), x_m^c). \quad (2.2)$$

Based on the hypothesis that the origins of patient samples are covered by the NC sample space, we deduce that if we apply the pathological process model to the NC samples \mathbf{X} , the transformed NC point distribution \mathbf{X}' will cover the patient point distribution \mathbf{Y} , as shown in Figure 2.1(B).

The matching of distributions \mathbf{Y} and \mathbf{X}' is found by a variant of the coherent point drift algorithm [111]. Each point x'_m is considered as a centroid of a spherical Gaussian cluster. All the clusters are assumed to have the same variance σ^2 , which is inferred by the method. Points y_n are treated as i.i.d. data generated by a Gaussian Mixture Model (GMM) [15] with equal weight $P(x'_m) = \frac{1}{M}$ for each cluster. The similarity between the two distribution is measured by the data likelihood of this mixture model, as presented in Equation (2.3).

In order to take covariate features into account, we adopt a multi-kernel setting. The distance between two points is measured by RBF kernels, where the kernel size of covariate features is r times larger than the kernel size of the imaging features. As a result, the likelihood of data \mathbf{Y} generated by centroids \mathbf{X}' can be described as follows:

$$\begin{aligned} P(\mathbf{X}, \mathbf{Y}) &= \prod_{n=1}^N \sum_{m=1}^M P(x'_m) P(y_n | x'_m) \\ &= \prod_{n=1}^N \sum_{m=1}^M \frac{1}{M} \frac{r^{D_2/2}}{(\sqrt{2\pi}\sigma)^{D_1+D_2}} \cdot \exp \left\{ \frac{\|y_n^v - \mathbf{T}(x_m^v)\|^2 + r\|y_n^c - x_m^c\|^2}{-2\sigma^2} \right\}. \end{aligned} \quad (2.3)$$

During our experiments, the hyper-parameter r was determined by the ratio of total variance of these two features.

We assume that there are K pathology directions T_1, \dots, T_K for a given disease. We

define the transformation for one NC point to the patient space as:

$$\mathbf{T}(x_m^v) = \sum_{k=1}^K \zeta_{km} T_k(x_m^v). \quad (2.4)$$

Ideally, if the disease subtypes were generated by distinct groups of NC points, ζ_{km} would be 1 for the transformation corresponding to the disease subtype that affects x_m , and value 0 otherwise. In this work, we assume that patients with different pathologies might correspond to the same point in the space of NC distribution, and we relax the variable ζ_{km} to sum up to 1 for each m . This relaxation leads us to consider the transformation \mathbf{T} for each NC point x_m as a convex combination of all possible transformations T_k .

Linear transformation were chosen to model T_k , in order to derive analytical solutions for the distribution matching. Each T_k was described by a pair of parameters $(A_k, b_k) \in (\mathbb{R}^{D_1 \times D_1}, \mathbb{R}^{D_1})$:

$$\mathbf{T}(x_m^v) = \sum_{k=1}^K \zeta_{km} (A_k x_m^v + b_k), \quad (2.5)$$

where $\sum_k \zeta_{km} = 1$ and $\zeta_{km} \geq 0$ for all m .

During our experiments, three different kinds of A_k matrices were chosen: (1) full matrices (CHIMERA-affine), (2) diagonal matrices, in order to restrict the transformations to the combinations of scaling and translations (CHIMERA-duo) and (3) the identity, in order to consider only the translations b_k (CHIMERA-trans).

Introducing the definition of transformations T_k into Equation (2.3) leads to the follow-

ing expression for the log-likelihood of the data:

$$\begin{aligned} \mathcal{L}(\mathbf{X}, \mathbf{Y}, \Theta) = & \sum_{n=1}^N \log \sum_{m=1}^M \frac{1}{M} \frac{r^{D_2/2}}{(\sqrt{2\pi}\sigma)^{D_1+D_2}} \exp \left\{ \frac{r\|y_n^c - x_m^c\|^2}{-2\sigma^2} \right\} \\ & \cdot \exp \left\{ \frac{\|y_n^v - \sum_{k=1}^K \zeta_{km}(A_k x_m^v + b_k)\|^2}{-2\sigma^2} \right\}. \end{aligned} \quad (2.6)$$

2.2.2 Model regularization

As defined in the previous section, in an imaging feature space of dimension D_1 , the dimension of parameter space of CHIMERA-affine is in the order of $\mathcal{O}(D_1^2)$, while for CHIMERA-duo and CHIMERA-trans is in the order of $\mathcal{O}(D_1)$. In the low sample size settings that are typically observed in medical imaging studies, this large dimension yields ill posed problems. This issue is commonly mitigated by regularizing/penalizing the parameters of the transformations [41, 132]. We have adopted this approach, which improves also the generalization and the robustness of our model. In order to derive an analytical solution, we have chosen to penalize the Frobenius norm of $A_k - \mathbf{I}$ and the ℓ_2 norm of b_k , where \mathbf{I} is the identity matrix. This regularization, is equivalent to posing Gaussian priors for the parameters.

$$\mathcal{R}(\Theta) = \frac{\lambda_1}{2\sigma^2} \sum_k \|b_k\|_2^2 + \frac{\lambda_2}{2\sigma^2} \sum_k \|A_k - \mathbf{I}\|_F^2. \quad (2.7)$$

Beside the explicit regularization term \mathcal{R} , our model can also be considered as being “implicitly” regularized. Instead of focusing on the points at the border between the different groups, like support vector machine [22] and relevance vector machine [149], our model always consider the entire point distributions. We aim, in that way, to reduce the sensi-

tivity of clustering produced with respect to the individual subject variability (population variability).

2.2.3 Optimization

In this work, we have used an Expectation-Maximization algorithm [15, 108] for optimizing the parameters $\Theta = (\mathbf{A}, \mathbf{b}, \zeta, \sigma^2)$ of our model, where $\mathbf{A} = \{A_1, \dots, A_K\}$ and $\mathbf{b} = \{b_1, \dots, b_K\}$. The algorithm introduces latent variables z indicating the posterior probability of data point n for each mixture component m , $q_{nm} = q(z_n = x'_m | y_n)$. By doing so, it provides a lower bound of the log-likelihood [15]:

$$\mathcal{F}_0 = \sum_{n,m} q_{nm} \log \left(\frac{P(y_n, x'_m)}{q_{nm}} \right). \quad (2.8)$$

The energy \mathcal{E} is minimized via an iterative scheme. In each iteration t , the algorithm alternates between calculating in the E-step the expected value of q with respect to the parameters obtained in the previous iteration $\Theta^{(t-1)}$, and updating $\Theta^{(t)}$ by minimizing the objective function (Equation (2.10)) in the M-step.

During our experiments, at the initialization, the parameters σ^2 was set to the mean distance between datasets \mathbf{X} and \mathbf{Y} , ζ was set to be uniformly distributed for each x_m , each A_k was set to the identity matrix \mathbf{I} , while the translation term b_k was sampled from a normal distribution $\mathcal{N}(0, 1)$. The E-step and M-step were performed as follows.

E-Step:

Using the parameters $\Theta^{(t-1)}$ estimated in the previous M-step, Equation (2.8) was op-

timized at $q_{nm} = P(z_n = x'_m | y_n)$:

$$q_{nm} = \frac{\exp\left(\frac{\|y_n^v - \sum_k \zeta_{km}(A_k x_m^v + b_k)\|_2^2 + r\|y_n^c - x_m^c\|_2^2}{-2\sigma^2}\right)}{\sum_{i=1}^M \exp\left(\frac{\|y_n^v - \sum_k \zeta_{ki}(A_k x_i^v + b_k)\|_2^2 + r\|y_n^c - x_i^c\|_2^2}{-2\sigma^2}\right)}. \quad (2.9)$$

M-Step:

We constructed our objective energy function $\mathcal{F}(\Theta)$ as an upper bound of our energy function \mathcal{E} . The minimization of $\mathcal{F}(\Theta)$ leads to the minimization of \mathcal{E} [112]:

$$\begin{aligned} \mathcal{F}(\Theta) = & \frac{1}{2\sigma^2} \sum_{m,n} q_{nm} \left(\|y_n^v - \sum_k \zeta_{km}(A_k x_m^v + b_k)\|_2^2 + r\|y_n^c - x_m^c\|_2^2 \right) \\ & + \frac{N(D_1 + D_2)}{2} \log \sigma^2 + \frac{\lambda_1}{2\sigma^2} \sum_k \|b_k\|_2^2 + \frac{\lambda_2}{2\sigma^2} \sum_k \|A_k - \mathbf{I}\|_F^2, \end{aligned} \quad (2.10)$$

subject to

$$\sum_{k=1}^K \zeta_{km} = 1 \text{ for } m = 1, \dots, M, \quad 0 \leq \zeta_{km} \leq 1.$$

The objective function is not globally convex but jointly convex in each parameter. Hence, we propose an iterative procedure by minimizing the objective sequentially with respect to σ^2 , ζ , \mathbf{A} and \mathbf{b} . We derived a closed form solution for σ^2 , \mathbf{A} and \mathbf{b} by setting the derivative of the objective function to zero. ζ was optimized using an advanced projected gradient descent algorithm that preserves the sum of the ζ_{km} [39]. The notation used in the following equations is presented in Table 2.1. We define $\mathbf{d}(\mathbf{u})$ as a square matrix with the diagonal elements to be the vector \mathbf{u} , and $\mathbf{1}$ as a vector with all elements being 1 with appropriate dimension.

Notation	Dimension	Description
\mathbf{X}_v	$D_1 \times M$	m th column is $\mathbf{T}(x_m^v)$
\mathbf{X}_c	$D_2 \times M$	m th column is x_m^c
\mathbf{Y}_v	$D_1 \times N$	n th column is y_n^v
\mathbf{Y}_c	$D_2 \times N$	n th column is y_n^c
\mathbf{A}	$D_1 \times K D_1$	concatination of A_k , $\mathbf{A} = [A_1, \dots, A_K]$
\mathbf{b}	$D_1 \times K$	concatination of b_k , $\mathbf{b} = [b_1, \dots, b_K]$
ζ	$K \times M$	$[\zeta]_{km} = \zeta_{km}$, i th row is $\zeta_{i\cdot}$, m th column is $\zeta_{\cdot m}$
\mathbf{Q}	$N \times M$	$[\mathbf{Q}]_{nm} = q_{nm}$

Table 2.1: Notation used in M-step.

1. Update σ^2 : The estimation of the variance is updated as follows:

$$\begin{aligned}
\sigma^2 = & \left[\text{tr}(\mathbf{Y}_v \mathbf{d}(\mathbf{Q}\mathbf{1}) \mathbf{Y}_v^T - 2\mathbf{Y}_v \mathbf{Q} \mathbf{X}_v^T + \mathbf{X}_v \mathbf{d}(\mathbf{Q}^T \mathbf{1}) \mathbf{X}_v^T) \right. \\
& + \text{tr}(\mathbf{Y}_c \mathbf{d}(\mathbf{Q}\mathbf{1}) \mathbf{Y}_c^T - 2\mathbf{Y}_c \mathbf{Q} \mathbf{X}_c^T + \mathbf{X}_c \mathbf{d}(\mathbf{Q}^T \mathbf{1}) \mathbf{X}_c^T) \\
& \left. + \lambda_1 \sum_k \|b_k\|_2^2 + \lambda_2 \sum_k \|A_k - \mathbf{I}\|_F^2 \right] / N(D_1 + D_2). \quad (2.11)
\end{aligned}$$

2. Update \mathbf{A} : depending on the nature of the matrix A , the following cases are discerned:

- When A_k is full matrix, let C_{ij} be the (i, j) th block of \mathbf{C} , and $\mathbf{G} = [G_1, \dots, G_K]$, where C_{ij} and G_i are obtained by:

$$C_{ij} = \mathbf{X}_v \mathbf{d}(\mathbf{Q}^T \mathbf{1}) \mathbf{d}(\zeta_{i\cdot}) \mathbf{d}(\zeta_{j\cdot}) \mathbf{X}_v^T \quad (2.12)$$

$$G_i = \lambda \mathbf{I} + \mathbf{Y} \mathbf{Q} \mathbf{d}(\zeta_{i\cdot}) \mathbf{X}_v^T - \mathbf{B} \mathbf{d}(\mathbf{Q}^T \mathbf{1}) \mathbf{d}(\zeta_{i\cdot}) \mathbf{X}_v^T. \quad (2.13)$$

By setting the derivative of $\mathcal{F}(\Theta)$ with respect to \mathbf{A} to be 0, we can calculate \mathbf{A}

using Equation (2.14):

$$\mathbf{A} = \mathbf{G}(\lambda \mathbf{I} + \mathbf{C})^{-1}. \quad (2.14)$$

- When A_k is a diagonal matrix: we first denote $A_i = \mathbf{d}(a_{i1}, \dots, a_{iD})$ and we will calculate all j th diagonal elements for all k together, $\mathbf{a}_j \in \mathcal{R}^{K \times 1}$. Let \mathbf{W}_{rs}^j be the (r, s) th element of $K \times K$ matrix \mathbf{W}^j , which can be derived as:

$$\mathbf{W}_{rs}^j = \sum_m x_{mj}^2 \sum_n q_{nm} \zeta_{rm} \zeta_{sm}. \quad (2.15)$$

Let $z_m = \sum_k \zeta_{km} b_k$, and \mathbf{U}_i^j be the i th element of $K \times 1$ vector \mathbf{U}^j :

$$\mathbf{U}_i^j = \sum_{m,n} q_{nm} \zeta_{im} (z_m - y_n)_j x_{mj}. \quad (2.16)$$

Then, we can obtain the update of \mathbf{a}_j as follows:

$$\mathbf{a}_j = (\mathbf{W}^j + \lambda \mathbf{I})^{-1} (\lambda \mathbf{1} - \mathbf{U}^j). \quad (2.17)$$

3. Update \mathbf{b} :

Let \mathbf{V}_{rs} be the (r, s) th element of $K \times K$ matrix \mathbf{V} , and \mathbf{Z}_r be the r th row of $K \times D_1$ vector \mathbf{Z} , respectively:

$$\mathbf{V}_{rs} = \sum q_{nm} \zeta_{rm} \zeta_{sm} \quad (2.18)$$

$$\mathbf{Z}_r = \sum_{m,n} q_{nm} \zeta_{rm} (y_n^v - \sum_k \zeta_{km} A_k x_m^v). \quad (2.19)$$

Thus, we can derive \mathbf{b} as:

$$\mathbf{b}^T = (\lambda \mathbf{I} + \mathbf{V})^{-1} \mathbf{Z}. \quad (2.20)$$

4. Update ζ :

We adopted a gradient descent method for optimizing ζ , and projected the obtained vector to the ℓ_1 simplex in order to satisfy the sum-to-one constraint. Let \mathbf{H}^m be the hessian for the m th column of ζ , which can be obtained element wise using:

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \zeta_{im}} &= \frac{1}{\sigma^2} \sum_n q_{nm} (y_n^v - \sum_k \zeta_{km} (A_k x_m^v + b_k))^T (-A_i x_m^v - b_i) \\ \mathbf{H}_{ij}^m &= \frac{\partial^2 \mathcal{F}}{\partial \zeta_{im} \partial \zeta_{jm}} = \frac{1}{\sigma^2} \sum_n q_{nm} (A_j x_m^v + b_j)^T (A_i x_m^v + b_i). \end{aligned} \quad (2.21)$$

Given the above hessian estimate, we performed gradient descent, and projected the new vector to the ℓ_1 simplex: [39].

$$\zeta_{\cdot m}^{new} = \zeta_{\cdot m}^{old} - (\mathbf{H}^m + \mu \mathbf{I})^{-1} \frac{\partial \mathcal{F}}{\partial \zeta_{\cdot m}}. \quad (2.22)$$

During our experiments, we stopped iterating when the objective difference between two iterations reached a predefined tolerance, which was set to 0.01. Because the EM algorithm only guarantees a local minimum solution, we ran the optimization several times, and we kept the solution with the lowest energy value.

The next section explains how a clustering can be derived from the coefficients ζ_{km} and the posteriors q_{nm} which were estimated during the optimization, and how a new sample can be assigned to these clusters.

2.2.4 Clustering

The coefficients ζ_{km} can be considered as the probability, for the NC sample x_m , to undergo the transformation T_k . Let $P(y_n|x_m)$ be the likelihood of a patient sample y_n to be associated with x_m . Then, the likelihood of a given patient sample, y_n , to have been generated by the transformation T_k can be estimated by:

$$P_k(y_n) = \sum_m P(y_n|x_m)\zeta_{km}. \quad (2.23)$$

Because the posteriors q_{nm} are proportional to $P(y_n|x_m)$, with a common denominator for each n (Equation (2.9)), they can be used for partitioning the patient samples according to their main transformation. Thus, each patient y_n can be assigned to the label l_n , which corresponds to the largest likelihood:

$$l_n = \operatorname{argmax}_k P_k(y_n) = \operatorname{argmax}_k \sum_m q_{nm}\zeta_{km}. \quad (2.24)$$

As long as the ζ_{km} are stored, the label can be estimated for a novel data s by: (1) computing the likelihood $P(s|x_m)$ based on the distances between the novel sample s and the transformed controls \mathbf{X}' , (2) computing $P_k(s)$, and (3) obtaining the label $l_s = \operatorname{argmax}_k P_k(s)$.

This strategy was adopted for clustering clinical data during our experiments.

2.3 Experiments

This section presents the experiments that were conducted for validating our approach. We compared first our approach with two standard clustering methods, i.e., K-means [97] and

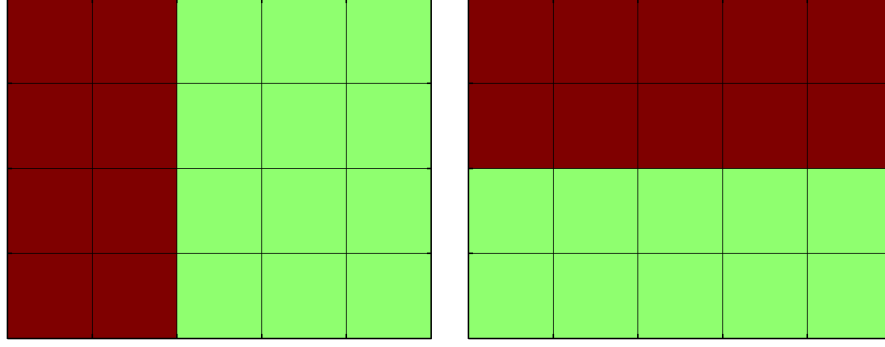


Figure 2.2: Atrophy patterns introduced (in red).

Ward hierarchical clustering [160], as well as two variants of these methods, on synthetic data and a real dataset of dementia patients with known subtypes. The promising results obtained incited us to analyze a clinical dataset where the ground truth is unknown.

2.3.1 Synthetic data

Our method was first validated using synthetic data simulating the effect of age and disease on brain volume. The brain was divided into 20 regions of interest (ROIs), where the atrophy was described by a normalized volume between 0 (the most serious atrophy) and 1 (largest possible ROI volume).

The simulated data was generated as follows:

1. 1000 samples were generated independently. For each sample, 20 ROI volumes were sampled randomly from a normal distribution, $\mathcal{N}(1, 0.1)$. In addition, each sample was associated with a random age, sampled from a uniform distribution between 55 and 85.
2. Age effect was introduced for each ROI volume and every sample, by subtracting the atrophy volume. The ROI volume atrophy was simulated by a normal distribu-

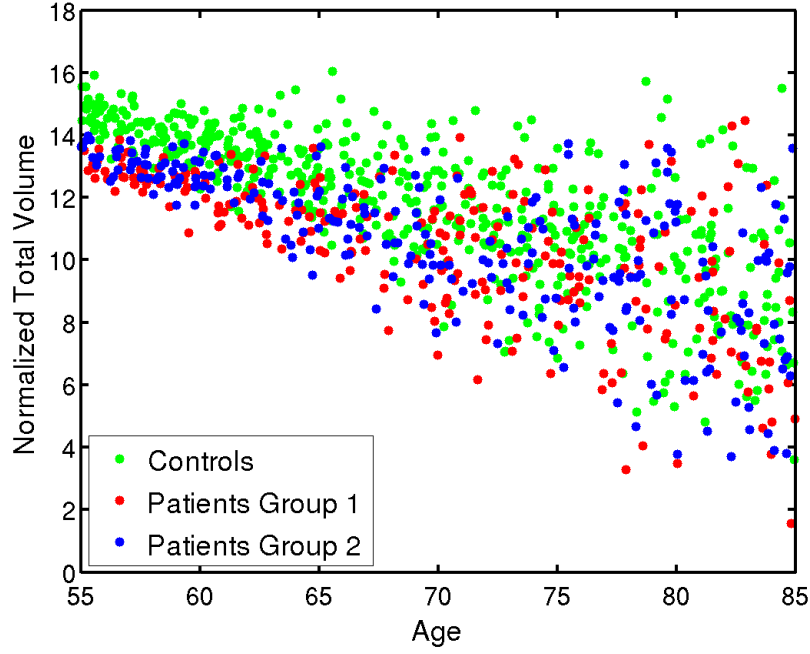


Figure 2.3: Simulated age effect on the normalized total volume. As age increases, the total volume linearly decreases and the variance of the ROI volumes increases.

tion $\mathcal{N}(0.01(t - 55), 0.005(t - 55))$, where t is the age. This simulation corresponds to a linear volume decrease with age (slope) equal to 0.01 per year; and a variance increase of slope equal to 0.005 per year.

3. The samples were randomly separated into two 500-sample groups, corresponding to a control group and a patient group. The patient group was further divided into two sub-groups of 250 samples. In each patient group, we introduced an atrophy pattern induced by a 15% decrease in volume in pre-selected regions. Some of the regions selected were common across the subgroups, while some others were distinct. This was done to simulate the effect of two distinct, but overlapping, variants of a same neurodegenerative disease. The two atrophy patterns are shown in Figure 2.2.

4. The ROIs volumes were then normalized independently, by scaling them between 0 (the most atrophied sample ROI volume) and 1 (the largest sample ROI volume).

The simulated data with heteroscedastic age effect is plotted in Figure 2.3. For both groups, the normalized total volume decreases as age increases. The patient group has smaller total volume due to the disease effect. However as the variance increases, the disease effect is overwhelmed by the age effect.

We compared our model with K-means [97] clustering and Ward hierarchical clustering [160]. However, standard clustering methods do not have access to the information of control group as CHIMERA does. For a fair comparison, we considered therefore two supplementary variants of these clustering methods. Similar to pattern-based morphometry [53], we computed a “profile” for each patient subject. That is, we computed the difference vector between each patient point and its nearest neighbor in the control group according to the Euclidean distance between features. These profiles were clustered instead of the original patient data. In these analysis, a general linear regression (GLM) [100] was performed on the imaging features in order to remove the age effects prior to the clustering. The three variants of our method were applied to the synthetic data. We set model parameters as follows, CHIMERA-affine: $(\lambda_1, \lambda_2) = (10, 100)$; CHIMERA-duo: $(\lambda_1, \lambda_2) = (10, 10)$; and CHIMERA-trans: $\lambda_1 = 10$.

The simulation was repeated 100 times independently. All the methods were applied to each simulated data set, with $K = 2$. The Dice score [35] of overlap between the ground truth and the clustering labels was generated for each run, and the box plots for different methods are presented in Figure 2.4. Given that the dice score is 0.5 when the labels are assigned randomly, our method performs better than clustering methods and their profile-

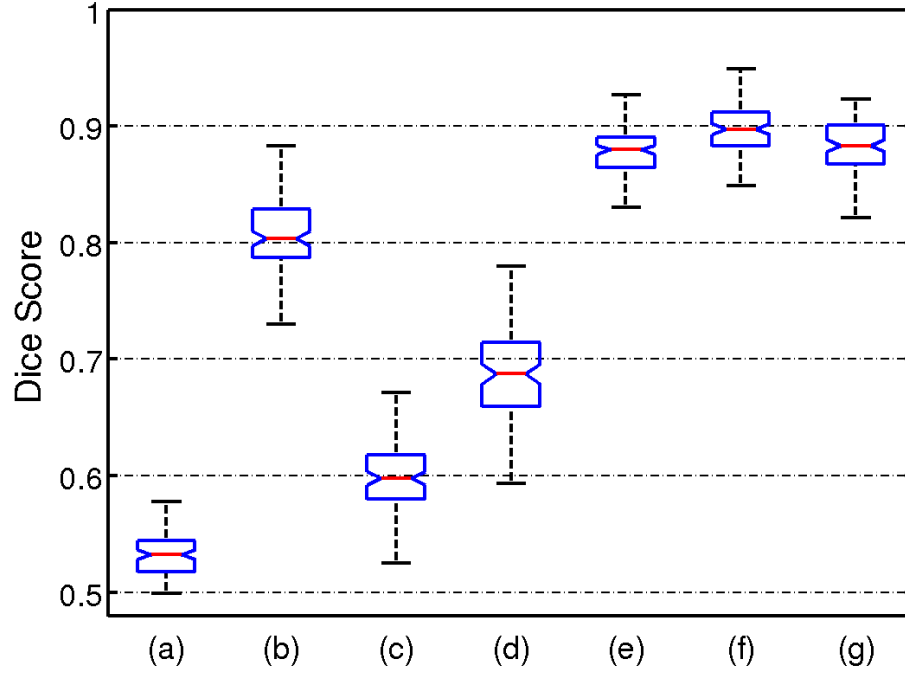


Figure 2.4: Box plot of dice scores on synthetic data between ground truth labels and outputs of clustering methods: (a) K-means, (b) K-means with profile, (c) Hierarchical clustering, (d) Hierarchical clustering with profile, (e) CHIMERA-affine, (f) CHIMERA-duo, and (g) CHIMERA-trans.

based variations. CHIMERA-duo outperformed the other CHIMERA variants. This result indicates that CHIMERA-duo model contains enough degrees of freedom for capturing the differences between patient and control groups, which cannot be expressed as a pure translation. At the same time, the model is much smaller than the affine model, which is hard to regularize.

2.3.2 Neurodegenerative disease data

Before using our method for exploring unknown heterogeneous imaging patterns, we validated our approach on a dementia dataset containing patients suffering from different dis-

eases generating distinct imaging patterns. We used a dementia clinical dataset of 317 T1 structural MRI scans corresponding to 148 Alzheimer’s Disease (AD) patients, 91 Parkinson’s Disease (PD) patients and 78 Normal Controls (NC). The images were skull-stripped [37], co-registered [118] and multi-atlas ROIs were generated [38], as described in Chapter 1. We listed the names of regions used in Appendix Table B.1. The volumes of 80 ROIs were calculated, as well as the volume of brain lesions present in the data [92]. The age and gender of each subject were utilized as covariate features.

The performances of the seven methods described in section 2.3.1 were estimated by performing one hundred 10-folds cross-validations on the dataset. For each cross-validation, the patient samples were partitioned randomly into ten folds. For each fold, the clustering was first established by using normal control samples and the remaining 90% patients. The 10% test samples of the fold were then assigned clustering labels. For K-means and Hierarchical clustering, the assignment was based on the distance to cluster centers. For our approach, the assignment procedure is explained in section 2.2.4. After this assignment, the dice score between the known subtype labels and the labels produced by the clustering methods was computed for the samples of the fold. A dice score for the entire cross-validation was obtained by averaging the dice scores obtained for the ten folds. Running the cross-validation one hundred times with different partitions of the patient data produced the distribution of dice scores shown in Figure 2.5. There is a significant performance gap between our approach and standard clustering methods. CHIMERA-duo and CHIMERA-trans worked comparably well, while the performance of the CHIMERA-affine model were a little lower.

This experiment confirms that our approach can identify distinct imaging patterns cor-

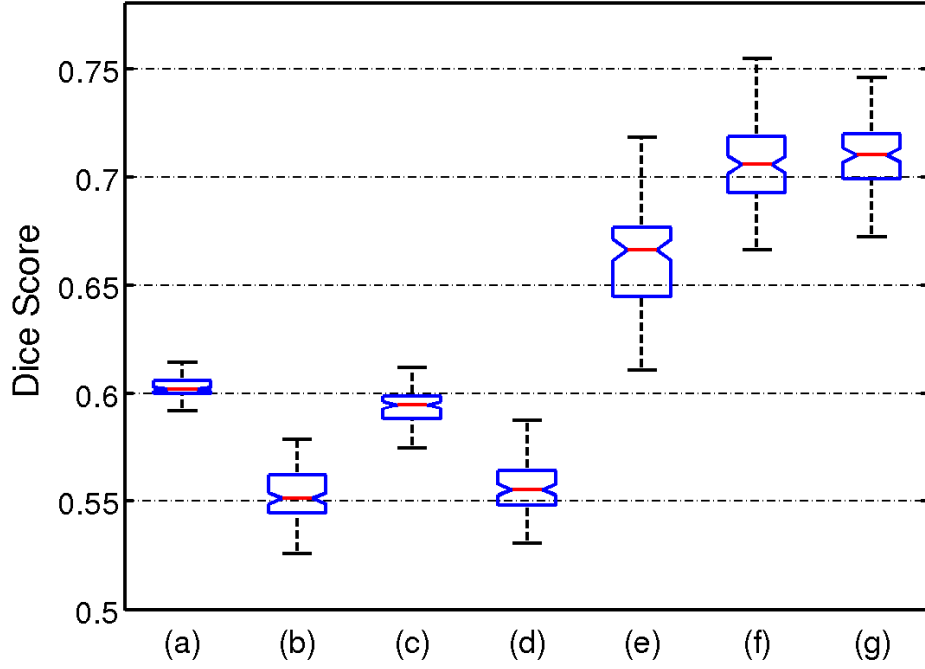


Figure 2.5: Box plot of dice scores on dementia dataset between ground truth labels and outputs of clustering methods: (a) K-means, (b) K-means with profile, (c) Hierarchical clustering, (d) Hierarchical clustering with profile, (e) CHIMERA-affine, (f) CHIMERA-duo, and (g) CHIMERA-trans.

responding to clinically heterogeneous populations using real imaging data. Thus, we used CHIMERA to investigate the existence of disease subtypes in Alzheimer’s disease. The results of our analysis are presented in Chapter 3.

2.4 Conclusion and discussion

CHIMERA is a novel generative clustering framework, which is capable of for identifying disease subtypes of heterogeneous diseases. CHIMERA relies on a point distribution mapping, while taking into account the influence of nuisance covariates, thus overcoming several methodological limitations of existing methods for the analysis of disease hetero-

geneity. We discuss here three main aspects that have not been presented in detail in the previous sections. We also discuss a way to address the main limitation of our current framework.

First, the soft assignment performed by our model provides a rich information about the pathology. Each normal/control point is transformed with a probability distribution ζ by all possible transformations. This notion implies that a healthy subject might make a transition to a diseased state via various pathological patterns/processes. The clustering of patients is based on the posterior probability q and ζ . Instead of a hard assignment for clustering outputs, our approach produces a probability-based soft assignment, which might better describe the disease effects.

Second, the framework is modular. In this work, we have used a linear transformation with scaling and translation that has $\mathcal{O}(D)$ degrees of freedom. Since the sample sizes of most neuroimaging studies are relatively small, we might improve the performance of the model by choosing a more constrained transformation. For instance, the transformation could be represented by the displacement of a few reference samples [111]. Such a transformation would exhibit much fewer degrees of freedom, which could further improve the robustness of the optimization/clustering. Hierarchical transformations could also be implemented, similarly to [118], for reducing the computational burden and/or better constraining the transformation.

Thirdly, we integrate the covariate features in a multi-kernel way. Our framework does not make any explicit assumption on the effect of covariates where GLM on the contrary assumes that the covariates have a linear relationship with the imaging features. With this strategy, our framework mitigates the effect of covariates non-linearly and softly, rather

than the common approach of stratifying and matching populations based on covariates before group analyses.

The large dimension of the transformations involved in our current framework constitutes its main limitation. The optimization instability induced was partially addressed by penalizing the transformations. However, this approach would not be suitable for high dimensional data, such as voxel-level image maps [58], or voxel-wise transformations. The use of sparser transformations, as explained above, will help reduce the dimension of our model. Stricter penalties, such as ℓ_{21} and ℓ_1 penalties, can be investigated in the future. However, we think that dimensionality reduction will probably remain necessary, in order to maintain the stability of the optimization and reduce the number of local optima. Another limitation of our current linear transformation formulation is that it does not take into account the covariance structure of the data, such as covariation between left and right side of the brain. Though we got symmetric results in the experiments, it might be beneficial to introduce this constraint into the framework. Lastly, the Euclidean distance adopted in the framework implicitly treats features with the same weight. This limitation could be addressed by using the Mahalanobis distance instead. These aspects will be further investigated in the future.

Chapter 3

Capturing Heterogeneity in Prodromal Alzheimer's Disease

3.1 Introduction

Alzheimer's disease (AD) presently affects over 5.5 million individuals in the United States and is projected to affect 13.8 million of people of age 65 and older by the mid-century [9, 65]. It poses a substantial healthcare, financial, and caregiver burden [69] to the community and healthcare system. Importantly, Alzheimer's disease is the most common cause of dementia, where several coincident pathologies, mainly vascular lesions, limbic TAR-DNA binding protein-43 (TDP-43) and Lewy body related pathology, also account for the cognitive symptoms of dementia subjects [131, 150, 153, 154, 165, 168]. The presence of different brain pathologic lesions and the variations in cognitive reserve, genetic background, and environmental exposures, can lead to differential vulnerability. This, in turn, can lead to large variations in the clinical presentation and course of the subjects. Therefore, it can

be expected that a group of cognitively impaired subjects is composed of different subtypes. Each subtype would present a specific disease course and characteristics. While AD is not treatable, an accurate identification of the disease in its early stage could be proved crucial towards leading to more effective therapeutic interventions. Towards this end, research into biomarkers that can precisely quantify the subtle and complex structural and functional changes that are induced in the brain during the early stages of AD is of particular interest and importance. Many studies have taken part in developing tools or applying established methodologies that use neuroimaging to improve diagnosis of AD.

Despite the increasing evidence of population heterogeneity [90, 110] and the potential benefits from accurately characterizing it, most of the imaging analysis approaches ignore phenotypic heterogeneity and define patterns of structural or functional changes based on clinical categorical definitions and summarizing them with a single imaging pattern. On the one hand, mass univariate tools such as voxel based morphometry and its variants [6, 10, 18, 81, 82, 169] are adopted in quantifying the differences between AD patients and normal control populations. On the other hand, multivariate pattern analysis [26, 42, 84, 106, 156] seeks to improve the specificity and sensitivity of computer-aided diagnosis by encoding relations across multiple variables within a discriminative imaging pattern. However, these imaging patterns are either incomplete or worst misleading, in the presence of heterogeneity.

Characterizing the heterogeneity of Alzheimer's disease using neuroimaging would allow elucidating the underlying neuropathological processes. By identifying patients with distinct imaging signatures, the disease diagnosis would be improved through precision diagnostic, which would further enhance therapeutic innovation in clinical trials through

appropriate patient recruitment.

Recognizing the limitations of the case-control setting, previous efforts have focused on using *a priori* defined neuropathological categories to identify subgroups of patients [77, 90, 110, 139]. However, such neuropathological or clinical categories may not be reliable enough for accurately distinguishing disease subtypes [93, 102]. Importantly, these approaches rely on a clinical “intuition”, thus being biased and prone to human error. Unbiased data-driven approaches show promise to be able to identify different patterns of pathology distribution based on the atrophy patterns inherent to the population [114, 115, 154]. However, commonly used standard clustering methods tend to group patients along the direction associated with the largest data variability, which may not be induced by the pathology, and which might reflect effects such as age, gender or disease stage. To avoid this, we should be steering the clustering algorithm to focus on the neurodegeneration patterns that drive cognitive impairment. Such a clustering is more likely to lead to grouping patients into relatively homogeneous groups, with potentially more predictable clinical outcomes and treatment responses.

In this chapter, we investigated the heterogeneity of neurodegeneration in mild cognitive impairment (MCI) and AD relative to cognitively normal (CN) individuals, by leveraging the methodology described in Chapter 2, and capitalizing on the large datasets of MCI and AD patients from ADNI (ADNI-1 and ADNI-GO/2). This work can complement and contribute to recently proposed dimensional approaches, such as the A/T/N (amyloid- β /tau/neurodegeneration) AD staging system [71], by evaluating and categorizing heterogeneity of disease-related neurodegeneration patterns in a data-driven way. Four distinct neuroanatomical subtypes were found in MCI and AD dementia relative to

CN, where the patients across ADNI-1 and ADNI-GO/2 cohorts showed consistent neurodegenerative signatures. Subtypes in this context are mainly meant to define the main dimensions of the heterogeneity of AD, rather than imply distinct imaging phenotypes. To investigate that, we studied in detail the cerebrospinal fluid (CSF) biomarkers, cognitive characteristics and white matter hyperintensity (WMH) volumes of these subtypes.

3.2 Materials and methods

3.2.1 Subjects

A total number of 1243 AD Neuroimaging Initiative (ADNI)¹ participants were included in the study, including 760 ADNI-1 subjects (213 CN, 370 late MCI (LMCI), and 177 AD subjects) and 483 ADNI-GO/2 subjects (186 CN, 160 LMCI, and 137 AD). Early MCI subjects from the ADNI-GO/2 were excluded because this group was only recruited in ADNI-GO/2. CN subjects included subjects with normal cognition, independently of the presence of memory complaints. Diagnoses of MCI and AD were established as described in [103, 121, 120]. The data for this study was downloaded in December 2015. The ADNI datasets have been extensively reviewed in [161]. To evaluate differences in cognitive performance, we studied the previously developed memory composite score [24], the executive composite score [56], and the Boston naming test scores. Median follow-up length for ADNI-1 and ADNI-GO/2 MCI subjects was 161.0 (1st quartile: 105.4 - 3rd quartile: 315.0) and 156.3 (1st quartile: 106.5 - 3rd quartile: 159.1) weeks, respectively.

¹<http://www.adni-info.org>

3.2.2 Cerebrospinal fluid (CSF) collection and measurement

CSF samples were processed as previously described [136, 137]. $A\beta_{1-42}$ and total tau (t-tau) were measured using the multiplex xMAP Luminex platform (Luminex Corp, Austin, TX) with Innogenetics (INNO-BIA AlzBio3; Ghent, Belgium; for research use-only reagents) immunoassay kitbased reagents.

3.2.3 MRI acquisition and processing

Acquisition of 1.5-T MRI (for ADNI-1) and 3.0-T MRI (for ADNI-GO/2) data at each study site followed a previously described standardized protocol that included volumetric 3D MP-RAGE (magnetization-prepared rapid gradient-echo imaging [109]) or Sagittal MP-RAGE with variable resolution around the target of 1.2mm isotropically. The scans went through the following correction methods: gradwarp, B1 calibration, N3 correction [141], and (in-house) skull-stripping [37]. See (www.loni.usc.edu/ADNI) and [72] for details.

For the heterogeneity analysis, T1 structural MRI images of all subjects were segmented into 80 anatomical regions of interest (ROIs). The ROI labels were obtained using the method described in Chapter 1. The names of the ROIs used are listed in Appendix Table B.1. To assess longitudinal neuroimaging changes, another set of surface ROI volumes provided by ADNI were used, which were processed using the FreeSurfer² software package version 4.4 [125, 126]. For visualizing disease patterns, regional tissue volumetric maps, termed RAVENS [29] were generated. Further details regarding image processing can be found in Chapter 1. We used the CN group data for linearly regressing age and gender from the RAVENS maps of the entire dataset. Adjusted hippocampal volume (aHV) was

²<http://surfer.nmr.mgh.harvard.edu/>

calculated and matched across ADNI-1 and ADNI-GO/2 cohorts using a set of matched MRIs as previously described in [155]. All the subjects were then divided into four quartiles in order to perform the stratification analysis based on hippocampal atrophy, which is considered to be a sensitive biomarker of dementia.

3.2.4 White matter hyperintensities (WMH)

WMH were segmented using different approaches in ADNI-1 [134] and ADNI-GO/2 [32]. The method applied on ADNI-1 utilized PD, T1, and T2 MR images. This method is based on a Bayesian Markov random field approach, where the joint posterior probability of the presence of WMH at each voxel is maximized. The posterior probability consists of a likelihood computed from image intensities, a spatial prior that regularizes the location of WMHs, and a contextual prior that encourages neighbor voxels to have the same labels. The method applied on ADNI-GO/2 utilized FLAIR and T1 images. This method operates first by co-registering the FLAIR MR image to the T1 image, and then performing inhomogeneity correction. The binary WMH mask is then estimated based on histogram fitting and thresholding at 3.5 standard deviations above the mean signal in brain matter distribution. The WMH mask is further refined by taking into account spatial prior and tissue class constraints in a Bayesian approach.

3.2.5 Heterogeneity and voxel based morphometry analysis

In this study, ADNI-1 and ADNI-GO/2 datasets were combined in order to delineate more consistent pathological imaging patterns. However, the structural MRI scans from these two cohorts have many discrepancies, such as different scan protocols and magnetic field

strengths. We took these discrepancies into account during our analyses by introducing the original recruitment cohort (ADNI-1 versus ADNI-GO/2) as a covariate in our model, in addition to age and gender. As a result, the patient and normal control distributions were matched within each cohort separately, but the pathological effects captured by CHIMERA were shared across datasets. We performed a 10-fold cross-validation using the combined dataset to evaluate the robustness of the method, which showed an 84.1% agreement. In addition, we applied our clustering approach separately in the ADNI-1 and ADNI-GO/2 cohorts, which showed a 63% and 74% overall agreement with the combined approach, respectively.

3.2.6 Statistical analysis

The demographic variables, *APOE* ϵ 4 genotype (apolipoprotein E type 4 [21]), CSF biomarker levels, cognitive test scores and WMH volumes were compared across clusters. For categorical variables, the Fisher exact test was used to identify differences between groups. In case of significant differences, the conducted pairwise comparisons between clusters were adjusted for multiple comparisons using the Bonferroni procedure. For quantitative demographic variables (e.g. age) an ANOVA analysis was conducted, followed by a Dunn test. For the clinical and CSF biomarker measures, an ANCOVA analysis was performed, which included age, gender, years of education and *APOE* ϵ 4 genotype as covariates. In case of significant differences, the Dunn's post-hoc test was performed. When the ANCOVA requirements were not met, a Box-Cox transformation was applied to the data. A longitudinal neuropsychological analysis was conducted for LMCI patients (AD subjects were excluded due to short follow-up) using mixed effect models that included subjects

and time as random effects and age, gender, time, $APOE\epsilon 4$ presence and years of education as fixed effects. A Cox hazards model including age, gender, $APOE\epsilon 4$ presence and years of education as covariates, was fitted for comparing the conversion of LMCI patient to AD in the different clusters. For the evaluation of the profile of longitudinal changes in MRI volumes, individual mixed effects models that included age, gender, time and $APOE\epsilon 4$ as covariates, were applied to estimate the yearly ROI volumetric changes in CN subjects and patients belonging to the different clusters. Baseline and 2nd year MRI scans were compared for this purpose, and ROI values were standardized to compare findings across the different areas. Analyses were performed using R v. 3.2.2 [122]. The visualization of imaging signatures of derived clusters (i.e., the clusters found by CHIMERA), of clinically-defined (AD/MCI/AD+MCI) groups, and of aHV-defined (aHV quantiles) groups was performed via VBM [6, 23] on RAVENS maps.

3.3 Experiments and results

3.3.1 Cluster demographic and genetic characteristics

We adopted a cross-validation strategy to find the optimal hyperparameters (i.e. number of clusters and regularization coefficient) of CHIMERA [36]. For each combination of hyper-parameters, 100 runs of leave-10%-out clusterings were performed. During each clustering, a random subset of 90% of the patient samples and all the normal control samples were used for generating the transformations and defining the patient clusters. The remaining 10% patient samples were assigned to one of the estimated clusters, based on their proximity with the transformed controls (Equation 2.24). We measured the Adjusted

Rand Index (ARI), which indicates the reproducibility of clustering memberships, between all the pairs of the 100 clusterings obtained for each hyperparameter set, and averaged the ARI for each clustering. The hyperparameters that yielded the best reproducibility were chosen to produce clustering memberships herein.

We finally partitioned the entire set of ADNI patients into four clusters that included in each case subjects from ADNI-1 and ADNI-GO/2. Subjects in different ADNI cohorts, but within the same cluster, exhibited similar atrophy patterns. The characteristics of clusters identified in ADNI-1 and ADNI-GO/2 cohorts are summarized in Table 3.1. In all ADNI cohorts, Cluster 2 subjects were older and had a greater proportion of AD dementia subjects compared to Cluster 1.

3.3.2 Cluster membership confidence

In our main analysis, we assigned each subject to the cluster with the highest probability. For most of the subjects, cluster membership was assigned with a probability ≥ 0.5 . However, in the remaining cases, membership was assigned with a probability < 0.5 . The “tightest” cluster was Cluster 2 (87% subjects had a probability ≥ 0.5), whereas Cluster 3 was the loosest one (66% subjects had a probability ≥ 0.5) (Figure 3.1), with most of the loose cases being close to Cluster 1. We summarize these findings using a Venn diagram in Figure 3.2.

3.3.3 Cross-sectional clinical and biomarker associations

CSF $A\beta_{1-42}$ defined groups, WMH volume and the studied cognitive measures differed between the four clusters (Table 3.1). Post-hoc group comparisons identified similar results

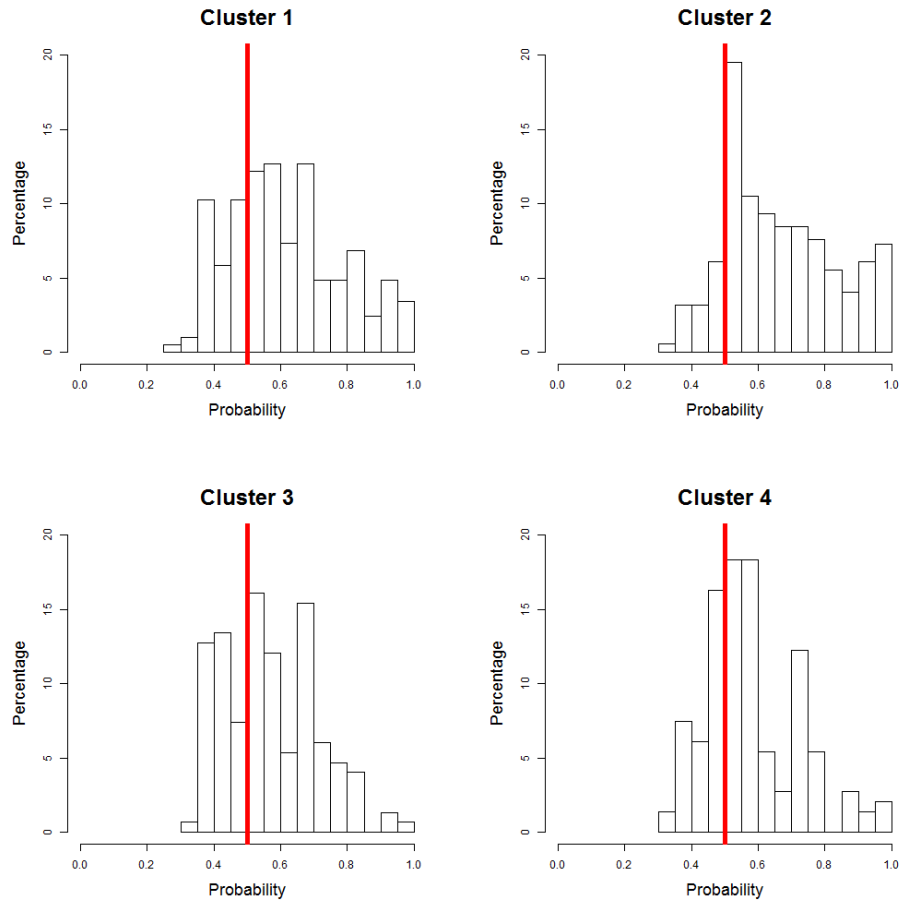


Figure 3.1: Cluster probabilities for subjects classified within each cluster. The vertical red line indicates a 0.5 probability of belonging to the cluster.

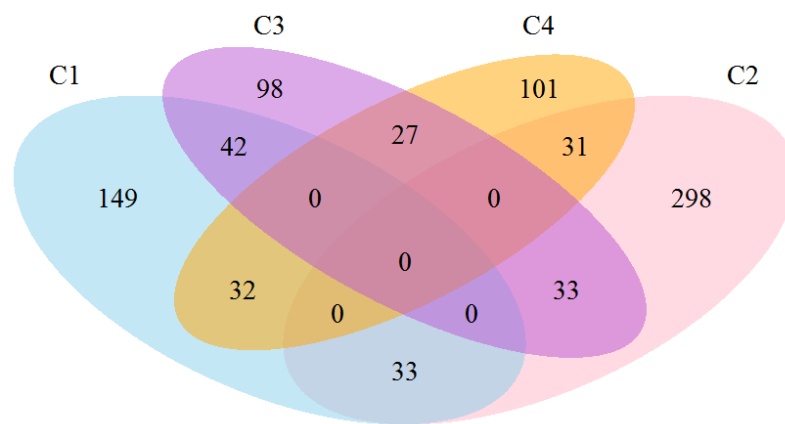


Figure 3.2: Venn diagram depicting the number of subjects classified tightly or loosely into each cluster. Subjects with a probability > 0.5 were included within a single cluster, whereas subjects with a highest cluster probability < 0.5 are depicted in the interphase of the two top clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	P-value ^c	C1 vs C2	C1 vs C3	C1 vs C4
Demographics								
ADNI-1								
Number of Subjects (%)	166 (30.35%)	213 (38.94%)	65 (11.88%)	103 (18.83%)				
AD (%)	34 (20.48%)	99 (46.48%)	19 (29.23%)	25 (24.27%)	<0.001	<0.0001	0.21	0.56
LMCI (%)	132 (79.52%)	114 (53.52%)	46 (70.77%)	78 (75.73%)				
Age (Median ^a)	75.25 (69.6-79.95)	77.3 (72.81-7)	71.6 (68-76.4)	72.7 (68.55-79.9)	<0.0001	0.021	0.006	0.51
Gender (Female%)	33.13	39.91	46.15	45.63	0.13			
APOE(% ϵ 4 presence)	52.41	63.85	55.38	58.25	0.15			
ADNI-GO/2								
Number of Subjects (%)	38 (12.79%)	131 (44.11%)	84 (28.28%)	44 (14.81%)				
AD (%)	9 (23.68%)	73 (55.73%)	38 (45.24%)	17 (38.64%)	0.004	<0.001	0.039	0.23
LMCI (%)	29 (76.32%)	58 (44.27%)	46 (54.76%)	27 (61.36%)				
Age (Median ^a)	69.9 (62-74.6)	75.9 (72.15-79.9)	71.45 (64.33-77.23)	72.85 (66.28-76.43)	<0.0001	<0.0001	0.31	0.30
Gender (Female%)	42.11	37.4	61.9	36.36	0.003	0.74	0.07	0.76
APOE(% ϵ 4 presence)	57.89	62.6	60.71	68.18	0.79			
CSF biomarkers and cognitive scores								
ADNI-1								
A β_{1-42} ^b (% \leq 192pg/mL)	63.3	89.5	96.8	75.9	<0.0001	0.0001	0.0007	0.40
T-tau	90.5 (65.5-125.2)	96 (67.75-135)	109 (86-150)	107.5 (69.75-141.8)	0.64			
P-tau	31 (20.5-45)	35 (24-46.75)	38 (29-47.5)	40 (25-50.5)	<0.0001	<0.0001	0.063	0.80
ADAS-Cog13	18.67 (14.33-22.33)	24.33 (18.67-29.67)	21 (16.17-28.41)	19.67 (16.16-24.67)	<0.0001	<0.0001	0.0025	0.91
Memory	-0.11 (-0.53-0.25)	-0.59 (-0.92-0.11)	-0.44 (-0.88-0.04)	-0.22 (-0.55-0.1)	<0.0001	<0.0001	0.028	0.34
Executive	-0.21 (-0.75-0.49)	-0.51 (-1.09-0.04)	-0.25 (-1.2-0.17)	0.03 (-0.47-0.47)	0.0022	0.0097	0.52	0.66
Boston naming test	25 (22-28)	24 (21-27)	27 (23-28)	26 (22-28)	0.015	0.29	0.95	0.20
WMH	0.25 (0.11-0.71)	0.45 (0.17-1.21)	0.15 (0.06-0.42)	0.32 (0.11-0.69)	0.0022	0.0097	0.52	0.66
ADNI-GO/2								
A β_{1-42} ^b (% \leq 192pg/mL)	68.6	89.3	83.8	71.4	0.007	0.018	0.25	1.0
T-tau	76 (46-109)	108.5 (82-146.5)	105 (67-172)	86.5 (54.25-160.2)	0.0079	0.0031	0.029	0.35
P-tau	37 (32-56.5)	48 (37-63.25)	46 (33-67)	40.5 (23.5-69.5)	<0.0001	<0.0001	0.016	0.28
ADAS-Cog13	17 (12-24)	27 (22-33.5)	21 (17-27.75)	20 (15-29)	0.0013	<0.0001	0.016	0.19
Memory	0.06 (-0.41-0.59)	-0.47 (-0.74-0.06)	-0.15 (-0.55-0.21)	-0.22 (-0.61-0.14)	0.0053	0.011	0.0051	0.54
Executive	-0.03 (-0.32-0.87)	-0.48 (-1.13-0.15)	-0.36 (-1.24-0.3)	-0.01 (-0.68-0.68)	0.017	0.029	0.89	0.84
Boston naming test	27 (24-28)	23 (20-26)	25 (22-27)	26 (23-28)	0.041	0.028	0.12	0.80
WMH	2.31 (1.28-5.36)	6.73 (3.22-14.13)	3.04 (1.66-6.9)	3.9 (2.11-5.69)	0.017	0.029	0.89	0.84

A β = Amyloid- β ; ADAS-Cog = Alzheimer's Disease Assessment Scale - Cognitive Subscale; P-tau = phosphorylated tau; T-tau = total tau.

^a Parenthesis include first and third quartile values.

^b Due to incomplete data, the percentages of subjects with CSF biomarkers collected are 50.8% in ADNI-1 and 94.3% in ADNI-GO/2.

^c CSF Amyloid- β group differences were tested using a Fisher exact test, all other comparisons were performed using an ANCOVA test.

Table 3.1: Cluster demographics and characteristics of CSF biomarkers and cognitive scores.

across cohorts. Subjects in Cluster 2 and 3 included a higher frequency of subjects with pathological CSF $A\beta_{1-42}$ values. Cluster 2 and 3 subjects presented worse performance in the memory composite and in ADAS-Cog (Alzheimer's Disease Assessment Scale - Cognitive Subscale) compared to Cluster 1. In addition, Cluster 2 subjects had worse executive composite, higher p-tau values, and greater WMH volume compared to Cluster 1 subjects. Only in ADNI-GO/2 did the clusters differ in terms of CSF t-tau values (Cluster 1 had lower values than Cluster 2 and 3).

3.3.4 Group-wise VBM results

The group-wise VBM results are presented in Figure 3.3. Cluster 1 showed the least amount and extent of atrophy. In the ADNI-1 cohort, the atrophy was mainly localized in the temporal lobe with additional involvement of the posterior cingulate cortex. However, in ADNI-GO/2 it was statistically indistinguishable from CN. Interestingly, Cluster 4 showed moderate localized atrophy in the hippocampus and the anterior-medial temporal cortex, without any significant involvement beyond these regions. Conversely, Cluster 2 and 3 showed widespread involvement, extending to the temporal, parietal, and occipital lobes in addition to subcortical gray matter. However, they presented markedly different patterns, as Cluster 2 presented relatively most severe temporal atrophy, whereas Cluster 3 exhibited a more diffuse atrophy pattern, with a comparatively less dominant involvement of the temporal lobe compared to Cluster 2. Group comparisons based on baseline diagnosis and aHV quartiles are shown in Figure 3.4 and Figure 3.5, respectively.

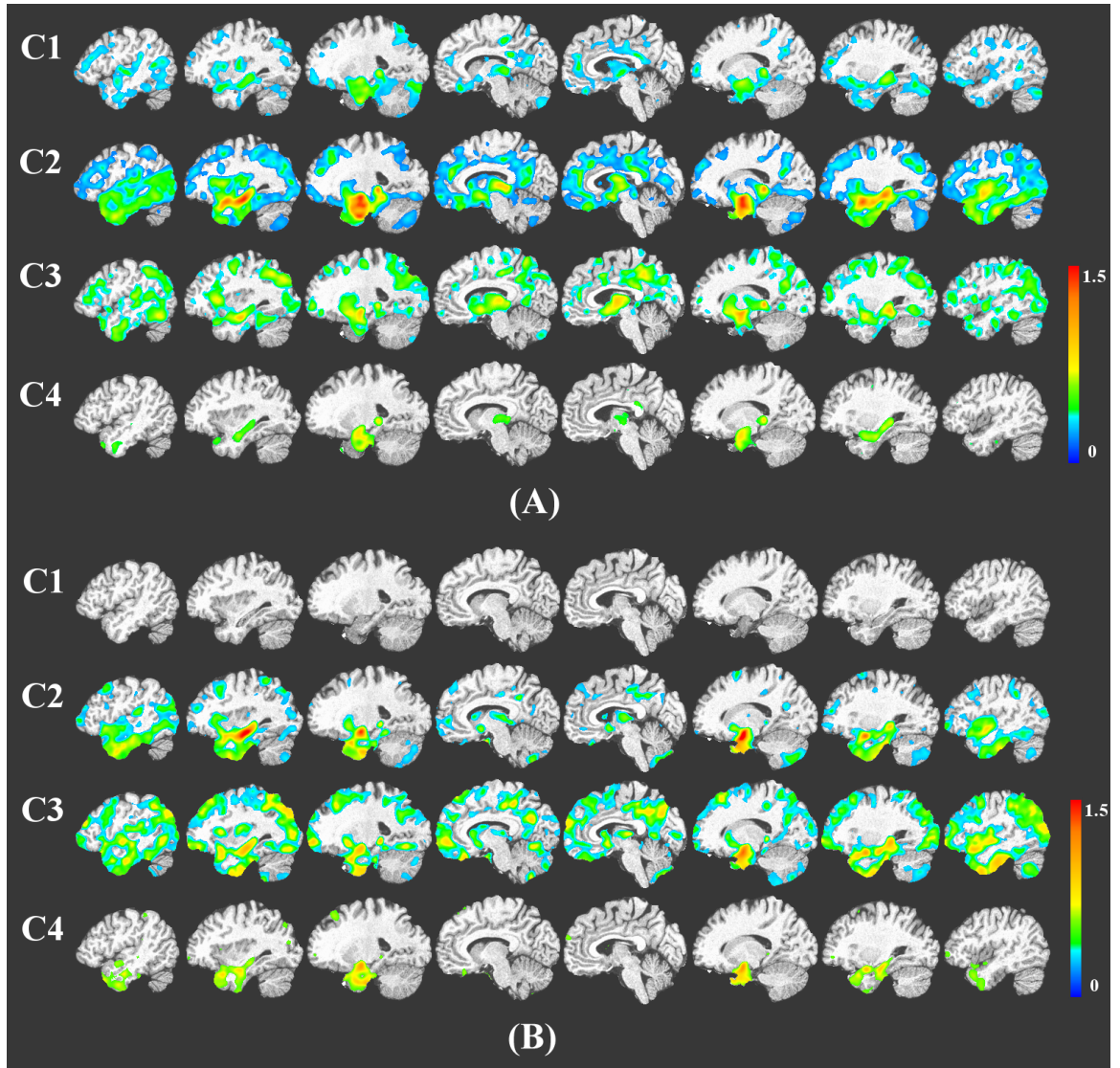


Figure 3.3: VBM between the identified clusters and the CN reference group for the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05.

3.3.5 Longitudinal changes

Cluster 2 and 3 showed a faster progression from MCI to AD, and a steeper cognitive decline in the studied cognitive measures compared to Cluster 1 (Table 3.2, Figure 3.6(A) and

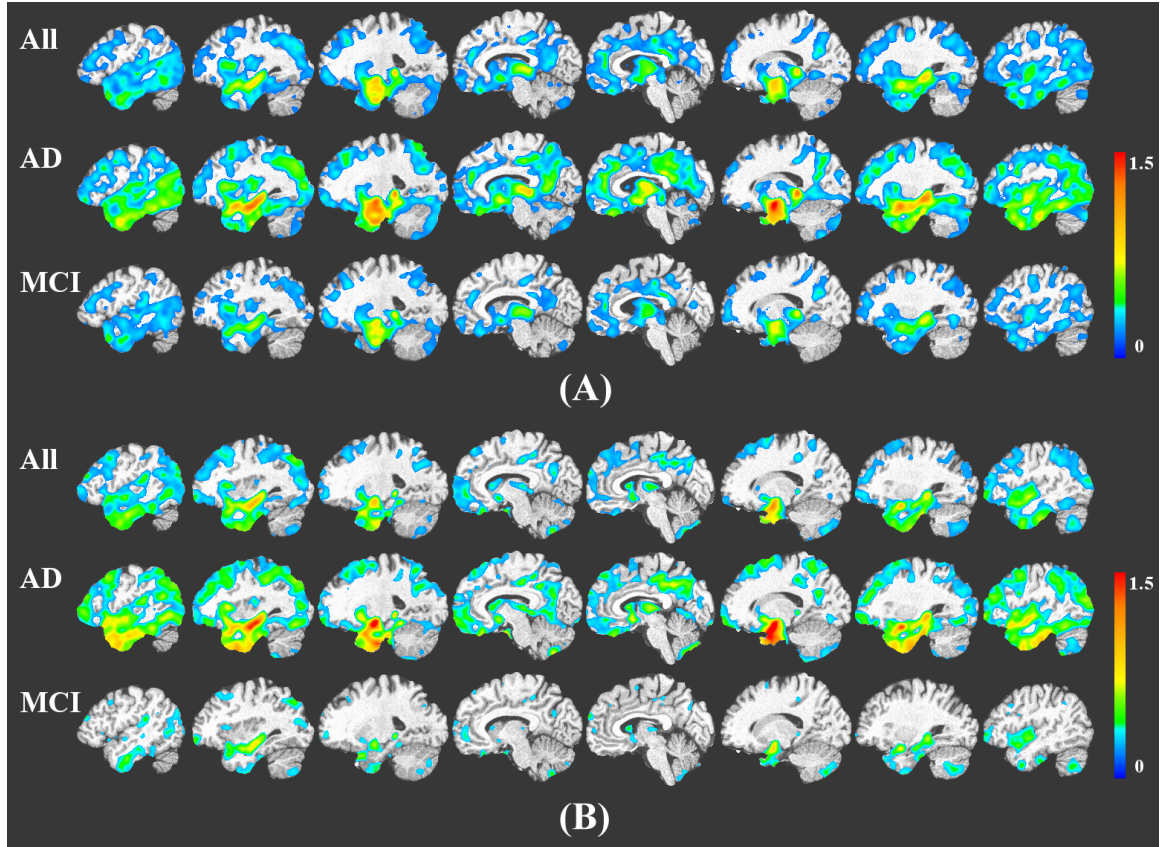


Figure 3.4: VBM between clinical groups (MCI and AD) and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05.

Figure 3.7). Similarly, Cluster 2 and 3 showed a more pronounced longitudinal cognitive decline than Cluster 1 (Figure 3.6(A)). Interestingly, whereas the rate of change of the executive profile was similar in Cluster 2 and 3 (p-value=0.75), Cluster 3 showed a lower rate of memory decline compared to Cluster 2 (p-value=0.039). In none of the analyses did Cluster 4 differ from Cluster 1. Tabel 3.3 and Figure 3.6(B) summarize the results for aHV quartiles.

Longitudinal MRI changes in the different ROIs and groups are summarized in Figure

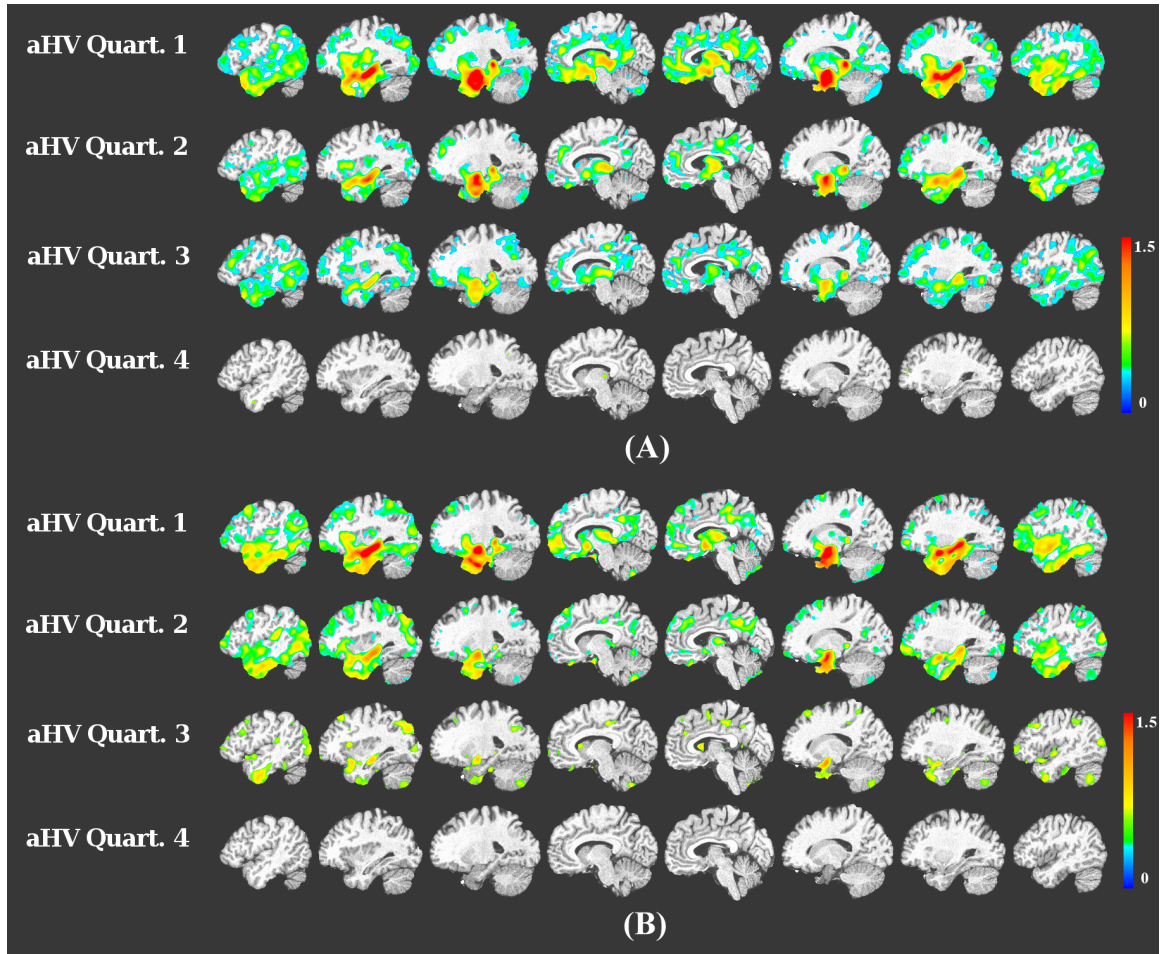


Figure 3.5: VBM between patient clusters, stratified by adjusted hippocampal volumes, and CN reference group in the ADNI-1 (A) and ADNI-GO/2 cohorts (B). Color scale represents the effect size of gray matter RAVENS maps of each comparison between a cluster and CN individuals. Red indicates greater atrophy (lower volume). Effect size maps are thresholded at false discovery rate (FDR) adjusted p-value of 0.05. Quartile 1 represents the lowest volume, whereas Quartile 4 is the highest.

3.8. In this figure, it can be appreciated that the Cluster 1 shows a similar pattern of atrophy as the CN group in all areas except the temporal lobe, where the atrophy rate is more pronounced for Cluster 1. Cluster 2 and 3 showed the fastest rates of atrophy, with the former showing a faster temporal atrophy, whereas in the latter case the preponderance was frontal. Finally, Cluster 4 showed an intermediate pattern.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
MCI to AD ^a	Ref.	2.26 (<0.0001)	1.87 (0.0024)	1.27 (0.21)
ADAS-Cog13	Ref.	0.20 (<0.0001)	0.09 (0.023)	0.04 (0.31)
Memory ^b	Ref.	-0.11 (<0.0001)	-0.06 (0.030)	0.004 (0.86)
Executive ^b	Ref.	-0.12 (<0.0001)	-0.11 (0.0005)	-0.04 (0.17)

Only late MCI subjects were included due to short Alzheimer's disease (AD) subjects follow-up.

Age, gender, education and *APOE* were included as covariates.

^a Hazard ratio (P-value).

^b Regression coefficient (P-value).

Table 3.2: Longitudinal neuropsychological associations of the clusters.

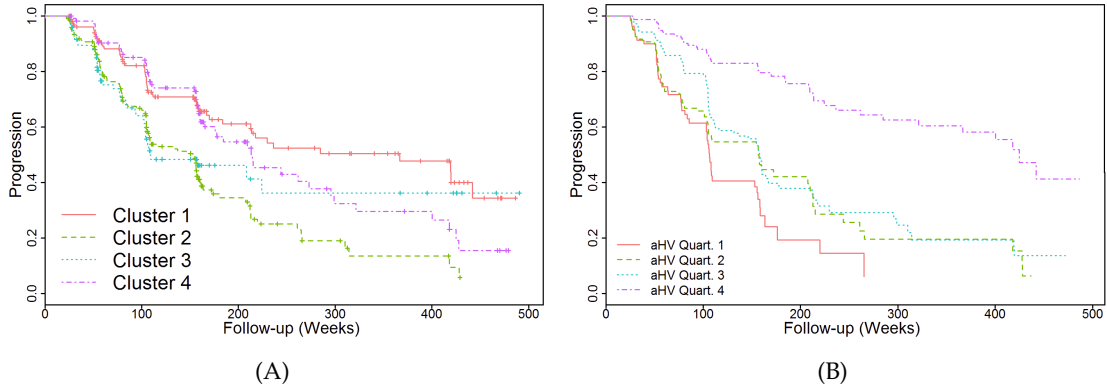


Figure 3.6: (A) Progression from MCI to AD stratified by MRI-defined clusters. (B) Progression from MCI to AD based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.

3.3.6 Prevalence of clusters as a function of age

We also investigated the prevalence of each of the 4 clusters, as a function of age. Figure 3.9 shows the number of subjects in each cluster, averaged over a 5-year bracket around each age, as well as the clusters relative frequency (which takes into account the variable number of subjects per age group) as a function of age. Curves were fitted using cubic splines with 3 control points. Taking into consideration potential boundary effects in these fits (small number of subjects on either end of the age spectrum), Figure 3.9 generally shows that the proportions of Cluster 2 and 3 increase and decrease, respectively, steadily.

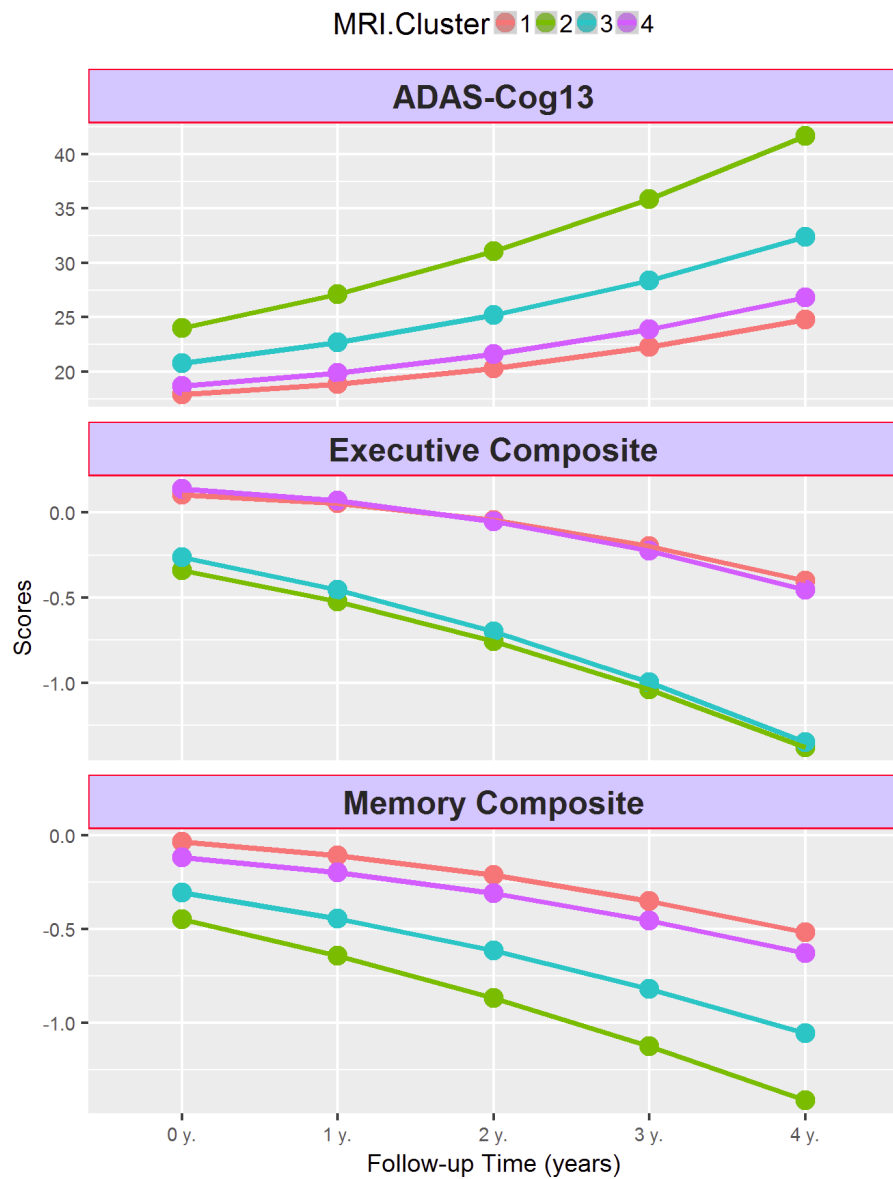


Figure 3.7: Longitudinal cognitive changes in ADAS-Cog13, memory and executive composite scores in MCI subjects stratified by MRI-defined clusters.

Cluster 4 is relatively stable throughout this age range, whereas the proportion of Cluster 1 seems to decrease steadily after the age of 63.

Quartile 1 as reference				
	Memory Composite		Executive Composite	
	Coefficient	p-value	Coefficient	p-value
Quartile 1	Ref.	Ref.	Ref.	Ref.
Quartile 2	-0.130	<0.0001	-0.122	<0.0001
Quartile 3	-0.147	<0.0001	-0.130	<0.0001
Quartile 4	-0.131	<0.0001	-0.154	<0.0001

Quartile 4 as reference				
	Memory Composite		Executive Composite	
	Coefficient	p-value	Coefficient	p-value
Quartile 1	-0.016	<0.0001	0.154	<0.0001
Quartile 2	0.001	0.97	0.032	0.36
Quartile 3	0.131	0.60	0.024	0.51
Quartile 4	Ref.	Ref.	Ref.	Ref.

Table 3.3: Regression coefficients and p-values of studying longitudinal associations of cognitive measures with aHV quartiles (Quartile 1 corresponds to the lowest volume, whereas Quartile 4 is the highest).

3.4 Discussion

We evaluated the heterogeneity of neurodegeneration present in MCI and AD, relative to CN individuals, by applying the proposed pattern analysis and machine learning method, which was described in Chapter 2, on data from ADNI. We found four distinct groups that best summarized this neuroanatomical heterogeneity. Besides having markedly different atrophy patterns, these groups also differed in the frequency of AD-like CSF $A\beta_{1-42}$ and tau levels, as well as in their clinical profiles (Table 3.4). In particular, Cluster 1 included individuals with generally normal anatomy, the lowest frequency of subjects with abnormal CSF $A\beta_{1-42}$ levels, normal CSF-tau levels, least baseline cognitive impairment and slowest rates of cognitive decline. Conversely, Cluster 2 was consistent with the typical AD-like neuroanatomical patterns and high frequency of AD-like CSF $A\beta_{1-42}$ levels, and fastest

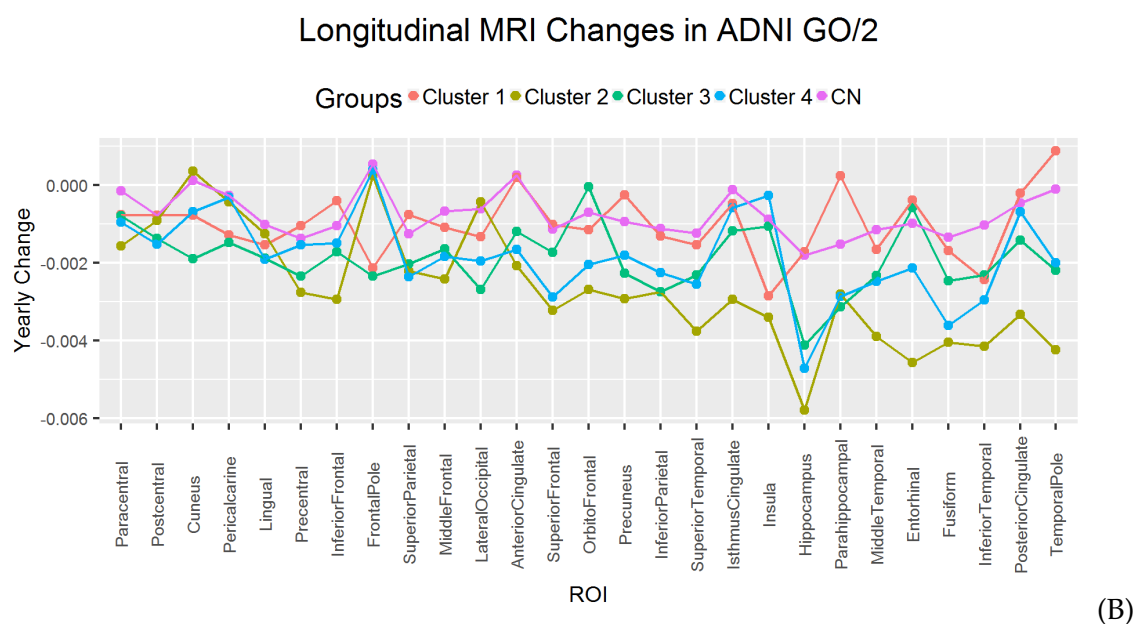
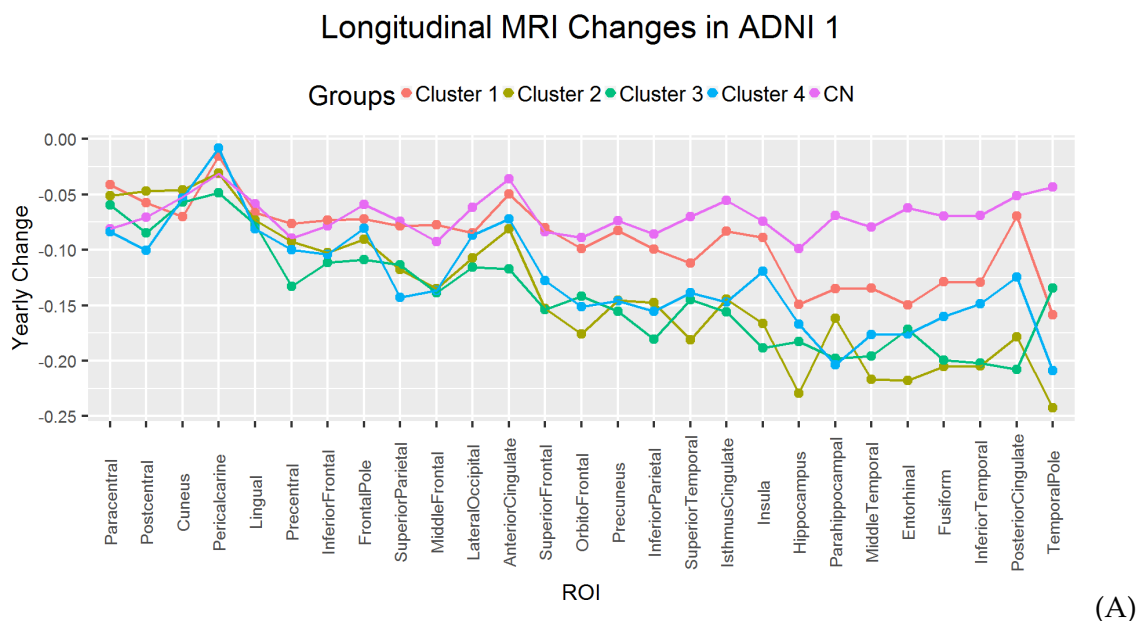


Figure 3.8: Standardized yearly MRI changes observed in CN subjects and MCI subjects belonging to the four identified clusters. A total of 124 CN, 57 Cluster 1, 44 Cluster 2, 18 Cluster 3 and 40 Cluster 4 subjects were included in the analysis in ADNI-1 (A). 84 CN, 15 Cluster 1, 17 Cluster 2, 17 Cluster 3 and 13 Cluster 4 subjects were included in the analysis in ADNI-GO/2 (B).

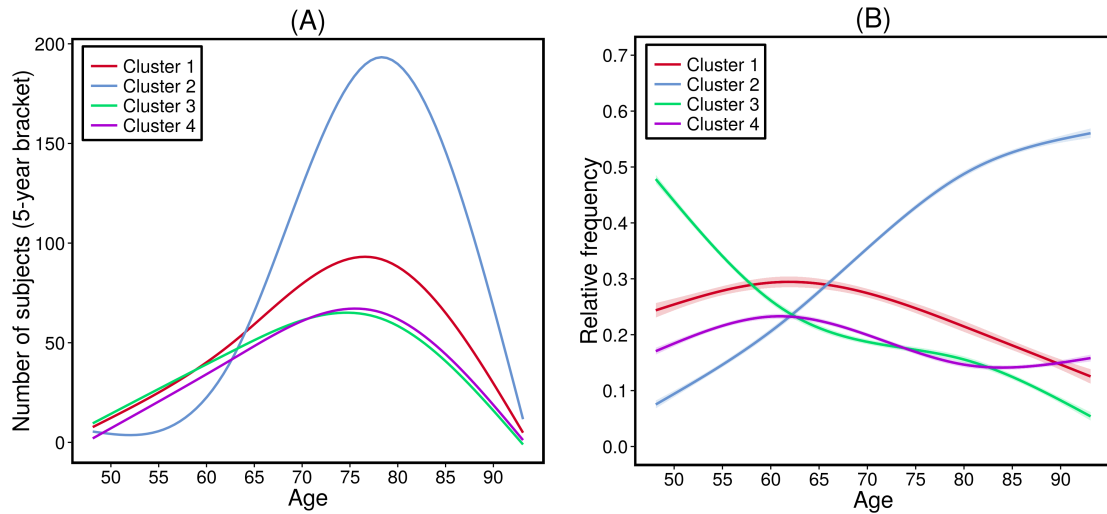


Figure 3.9: Prevalence of clusters as a function of age. (A) number of subjects with 5-year brackets, (B) relative frequency of clusters, fitted with cubic splines.

cognitive decline affecting executive and memory cognitive domains. Cluster 3 showed greater cortical atrophy in parietal and dorsolateral frontal cortex with proportionately lesser involvement of the limbic cortex, compared to Cluster 2. Although Cluster 3 was associated with fast cognitive decline, this decline was more marked for the executive rather than the memory composite score, which is consistent with the imaging findings. Notably, Cluster 3 MCI individuals did not show further progression to AD after four years, although this has to be interpreted cautiously due to the small number of subjects followed for that long a period. Finally, Cluster 4 included individuals with localized atrophy in the hippocampus and medial temporal lobe, although cognitive changes did not differ from the ones observed in Cluster 1.

In the A/T/N AD staging system [71], each dimension has only two status, presence/positive or absence/negative, which are determined by a single dichotomous classification/thresholding. This dichotomous classification may be appropriate for CSF Amyloid-

	Neuroanatomical atrophy pattern	Alzheimer’s disease-like CSF $A\beta_{1-42}$ levels	Cognitive decline
Cluster 1	Mild or none; non-focal	Lowest frequency	Least steep
Cluster 2	Widespread, greater temporal involvement	Higher frequency	Steepest for memory and executive
Cluster 3	Widespread, global	Higher frequency	Steepest for executive, intermediate for memory
Cluster 4	Localized, temporal	Lower frequency	Least steep

Table 3.4: Summary of characteristics of clusters.

β and tau values (“A” and “T” dimensions). These are typically considered as a single linear measure for AD, which is line with previous pathology studies that have described the typical Amyloid- β deposition in the brain and brainstem [146]. However, we consider that the structural imaging neurodegeneration patterns (“N” dimension) can be heterogeneous. Thus, the binarized “N” dimension in [71] might be an oversimplification that does not take advantage of the data richness offered by current neuroimaging approaches. As we describe below, none of the clusters captured specifically Amyloid- β positivity or negativity, although the clusters showed differences in the frequency of Amyloid- β positivity. It is well known that different neurodegenerative conditions can manifest similar clinical presentations leading to imperfect clinic-pathological correlations [152]. Therefore, it is not surprising that this extends to the patterns of brain atrophy captured by structural MRI, which ultimately represent neurodegeneration, which in turn is clinically expressed as impairment in different cognitive functions. Even in ADNI, which is a clinical trial-oriented cohort recruiting subjects with a typical AD profile, there has been a significant neuropathological heterogeneity in patients with classic amnesic AD clinical presentation [153].

The variability revealed by our analysis indicates that a dimensional approach to neurodegeneration in cognitively impaired subjects, including MCI and AD dementia stages, is important and consistent with previous observations of atypical AD presentations [4, 25, 62, 117]. The different patterns we observed might also relate to several coincident neurodegenerative and vascular pathologies [5, 131, 150, 152].

This dimensional approach is important not only for a personalized prognosis, but also for recruiting more homogeneous groups of individuals into clinical trials and tailored interventions. For example, MCI individuals who fall into Cluster 1 have relatively normal brain anatomy and better prognosis, suggesting that their cognitive decline might be highly influenced by factors other than AD pathology. Interestingly, although this group has a decreasing relative frequency with increasing age, as expected, it remains a significant $\sim 20\%$ of this population in ages 80 and older. Individuals in Cluster 2 present the typical AD profile and have rapidly increasing relative frequency with increasing age (Figure 3.9(B)). Individuals along the dimension of Cluster 4 are particularly interesting. These subjects showed focal and pronounced atrophy in the temporal lobe with preserved brain volumes elsewhere. Although subjects in this group showed a relatively slower cognitive decline, long-term follow-up indicated a steady progression from MCI to dementia. These individuals are likely to represent an earlier disease stage, as indicated by a higher frequency of LMCI subjects as opposed to AD dementia at the baseline visit which might later develop to a pattern similar to the one observed in Cluster 2. These changes would be consistent with Braak's tau pathology staging [16]. Cluster 4 could, therefore, be at the relatively early stage of disease, which is potentially an excellent target for clinical trials aiming to slow down disease progression. It is important to note that the cognitive perfor-

mance of these individuals was comparable to the subjects in Cluster 1, indicating that cognitive summary scores might not always capture regional differences in atrophy patterns and lack the ability to detect heterogeneous atrophy patterns. Interestingly, Cluster 4 had a rather stable relative frequency as a function of age (Figure 3.9(B)), which is consistent with the interpretation of this group as newly emerging, early stage AD cases who later move into Cluster 2 as new cases take their place in Cluster 4. Longitudinal analyses are required to further test this hypothesis. Finally, Cluster 3 subjects presented predominantly executive function decline and a more widespread and non-focal pattern of atrophy. Therefore, this cluster might be likely representing atypical AD presentations [117], or a mixture of pathologies, which are commonly present in demented subjects, and are associated with a relatively greater impairment of executive function [150, 152, 153, 154]. The decreasing prevalence of this group with increasing age is consistent with prior work that more “cortical”, or atypical presentations of AD, occur more commonly at a younger age of onset [46]. In addition, the profile of Cluster 3 is consistent with previous results indicating that hippocampal volume alone might be neither a sensitive, nor a specific biomarker in early disease stages [27, 151, 155]. This especially might be the case for atypical non-amnesic presentations without underlying AD pathology. Our results indicate that the entire pattern of brain atrophy needs to be taken into consideration. This also further emphasizes the potential value of such a clustering in clinical trial recruitment, as Cluster 3, similar to Cluster 2, represents a group that has a high likelihood of AD pathology based on CSF $A\beta_{1-42}$ levels, but in which memory and hippocampal measures would be less effective as markers of disease progression than, for example, executive measures.

A somewhat unexpected finding of our study was the fact that Cluster 3 had signifi-

cantly lower WMH load, compared to Cluster 2 and 4, and even to Cluster 1 in ADNI-1, although it was characterized by fast executive decline. This was contrary to our initial expectation that this group might present more of the small vessel ischemic disease phenotype, which would have been consistent with more widespread atrophy. It appears that Cluster 3 might reflect either more cortical presentations of AD, or potentially other kinds of comorbidities or mixed pathologies, which result in the atypical AD pattern of atrophy. For example, the initial study of ADNI autopsy cases identified that cases with coincident Lewy related pathology had a relative large executive impairment compared to memory scores [153]. This unexpected result could be partly explained by the exclusion of subjects with a high baseline Hachinski score, or imaging findings consistent with an infarct, which would result in excluding cases with overt vascular pathology from ADNI.

Somewhat unexpected was also the fact that Cluster 4 had significant WMH load. Since this group's imaging pattern seems to mainly indicate early and likely relatively purer AD pathology, we did not expect to have significant WMH load. This finding is consistent with a recently reported association between high WMH volume and temporal lobe atrophy in a large population based study [63]. Moreover, it has been described that the hippocampus might present relatively higher vulnerability to vascular changes [107]. Cluster 2 also displayed AD-like atrophy and high WMH load. These results indicate that lesion load and AD-like atrophy seem to be correlated, to some extent, even at seemingly early disease stages. This finding is in agreement with growing literature that shows an association between WMH load and AD pathology, albeit our study is not able to determine whether this association is due to shared risk factors or to a more direct relationship in pathophysiology.

Overall, Cluster 2 and 3 were characterized by the relatively highest frequency of

subjects with AD-like CSF $A\beta_{1-42}$ values, whereas Cluster 1 presented a higher number of subjects with normal CSF $A\beta_{1-42}$ values. On the other hand, the frequency of suspected non-amyloid pathology (SNAP) cases differed between the different clusters (4.2% - 36.7%), which indicates that the AD specificity varies across clusters. These findings confirm that different pathologies (or combinations of pathologies) [153] can also be characterized by classical amnesic MCI and AD dementia clinical presentations, while these similarities can even extend to MRI patterns of atrophy.

Finally, the longitudinal MRI changes did also differ between the clusters. Clusters 2 and 3 showed the fastest decline. Areas with the fastest decline corresponded to the same areas that were more involved in baseline comparisons, further reinforcing baseline findings. Interestingly, Cluster 1 showed a large overlap with the CN group, except in the temporal lobe, where Cluster 1 showed a faster progression. Therefore, Cluster 1 may mainly have AD pathology, but probably representing an early stage, at least for part of this group.

Taken together, our findings suggest that there is remarkable heterogeneity in the patterns of brain atrophy that distinguish CN from MCI and AD patients, even in a relatively homogeneous group of subjects, such as the one recruited in ADNI. In particular, a relatively normal group (Cluster 1) displays only mild atrophy. The cross-sectional and longitudinal profile of this cluster suggests that it is heterogeneous, with some individuals likely to progress to the other two clusters (Cluster 2 and Cluster 3), and some likely to remain stable for a relatively long time. Individuals belonging to the Cluster 4 are likely to be at early and rapidly progressing AD disease stages. Individuals in Cluster 3 either represent more cortical presentations of AD, perhaps reflected by their younger age, or

mixed pathologies other than small vessel ischemic disease. Importantly, some of these individuals are likely to progress to predominantly AD dementia, and others to a more executive-prominent cognitive decline and dementia. In addition to shedding light into the neuroanatomical heterogeneity of MCI and AD, our results suggest that patient recruitment into clinical trials might benefit from a finer characterization of the neuroanatomical phenotypes. Finally, our results suggest that a rigorous and quantitative dimensional neuroanatomical approach is necessary for neurodegenerative diseases, in view of the underlying heterogeneity seen, even in relative strictly selected groups of subjects like the ones of ADNI.

Figure 3.4 further underlines the limitations of commonly used voxel-based methods in characterizing the spectrum of neuroanatomical alterations in MCI and AD based on *a priori* diagnostic definitions that rely on disease severity. In particular, these figures indicate that voxel-based analysis of regional volumes detects the same form of brain atrophy pattern in AD dementia and MCI, with the MCI pattern being more spatially restricted and less pronounced. This picture is consistent with a single typical progression pattern for AD, presenting a milder involvement in MCI, thereby largely missing the remarkable heterogeneity unveiled by our results. Given that the underlying pathophysiological mechanisms leading to brain atrophy are complex and heterogeneous, personalized treatment decisions and selection into treatment trials are likely to benefit significantly from the dimensional approach followed herein.

The work presented in this chapter differs substantially from recent clustering-based approaches [114]. Our methodology [36] used herein does not apply direct clustering to the images themselves, which could cluster individuals according to anatomical character-

istics, such as brain or ventricular size, demographics, and other covariates, which are not related to disease effects. In contrast, CHIMERA estimates the disease effects by grouping the differences between patients and controls, after having matched for these confounding covariates. Also, previous findings [114] were derived from a much smaller sample drawn only from ADNI-1, which also uses a limited number of features to perform the clustering (i.e., 11 basic features, including brain, ventricular and hippocampal volumes). As a consequence, previous results seemed to have been significantly affected by outliers in the data (e.g. Cluster 4 of the aforementioned publication included only seven individuals of likely very extreme measurements, which are arguably outliers). However, our findings are generally in agreement with [114]. This is particularly true for the existence of a subpopulation of MCI which is almost entirely normal in all measures, and the existence of a subpopulation, which seems to display a typical AD-like pattern. Albeit their results reveal additional heterogeneity.

Another previous study described three patterns of neurofibrillary tangle deposition based on *a priori* definitions [110]. These groups were described as hippocampal-sparing, limbic-predominant and typical AD, and were later characterized using structural MRIs [163]. Our findings also confirm the presence of subpopulations in MCI and AD-dementia subjects with different degrees of limbic and extra-limbic pathologies, which in some cases do not follow the pattern expected based on tau neurofibrillary tangle tau staging [16]. In our study, cases with a lesser limbic involvement (Cluster 3) were also younger than cases with typical AD atrophy, despite exhibiting larger dorsolateral prefrontal cortex atrophy. As noted above, Cluster 3 is likely to include patients with more cortical disease in young onset cases as described in [46].

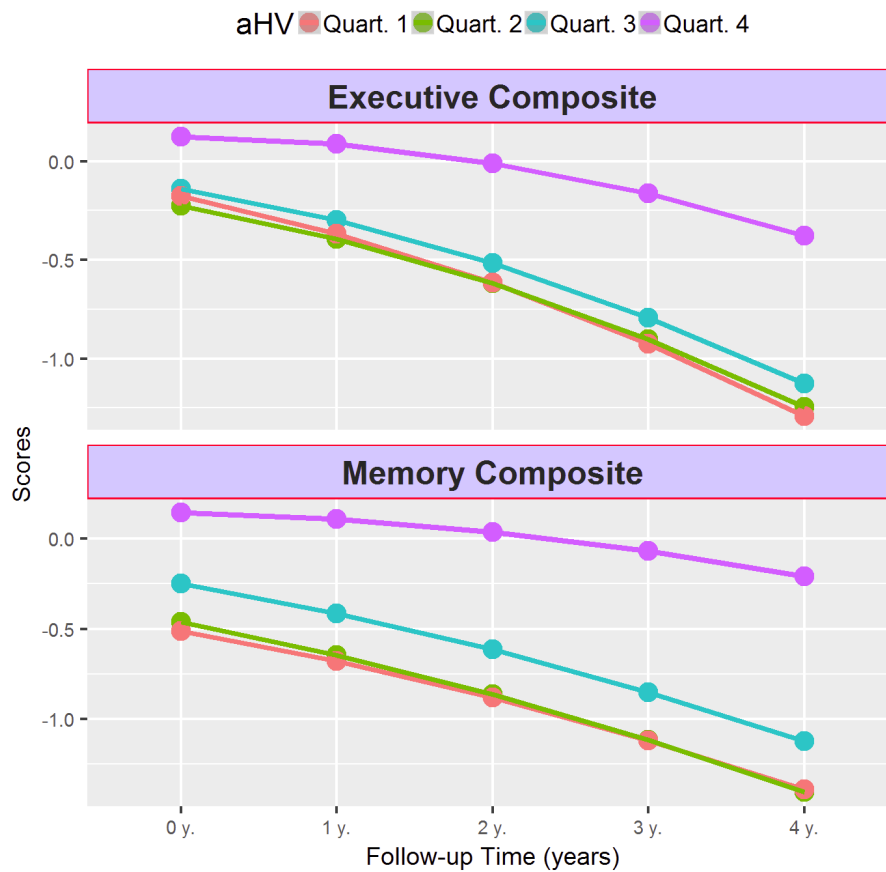


Figure 3.10: Cognitive longitudinal changes based on aHV quartiles. Quartile 1 represents lowest volume, whereas Quartile 4 is the highest.

Grouping subjects based on aHV quartiles led to a linear staging of patient atrophy, primarily reflecting hippocampal volume, as expected. However, these maps failed to reveal the remarkable heterogeneity highlighted by our clustering analysis, including Cluster 3, which was characterized by substantial and widespread cortical atrophy, but relatively preserved hippocampal volumes. We have also calculated conversion and longitudinal cognitive models, as shown in Figure 3.6(B) and Figure 3.10. Although four aHV quartiles were studied, these analyses showed that patterns effectively represented two subgroups. Specifically, one subgroup comprised 75% of the subjects who progressed rapidly, while

the second subgroup consisted of 25% of the subjects, who remained more stable. This suggests that while hippocampal volume is an important biomarker of memory decline, it is not sufficiently rich or specific in capturing the heterogeneity of AD. In order to disentangle the heterogeneity, our analysis suggests that it is important to delineate imaging patterns throughout the brain.

Limitations of this study are: 1) the smaller sample size for ADNI-GO/2 compared to ADNI-1, which led to small numbers in some clusters; 2) the relative homogeneity of the subjects recruited in the study, which were recruited either as late MCI subjects with AD-like clinical presentation or early AD dementia type subjects. Future studies should expand to more heterogeneous populations.

Chapter 4

Parsing Heterogeneous Longitudinal Trajectories

4.1 Introduction

Longitudinal studies are indispensable tools for tracking structural or functional brain changes over time [30, 47, 61, 76, 88, 142]. In longitudinal studies, each subject serves as his/her own control, and change is assessed directly over repeated evaluations. This is in contrast to cross-sectional approaches, where change is evaluated by comparing different individuals, and thus may be confounded by normal variations across the population. The within-subjects comparisons in longitudinal studies have the advantage of reducing inter-subject variability, which results in increased statistical power for quantifying change over time. This allows the extraction of subtle brain imaging patterns that accurately characterize temporal dynamics.

The above properties make longitudinal studies particularly suitable for studying the

dynamics of pathological processes [76, 128]. As a consequence, longitudinal designs have found numerous applications in computational neuroimaging. Among their most important applications, one may cite the identification of the neuroimaging pattern that reflects disease processes [40, 130, 133, 147], the accurate description of age trajectories of brain region volumes [104, 164], as well as the study of the associations between cognitive change and brain structure and function [40, 123, 135].

Longitudinal studies typically make use of analytical tools that fall into two main categories. The first group of methods puts emphasis on differentiating distinct longitudinal processes by performing group comparisons between categorizations of subjects (e.g., comparing individuals before and after treatment). These categorizations are commonly based on clinical diagnosis [19, 73], or on individual characteristics extracted from the imaging data [66, 99]. The second group of approaches, on the other hand, focuses on describing the brain temporal dynamics by employing linear/nonlinear mixed effects regression models [12, 13, 14, 124, 157]. The mixed effects model can effectively model the global trend that spans the time range of the entire study as fixed effect, while also accounting for individual variability using random effect variables.

The overarching assumption behind the above methods is that the population of interest is a homogeneous entity, resulting from a single unifying (patho)physiological temporal process, which is characterized by a single imaging signature. This assumption effectively ignores ample evidence for the heterogeneous nature of both neurodegenerative diseases [90, 110] and neuropsychiatric disorders [43, 78, 113]. In such cases, the members of a heterogeneous population may present different phenotypes, which may be characterized by multiple distinct disease processes. As a consequence, common approaches that ignore

heterogeneity may only find a common/average direction of longitudinal brain changes. If we assume that two groups, or two conditions, differ by one longitudinal pattern in a subpopulation and a different longitudinal pattern in another subpopulation, such a common denominator is at best incomplete, and at worst misleading. Accurately identifying and quantifying heterogeneous longitudinal trajectories is essential for elucidating underlying heterogeneous neurobiological mechanisms, for improving disease diagnosis and prognosis, for targeted treatment interventions, as well as for enhancing therapeutic innovation.

In order to disentangle the heterogeneity of temporal (patho)physiological processes, clustering of longitudinal trajectories can be performed. There is a rich literature on the topic of curve clustering [2, 51, 52, 94]. However, most approaches [2, 51, 52] commonly require the trajectories to have the same length in order to be able to compare them effectively. This requirement is hard to fulfill in longitudinal studies, where subjects may be recruited at different time points, and may also have different lengths of follow-up. On the other hand, methods that can account for uneven-length trajectories, such as [94], seek for common groups of directions based on linear segments of individual trajectories. The continuity of individual curves is disregarded, which is crucial when applied to longitudinal studies, and thus not applicable to our needs.

To address the above limitations, we propose a novel method for analyzing the Heterogeneity of Longitudinal Imaging trajectories by integrating clustering and spatiotemporal Spline modeling of trajectories, termed HELIOS. The proposed framework estimates simultaneously multiple trajectories at the population level, and performs unsupervised multivariate clustering by assigning each individual to the population trajectory, whose segment fits best the individual's data. Population trajectories span the full time range

of the study and capture the temporal dynamics of respective variables/features. In estimating global trajectories, we assume that each individual trajectory is a segment of the global trajectory. This allows us to account for differences in recruiting time and length of follow-up. Given this assumption, the method attempts to reconstruct the global trajectories using the partial information from all the subjects' individual trajectories. Global trajectories are modeled using spatiotemporally regularized cubic B-splines. Polynomial splines are used to account for non-linearities, while temporal regularization controls the variance of the estimated trajectories. We additionally introduce spatial regularization to take advantage of the spatial structure of the data toward improving model estimation and statistical power. Multiple trajectories per variable/feature are estimated to model the underlying heterogeneous processes. As a consequence, heterogeneity of brain changes is modeled as multiple sets of trajectories, with each set representing a distinct multivariate imaging pattern of longitudinal change. Heterogeneity is dissected by clustering subjects based on how well corresponding segments of the global trajectories fit their own individual data time points. The resulting optimization problem is solved through Expectation-Maximization (EM) [15], where a multi-initialization strategy is used and the clustering result that yields the lowest objective is kept.

4.2 Method

In this chapter, we analyze the trajectories of imaging features across the temporal dimension, by modeling non-linear trajectories using cubic B-splines. In order to take into account the underlying heterogeneity, we assume that there are K subgroups of subjects following different longitudinal multivariate patterns of change. Each multivariate pat-

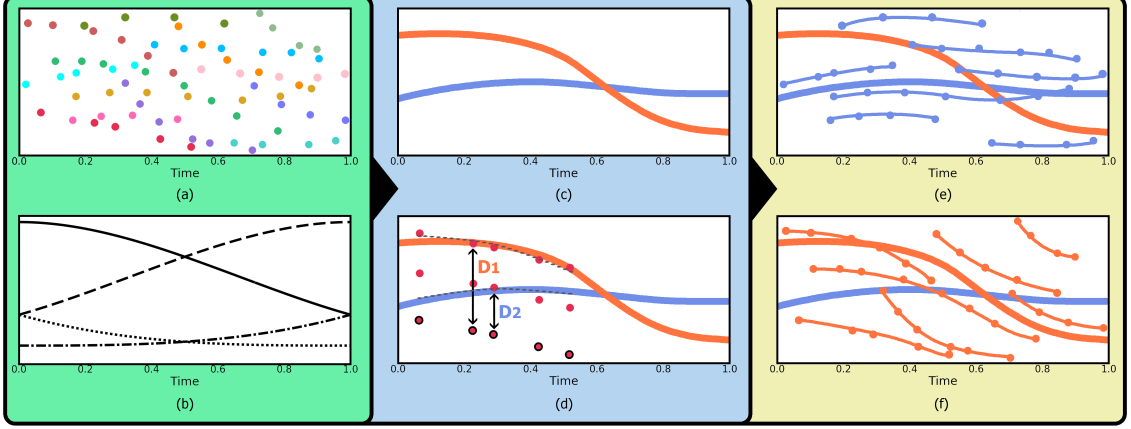


Figure 4.1: HELIOS takes as input a longitudinal dataset (illustrated in (a) by coloring all time points of one subject the same), a set of uniform cubic B-spline bases spanning the entire time range (shown in (b)), and the number of subgroups ($K=2$ here). Given these, HELIOS operates by simultaneously (c) estimating the subgroup-specific global trajectories as an optimal linear combination of the four bases, and (d) assigning each subject to the global trajectory that is most similar to it. The similarity is evaluated after accounting for differences in the vertical direction through the use of offset variable D . At the end of the algorithm, clustering and fitting for all subjects with respect to the two global trajectories has been performed ((e) and (f), respectively).

tern is described by a set of feature trajectories, and each subject is assigned to the subgroup whose imaging trajectories best represent its own individual changes. The above procedure is illustrated in Figure 4.1, where longitudinal trajectories are estimated for two groups.

This concept is formulated in a probabilistic setting in our framework. Optimal trajectories and clustering are estimated simultaneously by minimizing the following objective:

$$\mathcal{E}(X, \theta_1, \dots, \theta_K) = \mathcal{L}(X, \theta_1, \dots, \theta_K) + \mathcal{R}(\theta_1, \dots, \theta_K),$$

where \mathcal{L} is the least squares loss of data X fit by the cubic B-splines. The splines are described using K sets of parameters Θ , while the regularization \mathcal{R} of Θ encourages temporal smoothness of the splines and incorporates spatial covariance structure of the imaging fea-

tures. We present the two parts in detail in the following sections.

4.2.1 Loss term

Let the longitudinal study include N subjects, x_1, \dots, x_N , where each subject x_i has n_i imaging scans acquired at time points $t_{i1}, t_{i2}, \dots, t_{in_i}$. The number of follow-up scans varies for different subjects. An illustration of the longitudinal dataset is shown in Figure 4.1(a). For each imaging scan, a number of P features (e.g., region volumes, tissue densities, activation maps, etc.) are extracted. We denote the p th feature of j th time point of subject x_i by $x_{ip}^{t_{ij}}$.

In this work, the longitudinal trajectories of every feature are modeled using B-splines, which can be represented by an M dimensional basis vector $B(t)$, which spans the full time range of the study. Each spline can be uniquely described using a coefficient vector C of dimension M , i.e., the parametrized B-spline can be expressed by $B(t)C$. For the experiments herein, we chose cubic B-splines ($M=4$, see Figure 4.1(b)), which provides us with sufficient non-linearity to model the trajectories, but also with a reasonable number of parameters to optimize. Given time t , the bases of cubic B-splines can be obtained with the pre-defined functions:

$$B(t) = \begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix} \cdot \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}. \quad (4.1)$$

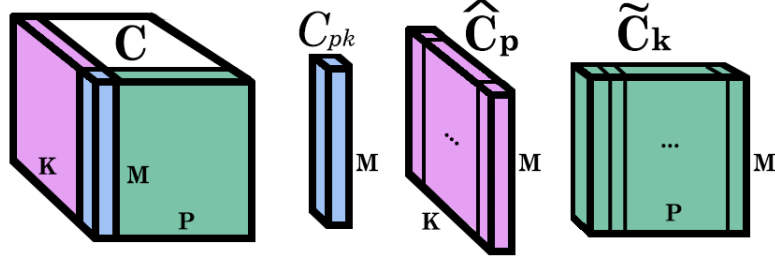


Figure 4.2: Illustration of the coefficient tensor \mathbf{C} . C_{pk} denotes the coefficient vector for a single spline; the $\hat{\mathbf{C}}_p$ cross section contains K sets of coefficients; the $\tilde{\mathbf{C}}_k$ cross section contains P sets of coefficients.

All the spline coefficients in our approach are represented together by a 3D tensor $\mathbf{C} \in \mathbb{R}^{K \times P \times M}$, which is illustrated in Figure 4.2. Every column of \mathbf{C} , $C_{pk} \in \mathbb{R}^{M \times 1}$, corresponds to a single spline that describes the trajectory of the p th feature of cluster k . To ease the notation, we denote two relevant cross sections of the tensor \mathbf{C} by $\hat{\mathbf{C}}$ and $\tilde{\mathbf{C}}$. Specifically, for each feature p , there is a set of K coefficients $\hat{\mathbf{C}}_p \in \mathbb{R}^{M \times K}$, describing the behavior of this feature in K different subgroups. On the other hand, for each subgroup k , there is a set of P coefficients $\tilde{\mathbf{C}}_k \in \mathbb{R}^{M \times P}$, which describe the multivariate feature trajectories for this subgroup. These two views of \mathbf{C} are used in the loss term \mathcal{L} and the regularization term \mathcal{R} , respectively.

Our goal is to simultaneously find the multivariate longitudinal patterns for all subgroups, while fitting the time points of every subject x_i according to which subgroup it belongs to. This fitting is evaluated and optimized by adopting a least square loss objective function. Without loss of generality, we describe below the fitting scheme for the p th feature of the multivariate pattern.

First, let us recall that the coefficients for the trajectories of the K different subgroups are stored in the matrix $\hat{\mathbf{C}}_p$. For example, in Figure 4.1(c), there are two sets of the M

coefficients describing the red and blue trajectories. For each subject x_i , we introduce a variable $\zeta_i \in \{0, 1\}^{K \times 1}$ to indicate its clustering membership to one of the K subgroups, where ζ_i is a vector with all elements being zero except one, which is equal to 1. The indicator selects the appropriate coefficients from $\hat{\mathbf{C}}_p$ for subject x_i , and it is shared among all features of x_i . We effectively assume that every subgroup is characterized by a single multivariate pattern (illustrated by the blue columns shown in Figure 4.3). Let us denote by t_{ij} the time of the j th follow-up scan of subject x_i . Accordingly, we can compute the B-spline basis at this time point as $B(t_{ij})$. Given the above notation, the least squares loss for fitting the p th feature of x_i to its corresponding global trajectory (as determined by ζ_i) can be written as $(x_{ip}^{t_{ij}} - B(t_{ij})\hat{\mathbf{C}}_p\zeta_i)^2$.

However, this loss emphasizes differences due to absolute value, which may reflect inter-subject variability. Thus, in order to focus on differences in the shape of trajectories, we introduce a subject-specific offset variable $\mathbf{D} \in \mathbb{R}^{N \times P \times K}$. This can be understood as a random intercept in a mixed effects model, where \mathbf{D}_{ip} contains K shifts for the p th feature of subject x_i . Each shift translates globally the values of the feature for all time points of the subject, such that they best align with the corresponding global trajectory. This is illustrated in Figure 4.1(d), where the longitudinal points of a single subject are shifted according to two offsets toward two different global trajectories. The subject is assigned to the global trajectory that best explains its own data (the orange curve in this case).

Summing up the least square fit loss for all N subjects, all P features, and all time points of each subject constitutes the loss term \mathcal{L} of our model:

$$\mathcal{L} = \sum_{i=1}^N \sum_{p=1}^P \sum_{j=1}^{n_i} \left(x_{ip}^{t_{ij}} - B(t_{ij})\hat{\mathbf{C}}_p\zeta_i - \mathbf{D}_{ip}\zeta_i \right)^2.$$

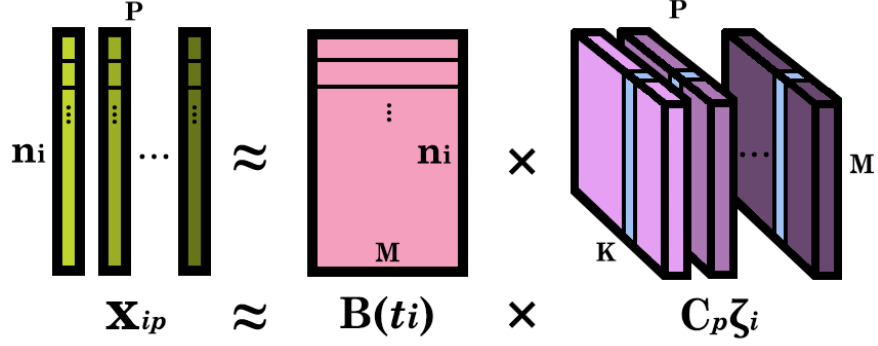


Figure 4.3: Illustration of the least squares fit: all features of the same subject share basis $\mathbf{B}(t_i)$, but have different coefficients from \mathbf{C} that are chosen by the indicator ζ_i . To ease the illustration, the offset variable \mathbf{D} has not been included here.

To reach a more compact form of the loss function, we denote the p th feature of all time points of x_i as $\mathbf{x}_{ip} \in \mathbb{R}^{n_i}$. We also introduce the $n_i \times M$ matrix $\mathbf{B}(t_i) = [B(t_{i1}), \dots, B(t_{in_i})]^T$ to represent the B-spline basis for all time points of x_i . The above notation provides the vectorized loss function, which will be used in the following sections:

$$\mathcal{L} = \sum_{i=1}^N \sum_{p=1}^P \left\| \mathbf{x}_{ip} - \mathbf{B}(t_i) \hat{\mathbf{C}}_p \zeta_i - \mathbf{D}_{ip} \zeta_i \mathbf{1} \right\|_2^2 \quad (4.2)$$

where $\mathbf{1}$ is an n_i dimensional vector with all elements being 1. Figure 4.3 illustrates how the tensor \mathbf{C} is used in the spline fitting with this compact form (without \mathbf{D}).

4.2.2 Regularization term

In order to avoid overfitting, we promote both the spatial and temporal smoothness of the estimated B-spline coefficients through the introduction of two regularization terms in the proposed framework.

Temporal smoothness of splines is encouraged because we expect individuals to change gradually. Therefore, we penalize large spline coefficients for all subgroups and features

by including the following quadratic penalty:

$$\mathcal{R}_T = \sum_{k=1}^K \sum_{p=1}^P \|C_{pk}\|_2^2 = \sum_{k=1}^K \text{tr}(\tilde{\mathbf{C}}_k^T \tilde{\mathbf{C}}_k), \quad (4.3)$$

The second regularization takes into account the spatial covariance of imaging features. This term effectively models the assumption that imaging features, which correspond to spatially close brain regions, are also biologically connected, and thus they are likely to undergo similar longitudinal change. Imposing a spatial regularization can effectively improve the robustness of the method to noise, which might have been introduced by preprocessing steps. Hence, in the framework, the difference of coefficients of features a and b within the same subgroup is penalized according to their spatial distance. When two features have small spatial distance, their B-spline coefficients are encouraged to be similar. We use the spatial covariance matrix $\Sigma \in \mathbb{R}^{P \times P}$ introduced by [13], comprising elements $\sigma_{ab} = e^{-d_{ab}}$, where d_{ab} is the Euclidean distance between the coordinates of brain regions/features a and b . Thus, the second regularization term can be written as:

$$\begin{aligned} \mathcal{R}_S &= \sum_{k=1}^K \left(\sum_{a=1}^{P-1} \sum_{b=a+1}^P \sigma_{ab} \|C_{ak} - C_{bk}\|_2^2 \right) \\ &= \sum_{k=1}^K \text{tr}\{(\tilde{\Sigma} - \Sigma) \tilde{\mathbf{C}}_k^T \tilde{\mathbf{C}}_k\}, \end{aligned} \quad (4.4)$$

where $\tilde{\Sigma}$ is a diagonal matrix with its diagonal elements calculated by $\tilde{\Sigma}_{ii} = \sum_{p=1}^P \sigma_{ip}$.

4.2.3 Optimization

We combine the loss term and the regularization term, with balancing parameters η (between loss and regularization terms) and λ (between temporal and spatial regularization terms), to obtain the objective function:

$$\mathcal{E} = \sum_{p=1}^P \sum_{i=1}^N \|\mathbf{x}_{ip} - \mathbf{B}(t_i) \hat{\mathbf{C}}_p \boldsymbol{\zeta}_i - \mathbf{D}_{ip} \boldsymbol{\zeta}_i \mathbf{1}\|_2^2 + \eta \sum_{k=1}^K \text{tr} \left\{ (\lambda \mathbf{I} + \tilde{\Sigma} - \Sigma) \tilde{\mathbf{C}}_k^T \tilde{\mathbf{C}}_k \right\}. \quad (4.5)$$

In order to estimate the offset variable \mathbf{D} , the clustering memberships $\boldsymbol{\zeta}$ and the spline coefficients tensor \mathbf{C} , we employ an adapted Expectation-Maximization algorithm to minimize the objective function (4.5). Specifically, we iterate between the following three steps until the difference of objective values between two iterations is smaller than a tolerance criterion.

1. **O-step:** The offset variables are calculated for each subject based on the splines obtained in the previous iteration. We denote \mathbf{D}_{ip}^k as the offset scalar for the p th feature of subject x_i with respect to the k th spline C_{pk} . The offset \mathbf{D}_{ip}^k acts as an “intercept” that is optimized to yield the minimal residual of the fit. Hence, it is computed as the average of the sum of residuals when fitting the B-spline with coefficient C_{pk} :

$$\mathbf{D}_{ip}^k = \frac{1}{n_i} \mathbf{1}^T (\mathbf{x}_{ip} - \mathbf{B}(t_i) C_{pk}). \quad (4.6)$$

2. **E-step:** The E-step is the clustering phase of the optimization. Given the global trajectories estimated in the previous step, and the offset variables calculated in the

O-step, we assign each subject to the subgroup k that provides the best fit:

$$\zeta_i = \underset{\zeta_j}{\operatorname{argmin}} \left\{ \sum_{p=1}^P \|\mathbf{x}_{ip} - \mathbf{B}(t_i) \hat{\mathbf{C}}_p \zeta_j - \mathbf{D}_{ip} \zeta_j \mathbf{1}\|_2^2 \right\}. \quad (4.7)$$

3. **M-step:** Because the loss term \mathcal{L} and the regularization term \mathcal{R} use two different views of the coefficient tensor \mathbf{C} , it is particularly difficult to derive an analytical solution for \mathbf{C} . Therefore, we adopted gradient descent method to estimate the coefficients. The two cross sections, $\hat{\mathbf{C}}$ and $\tilde{\mathbf{C}}$, share columns, thus allowing the optimization to be achieved in a column wise fashion. Let us denote by $\mathbf{e}_k \in \{0, 1\}^{1 \times K}$, a K dimensional vector with only the k th element being 1, and by $\boldsymbol{\epsilon}_p \in \{0, 1\}^{1 \times P}$, a P dimensional vector with only the p th element being 1. Thus $\hat{\mathbf{C}}_p$ can be rewritten as: $\hat{\mathbf{C}}_p = \sum_{k=1}^K C_{pk} \mathbf{e}_k$ while $\tilde{\mathbf{C}}_k$ can be written as $\tilde{\mathbf{C}}_k = \sum_{p=1}^P C_{pk} \boldsymbol{\epsilon}_p$. Each column C_{pk} can then be found using a quasi-Newton gradient descent method, where the partial derivative is derived as:

$$\frac{\partial \mathcal{E}}{\partial C_{pk}} = - \sum_{i=1}^N 2 \mathbf{e}_k \zeta_i \mathbf{B}(t_i)^T (\mathbf{x}_{ip} - \mathbf{B}(t_i) \hat{\mathbf{C}}_p \zeta_i - \mathbf{D}_{ip} \zeta_i \mathbf{1}) + 2 \eta \tilde{\mathbf{C}}_k (\lambda \mathbf{I} + \tilde{\Sigma} - \Sigma) \boldsymbol{\epsilon}_p^T. \quad (4.8)$$

Initializing \mathbf{C} randomly may lead to a slow convergence of this optimization process. In order to overcome this, we randomly partition the data samples into K subgroups and initialize each $\tilde{\mathbf{C}}_k$ using one of the random subgroups. The clustering is performed multiple times with different initializations, and we report the clustering result that yields the lowest objective.

4.3 Experiments

In this section, we present two sets of experiments. In the first sets of experiments, we evaluated our approach using simulated data. We examined the capability of HELIOS to cluster heterogeneous data by varying 1) the average length of individual trajectories, and 2) the signal to noise ratio. Lastly, we quantitatively compared the proposed approach with two baseline curve clustering methods using the synthetic data. In the second set of experiments, we applied HELIOS on a longitudinal study comprising cognitively normal elderly individuals to analyze the heterogeneity in aging. Our analysis revealed two stable/reproducible subtypes, which follow distinct imaging trajectories, and also have different cognitive profiles. Detailed investigations that are outside the scope of this paper will be carried out in the future for elucidating all the medical implications of this finding.

4.3.1 Synthetic data

We validated our approach on a series of simulated datasets modeling heterogeneous trajectories. Three independent trajectories were simulated as shown in Figure 4.4(A): 1) a slow decreasing pattern; 2) a late fast decreasing pattern; and 3) an early fast decreasing pattern. Time was normalized to range $[0,1]$. For every simulation conducted, we generated 400 samples, with 100 samples for each subgroup. Every sample has five features, and each feature follows one of the three trajectory patterns as indicated by the color in Figure 4.4(B). Random shifts (uniformly distributed) of feature values were also introduced for each individual trajectory.

In order to demonstrate the strength of our proposed method, we compared its performance against two variants of k-means clustering [97]. The first variant, referred to here-

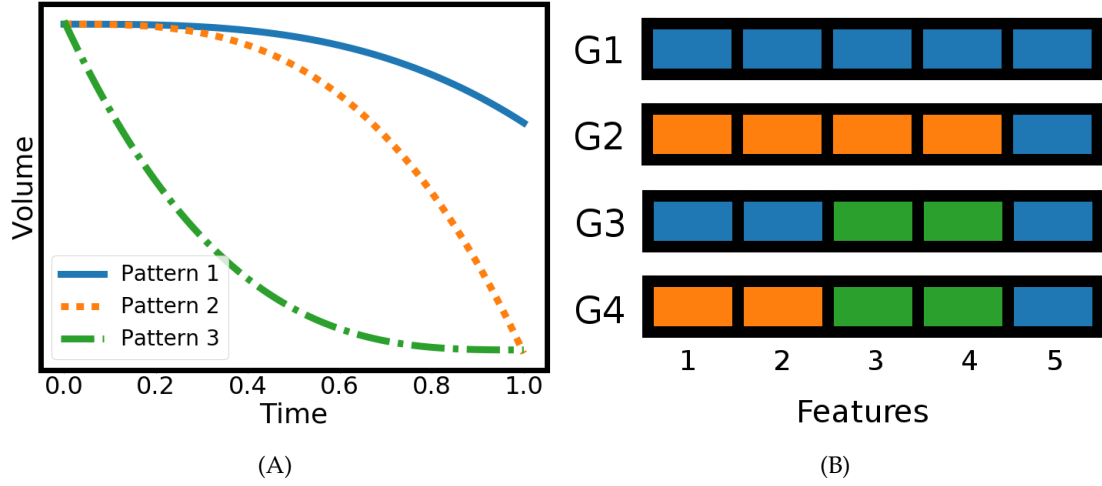


Figure 4.4: (A) The three simulated trajectory patterns. (B) Construction of four subgroups following different multivariate patterns, where each subject has five features. The color of each feature indicates the trajectory it follows.

after as Linear-KM, conducts linear regression separately on each feature trajectory of a sample, and uses the estimated slopes to form a 5-dimensional clustering feature. The second variant, termed hereafter Nonlinear-KM, aims to take into account the non-linearity of the trajectories by first removing the mean of each feature trajectory, and then fitting the trajectory with a regularized cubic B-spline. The estimated coefficients of all splines for one sample are concatenated in a 20-dimensional clustering feature. Lastly, k-means clustering is performed using the derived features to generate memberships for all the samples.

In our first experiment, we aim to evaluate the impact of the length of individual trajectories in estimating the global trajectories and clustering the population. Global trajectories, which span the entire time range of the longitudinal study, are estimated by taking into account information from all subjects. However, each subject may only contribute to the segment of the trajectory that corresponds to the period during which it was scanned. Given that the longitudinal studies often recruit participants at different time points, while

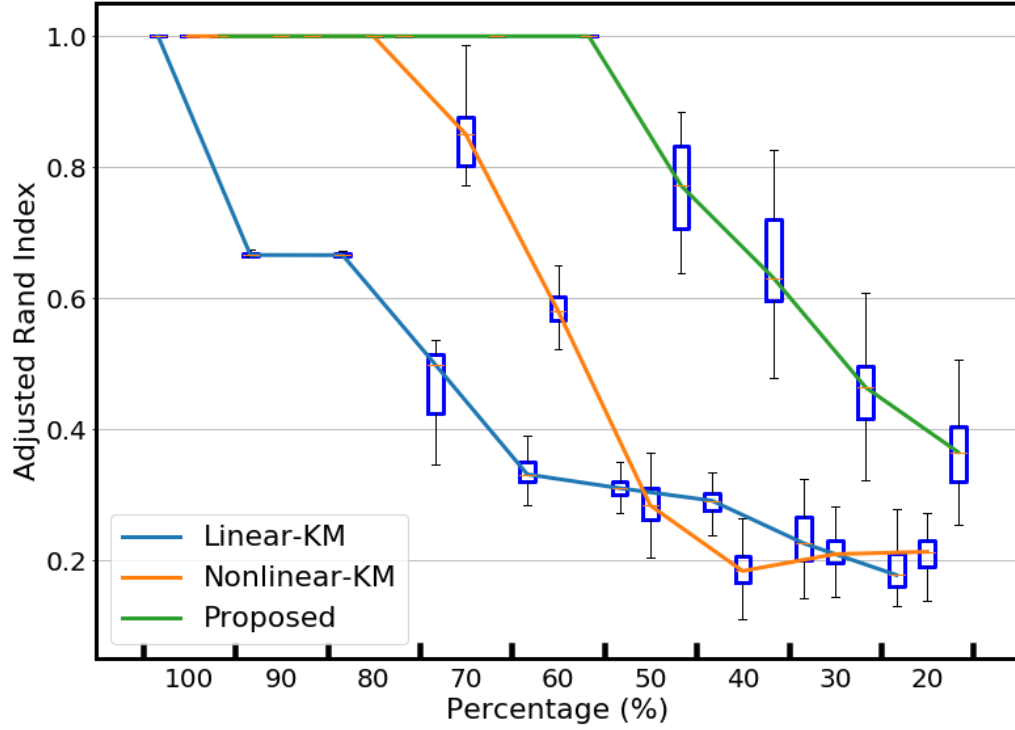


Figure 4.5: Comparison between the proposed method and the two k-means variants on simulated data. The performance is quantified by the adjusted rand index, where results are obtained by varying the lengths of individual trajectories.

also different numbers of serial scans are obtained for each participant, it is important to understand how the length of the follow-up for each participant affects the performance of the algorithm. This information will help us determine whether our model can successfully parse the heterogeneity within specific datasets.

Toward this end, we simulated individual trajectory lengths ranging from 100% to 20% of the global trajectory. For each length setting, we drew samples with a random starting time point and a fixed length of trajectory. These were afterward clustered using our proposed method, as well as the two baseline methods. The performance of the clustering result was evaluated by measuring the agreement between the obtained memberships and the true memberships using the Adjusted Rand Index (ARI) [68]. We repeated this process

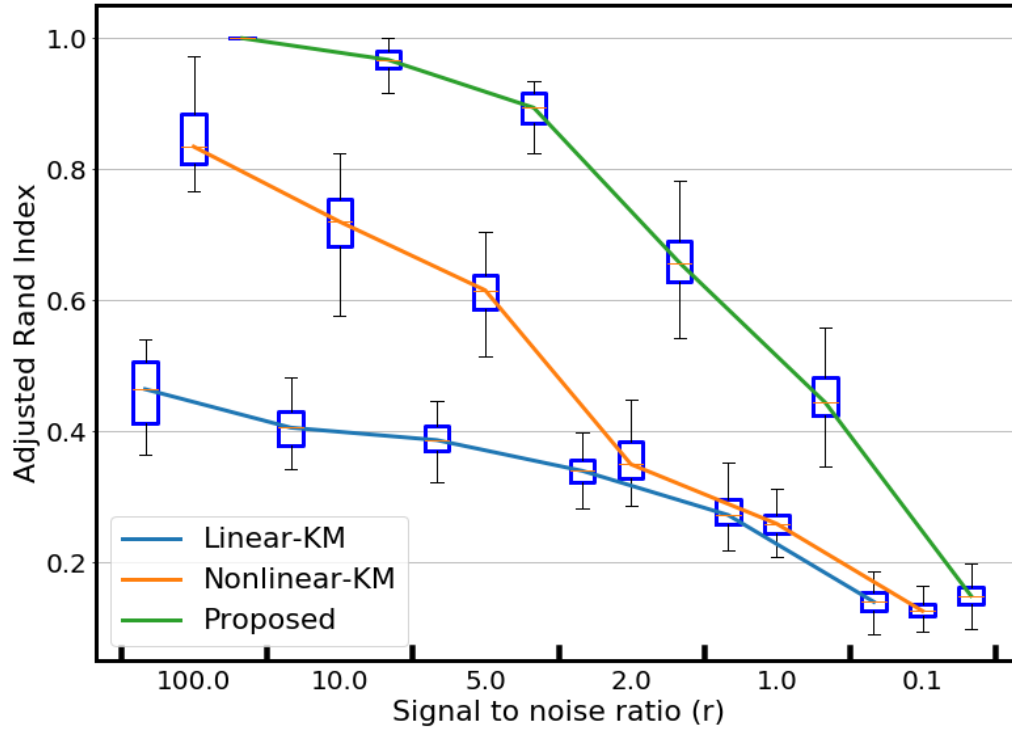


Figure 4.6: Comparison between the proposed method and the two k-means variants on simulated data. The performance is quantified by the adjusted rand index, where results are obtained for different signal to noise ratio levels, while fixing the individual trajectories length to be equal to 70% of the time range.

100 times to estimate the distribution of the performance shown in Figure 4.5. The figure shows that all the methods could perform perfectly when each individual trajectory had full length. However, the performance started decreasing for shorter follow-up durations. Specifically, Linear-KM was the first to drop accuracy for individual lengths equal to 90% of the global trajectory. Nonlinear-KM followed next for individual trajectories covering 70% of the full time range. The proposed approach was the most robust one, dropping performance when individual trajectories were equal to half the global trajectory. Importantly, HELIOS produced reasonable clustering results even when the length of individual trajectories was at 20% of the global trajectory.

Secondly, we analyzed the influence of noise in disentangling heterogeneous trajectory patterns. To evaluate the effect of noise, we fixed the length of individual trajectories to be equal to 70% of the range of the simulation study, which is the point when we first observed discrepancies between the three approaches in the previous experiment. Then, we introduced Gaussian noise $\mathcal{N}(0, \tau^2)$ to all the features by varying the signal to noise ratio r (defined as $r = \sigma^2/\tau^2$) from 100 to 0.1. We repeated the simulation for each noise level setting 100 times. The results are reported in Figure 4.6. The results demonstrate the improved robustness of the proposed method compared to the k-means variants, and suggest that a signal to noise ratio equal to 1.0 is sufficient for obtaining a reasonable clustering.

The above experiments clearly demonstrated that our proposed approach outperforms the baseline methods by a large margin. This encouraged us to apply HELIOS to a large longitudinal study of aging.

4.3.2 Longitudinal aging data

The Baltimore Longitudinal Study on Aging (BLSA) [140] is America’s longest-running scientific study of human aging conducted by the National Institute on Aging (NIA). It comprises healthy elderly adults that are followed annually with rich radiologic examinations, including structural and functional MR images. Imaging data are complemented with neuropsychological tests aiming to evaluate participants in different cognitive domains. Specifically, verbal memory is assessed using the California Verbal Learning Test (CVLT) and visual memory with the Benton Visual Retention Test (BVRT); category fluency (FLUCat) and letter fluency (FLULet) assess executive function, while the card rotations test (CRDRot) assesses visuospatial function. More details about the cognitive tests

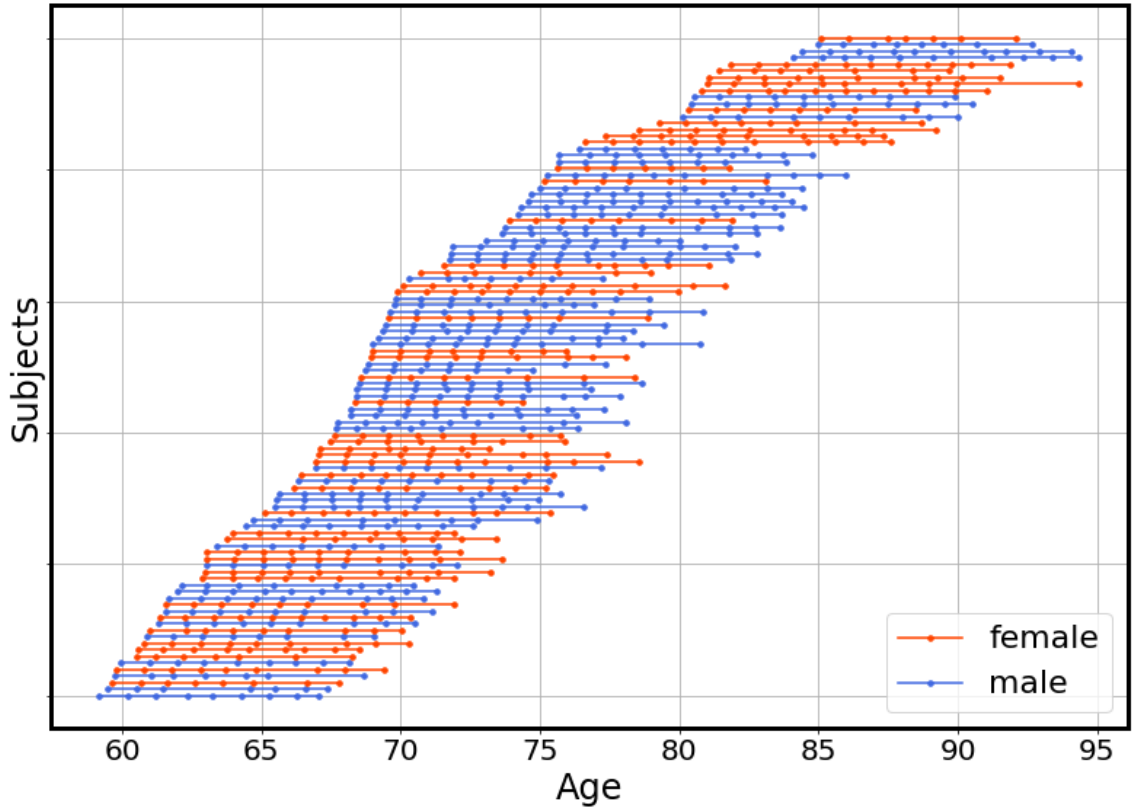


Figure 4.7: Distribution of scans per person by sex across the study age span. Each point denotes a scan; horizontal lines connect scans from the same individual. Red, female; blue, male.

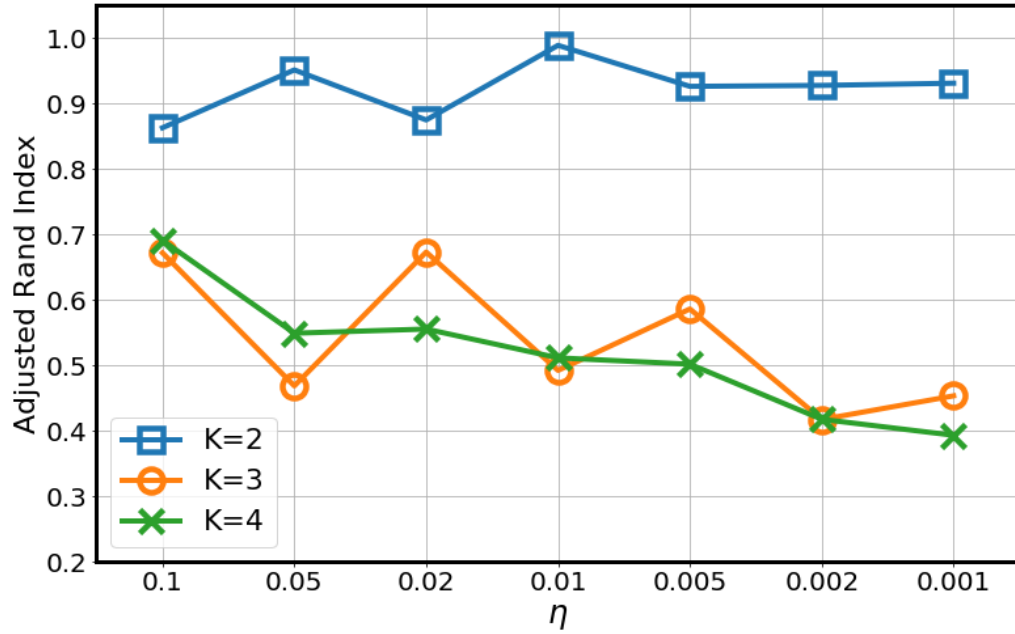
can be found in Appendix C, while more details about the study can be found in Chapter 1.

We applied our approach to a subset of BLSA dataset, consisting of 102 subjects who have been followed for more than six years. This subset contains 895 T1 structural MR imaging scans in total. We present in Figure 4.7 basic demographic information (i.e., age, gender, as well as the number of time points) for the subjects used in this study. The average follow-up time is more than 25% of the entire age range of the study, which, according to our simulations, provides enough information for our method to delineate heterogeneous aging patterns within this population.

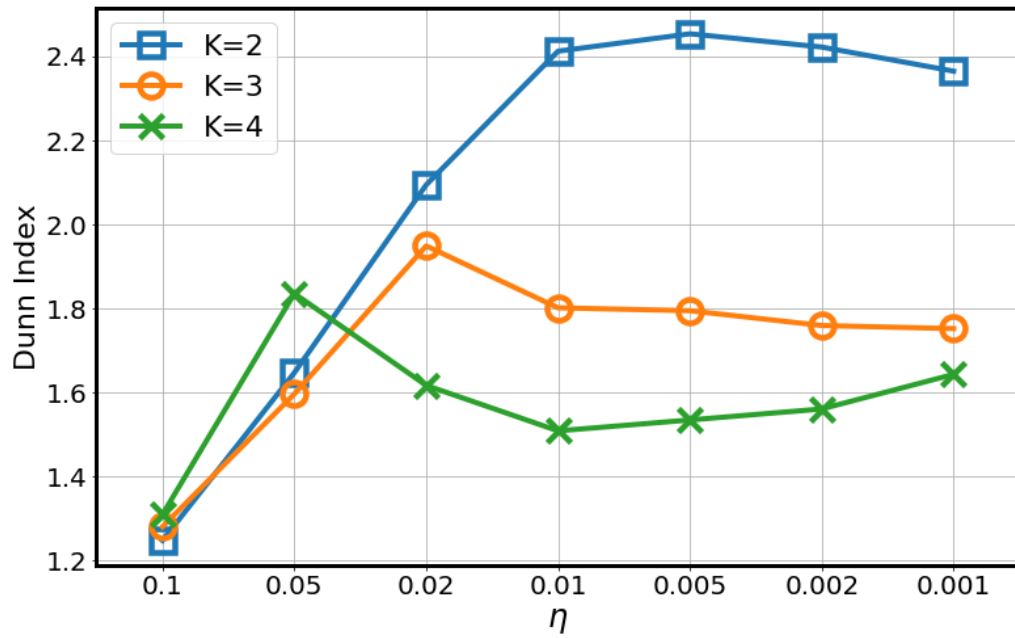
All the brain images were skull-stripped and segmented into 34 anatomical regions of interest (ROI), using a previously described multi-atlas segmentation protocol [38]. The dictionary of the employed ROIs is presented in Appendix B.2. In this experiment, we chose relatively large ROIs in order to reduce the noise level, which may be introduced during image acquisition and preprocessing. We normalized each ROI volume by the total intracranial volume (ICV) of the baseline visit scan. Moreover, each ROI was standardized to a normal distribution $\mathcal{N}(0, 1)$ across the entire population. The individual trajectories were smoothed afterward using a regularized cubic spline in order to further reduce noise in the temporal dimension. A spatial covariance matrix was computed by calculating the Euclidean distance between all pairs of ROIs in the MNI coordinate system.

The model parameters (i.e., the number of subgroups K , the regularization coefficient η , and the balancing weight for the two regularization terms λ) were selected by evaluating the clustering results in terms of their reproducibility across multiple runs, and their quality as quantified by the Dunn Index (DI).

In our analysis, we favor solutions exhibiting higher reproducibility as we assume that, as one gets closer to the intrinsic dimension of the solution subspace, the clustering algorithm should obtain similar results across different runs. The clustering reproducibility was measured as follows. For each combination of model parameters, our method was applied to the dataset for determining the optimal trajectories, as well as the clustering labels for all the subjects. This procedure was repeated 100 times for each model parameter set, and thus 100 different clustering labels were generated. The clustering reproducibility was finally measured by computing the ARI for each pair of clusterings. The average ARI value across all pairs was used to summarize the performance for each repetition.



(A)



(B)

Figure 4.8: Clustering performance measured by (A) the Adjusted rand index, and (B) the Dunn index, for different sets of hyperparameters K and η .

Additionally, we favor solutions comprising clusters that are compact, and at the same time, well separated. As a consequence, we prefer parameter sets that lead to higher Dunn Index. The Dunn index was calculated as the ratio between the 5th percentile of the inter-cluster distance and the 95th percentile of the intra-cluster distance, thus larger DI indicates a better clustering. In our analysis, we adapted the calculation of the Dunn Index to take into account the 5th percentile and 95th percentile instead of the respective minimum and maximum values to reduce the sensitivity of the ratio to outliers, and thus improve its stability. The intra-cluster distance measure was defined to be the sum of squared residuals (SSR) estimated by fitting each global trajectory to the time points of individuals that have been assigned to it. The inter-cluster distance was calculated by estimating the SSR by fitting global trajectories to the time points of individuals that have been assigned to different global trajectories.

In this experiment, we set λ to be 1 in order to provide the same level of regularization for both the spatial and temporal dimensions of the model. The following combinations of model parameters were considered: $K = \{2, 3, 4\}$ and $\eta = \{0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001\}$. The median value of the average ARI was calculated across all iterations for each model parameter set and is shown in Figure 4.8(A). The median value of the DI results for each parameter combination is shown in Figure 4.8(B). By taking into account the two criteria, we chose $K = 2$ and $\eta = 0.01$, which yielded two subgroups consisting of 29 and 73 subjects, respectively.

Individual trajectories for all participants, as well as the global population trajectories which were estimated by the proposed method, are shown in Figure 4.9. The two subgroup trajectories are colored red (G1) and blue (G2), respectively. We selected six ROIs from the

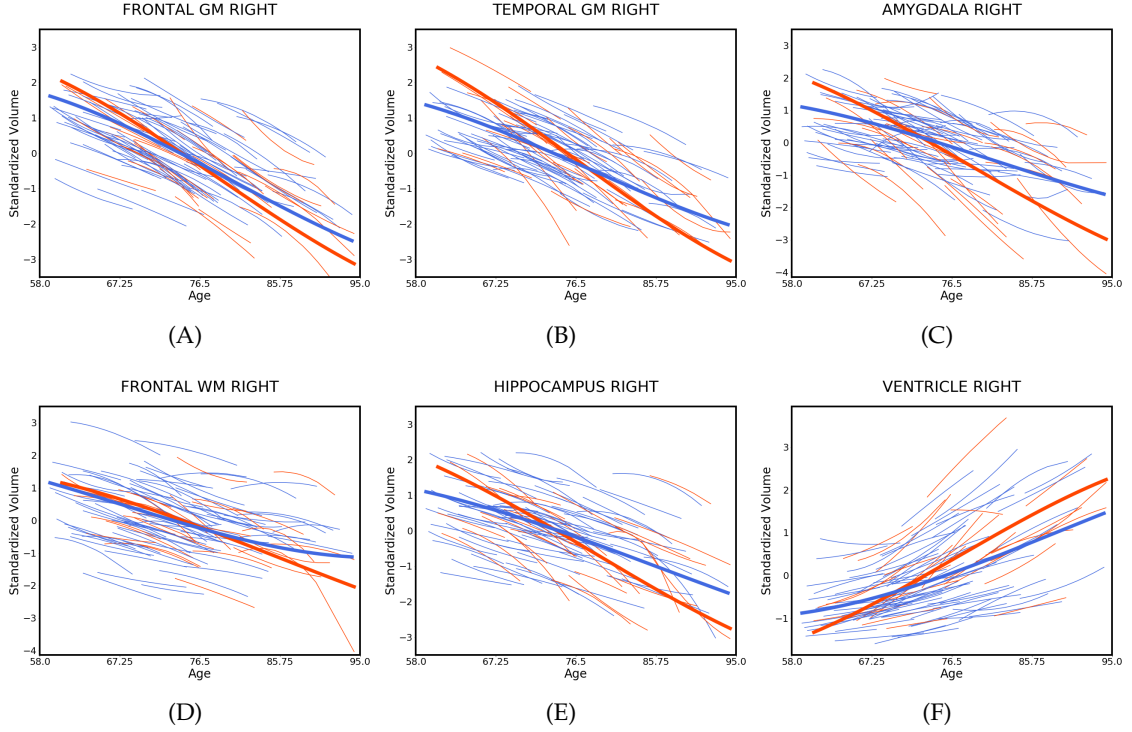


Figure 4.9: Estimated trajectories for six different ROIs. The two identified subgroups are colored red ($N=29$) and blue ($N=73$), respectively. Subgroup trajectories are thick, while the trajectories for each participant are thin.

right hemisphere to illustrate the longitudinal aging patterns. In order to quantify the differences of the trajectories between the two subgroups, we defined a distance statistic as the median of the absolute difference between the two trajectories. To estimate the null distribution of the statistic, we proposed a nonparametric permutation test. In each permutation, we first partitioned the entire dataset into two random groups, and then fitted the individuals in each group using our approach with $K = 1$ and $\eta = 0.01$ to derive the null hypothesis (i.e., a random fit for two global trajectories). Given the null distribution estimated from 1000 permutations, we obtained the p values of the statistic between the trajectories of G1 and G2. The ROIs for which the differences between the two groups were statistically significant (False discovery rate (FDR) corrected $p < 0.01$) are

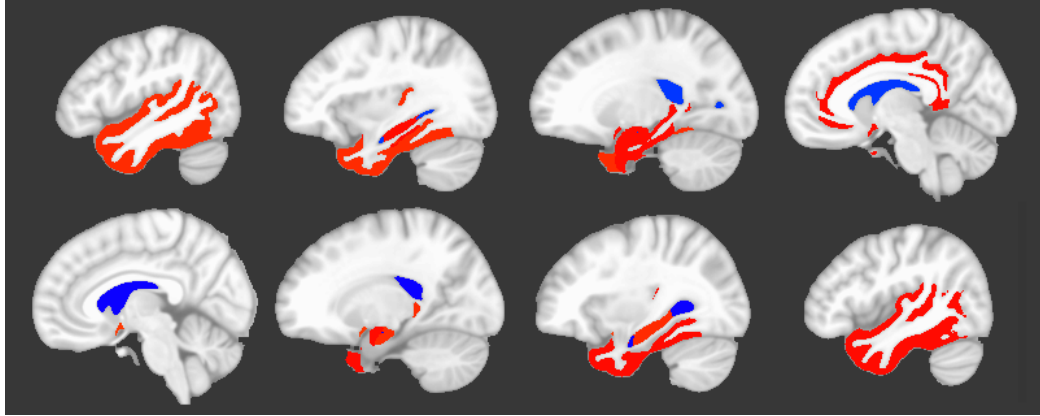


Figure 4.10: The brain regions that follow statistically significant (FDR corrected p value < 0.01) different trajectories between the two estimated subgroups found are shown. Red indicates G1 has a faster shrinking rate than G2, while blue indicates G1 has a faster expansion rate than G2.

shown in Figure 4.10, where red indicates that G1 has a higher rate of volume decrease than G2, and blue indicates the opposite. This nonparametric test shows that G1 is characterized by a significantly faster volume decrease than G2 in the temporal grey matter as well as in subcortical structures such as amygdala and hippocampus. At the same time, G1 is characterized by a faster ventricular expansion than G2.

To further analyze the difference between the two identified subgroups, we compared cognitive performance across clusters. Associations between cluster membership and cognitive change over time were examined using linear mixed effect models (R version 3.3.2 [122]). Age, sex, subgroup, interaction between age and sex, as well as interaction between age and subgroup, were included as fixed covariates. Random effects included intercept, age, and subject ID. The Table 4.1 summarizes the regression results. Significant associations between subgroups and cognitive change were found for the three California Verbal Learning Tasks (CVLtca, CVLfrs and CVLfrl) [33] and FLUCat [3]. For these cognitive tests, G1 was characterized by a faster decline than G2.

Table 4.1: Longitudinal cognitive score difference between two subgroups

	p value		p value
CVLtca	0.040*	CVLfrs	0.042*
CVLfri	0.004**	BVRTot	0.560
CRDRot	0.613	FLUCat	<0.0001***
FLULet	0.295		

4.4 Discussion and conclusion

In this chapter, we proposed a novel framework for parsing heterogeneity in longitudinal studies by simultaneously fitting multiple multivariate imaging trajectories, and clustering individuals. Our approach is able to delineate global trajectories using the partial information from individuals, while modeling explicitly heterogeneity in longitudinal processes. The trajectories are modeled as spatiotemporally regularized cubic B-splines, which allows for the nonlinearities and avoids overfitting.

By conducting simulations under multiple conditions, we showed that HELIOS has a high tolerance to noise, and performs reasonably well even when the length of the follow-up is small compared to the age range of the study. When we applied HELIOS to a BLSA data, two subgroups were found. Subgroup G1 was characterized by an imaging pattern of accelerated volume loss in regions that have been associated with neurodegenerative symptoms in aging [74, 124, 129, 147], as well as in early stage of Alzheimer’s disease [75, 89]. Moreover, individuals part of G1 performed statistically significantly worse than G2 in CVLT and FLUCat tests. These results complement the neuroanatomical profiles, and are in line with previous results associating increased atrophy in temporal pole with decreased performance in CVLT [31, 34, 96]. Taking together, these evidence suggest that G1 is likely to be an advanced aging group, or at the prodromal stages of neurodegenera-

tive disease, while G2 is more likely to be a resilient aging group, which follows a normal aging process. Additional analyses are required to investigate this hypothesis in the future, where a longer follow up would be crucial for boosting our confidence in the results.

Though longer follow up could allow us to better discern subtle brain changes from the confounding variability introduced by the processing, collection of longitudinal data is expensive and time consuming, which limits the potential use of HELIOS. Nonetheless, HELIOS is a general method that can handle any type of sequential data. Thus, it can be readily adapted to study the temporal dynamics in other types of data, such as time series data from resting-state functional magnetic resonance imaging (rsfMRI), where the brain activation of a heterogeneous population could be examined and analyzed.

Some limitations should be noted despite the advantages of this novel sophisticated method. The large number of parameters limits our current study to the level of ROIs. Specifically, the number of parameters in the spline coefficient tensor \mathbf{C} is in the order of imaging features P . However, the intrinsic dimensionality of the problem is much lower. As observed in our experiments, the corresponding regions in left and right hemisphere exhibit similar trajectories. Thus, a promising way to address the dimensionality problem is to adopt a sparse representation of \mathbf{C} , where we assume that there is a low rank decomposition of \mathbf{C} such that all the coefficient C_{pk} can be represented using a few bases. This approach may reduce the number of parameters significantly, leading to a model with lower complexity. Alternatively, this problem might be addressed by proposing an end-to-end framework which integrates the segmentation of high dimensional images with the characterization of heterogeneous trajectories.

Despite the aforementioned limitations, we were able to demonstrate that the proposed

method was able to dissect heterogeneity in both simulated and real longitudinal data. As there is an increasing number of longitudinal studies, HELIOS will become increasingly pertinent. Importantly, as current studies continue to collect data, HELIOS would benefit from the longer follow-ups to better estimate group trajectories and characterize the underlying heterogeneity. As a consequence, HELIOS can largely complement current literature and expand our understanding of longitudinal neurobiological processes.

Chapter 5

Summary and Future Work

5.1 Summary

The increasing evidence for heterogeneity in neurodegenerative diseases and neuropsychiatric disorders underlines the need for accurate categorization of patients in relatively homogeneous subgroup towards deepening our understanding of neurobiological mechanisms and improving disease diagnosis. The aim of this dissertation is to address the problem of disentangling heterogeneity of brain processes in neuroimaging studies. Depending on the study design, the heterogeneity can be addressed in two ways: 1) by learning the heterogeneous disease effect via group analysis using cross-sectional datasets, and 2) by delineating the heterogeneous (patho)physiological process that drives the differentiation of subgroups using longitudinal datasets. For each of these settings, we proposed analytical tools, termed CHIMERA and HELIOS respectively, which elucidate underlying heterogeneity, as this is reflected by imaging measures. These methods leverage machine learning approaches to best exploit the rich information provided by advanced neuroimaging

techniques, toward paving the way for precision medicine and therapeutic innovations.

In Chapter 2, we proposed CHIMERA, which is a semi-supervised clustering method. CHIMERA models two populations (e.g. patients and controls, responders to treatment and non-responders) as high-dimensional point distributions, and aligns them by solving a probability density estimation problem, thus accounting for inter-subject variability. Multiple transformations are estimated to model the underlying heterogeneity. The effect of covariates is taken explicitly into account in the matching criterion, thus reducing the effect of confounding variations. In order to fully explore the capability of CHIMERA, we applied it to a large Alzheimer’s disease cohorts in Chapter 3. Four distinct reproducible subgroups with significantly different imaging signatures were delineated. We conducted detailed analyses evaluating the differences between the four subgroups on the difference of cognitive scores, CSF biomarkers, white matter hyperintensities load, as well as other available demographic and clinical variables. Our findings largely complement the current literature of heterogeneity analysis of Alzheimer’s disease.

In Chapter 4, we proposed HELIOS, which aims to tackle heterogeneity in longitudinal designs. We expect heterogeneous neurobiological processes in neurodegenerative and psychiatric disorders to influence patients across their life-span, leading them to exhibit distinct longitudinal neuroanatomical trajectories. We proposed HELIOS for disentangling this heterogeneity by simultaneously fitting non-linear global trajectories and clustering individuals according to their multivariate trajectory patterns. HELIOS was applied to a large longitudinal study of aging where two subgroups were found. The subgroups were characterized by distinct imaging trajectories and significantly different performance in cognitive tests.

5.2 Future work

The methods described in this thesis constitute a new paradigm for data-driven analysis of heterogeneity in neuroimaging. There are several avenues for improving and extending these tools which are left to investigate for future work. We detail below a number of these avenues.

1. High dimensional imaging features.

In this thesis, the number of parameters to be optimized by the proposed methods is related to the number of features used. If we denote the number of features by n , then CHIMERA has $O(n^2)$ parameters and HELIOS has $O(n)$ parameters. In medical imaging analysis, the sample size is usually small compared to the number of features, which makes it difficult to optimize for a large number of parameters. Therefore, our analysis across the thesis was limited to using a partition of the brains in regions of interest, which give a coarse description of brain anatomy. Leveraging the power of high resolution MR images by exploiting the rich voxel-wise information could refine our characterization of disease subtypes and reveal previously unappreciated, subtle alterations in the brain structure. Toward this end, investigating sophisticated regularization strategies is a promising strategy. The ℓ_2 norm, as well as the Frobenius norm, which are currently used, put equal emphasis on all the parameters. As a consequence, they are quite sensitive to the number of parameters. However, it is possible to significantly reduce the degrees of freedom of the parameter space by introducing some prior knowledge to the parameters, such as spatial regularization for CHIMERA and sparse representation for HELIOS. Further im-

provements may also be achieved by exploring kernelized similarity measures for HELIOS, which could provide a dual form of the optimization problem, making feasible the inference in high-dimensional settings.

2. Reducing variations in longitudinal studies.

Longitudinal studies aim to detect subtle brain changes across time. However, these changes are often confounded by variations introduced by different imaging scanning protocols or preprocessing pipelines. Toward reducing confounding variations, a number of 4D methods [159, 166] have been proposed for longitudinal analysis. These methods typically operate by regularizing the temporal variations of tissue segmentation. However, this enforced temporal smoothness may jeopardize the heterogeneity analysis by reducing the underlying signal of interest. To overcome this problem, one promising approach is to propose an end-to-end framework, which unifies the segmentation process and the longitudinal heterogeneity analysis. This approach would potentially optimize for delineating heterogeneous temporal dynamics in the population, while also regularizing the temporal smoothness for each individual.

3. Estimating the statistical power of the clusters.

Our suite of tools (i.e., CHIMERA and HELIOS) allow for the heterogeneous population to be grouped into a user-specified number of different subgroups. In order to quantify the imaging signature of each subgroup, we conducted voxel based analysis (VBA) for comparing the patients in each subgroup with the normal controls. A false discovery rate (FDR) correction was applied to the statistical p value maps to determine the significance level for each voxel after accounting for the multiple

comparisons. In our analysis, we have yet to provide the multiple comparison correction on the cluster level. In other words, by clustering the patient population, we obtained by construction subgroups that are different from each other in terms of imaging presentations, and we discussed these differences based on the VBA. However, the VBA results should be discussed with care given the absence of a rigorous procedure for correcting p values in a cluster wise manner, acknowledging that differences at a certain level may be introduced by the clustering itself. This problem could be partially addressed by performing VBA between each patient subgroup and the corresponding group of normal controls, i.e., healthy subjects who undergo the same estimated pathologic transformation towards the patient distribution during the distribution matching process. Methods that can fully uncover the statistical power of clusters remain an open question to be investigated.

4. Applying HELIOS to time series data.

HELIOS was currently applied to longitudinal studies, where each subject often has a follow-up interval of several months. The collection of a longitudinal dataset is expensive and time-consuming, which limits the use of HELIOS. On the other hand, HELIOS is a general method that can readily handle time series data, such as the ones produced by resting-state functional magnetic resonance imaging (rsfMRI), electroencephalogram (EEG) and magnetoencephalography (MEG). As a consequence, HELIOS could be useful for studying the temporal dynamics of functional activation and how these vary across populations that are heterogeneous, uniquely contributing to this topic of growing interest.

Appendices

Appendix A

Software

CHIMERA is an open-source software, implemented using Python 2.7. The software package is available on NITRC¹. It takes a comma separated values (csv) file as input. This file contains imaging features of the subjects, covariate information (optional), as well as group information (i.e., whether the subject is a control or a patient). The program performs clustering using multiple random initializations, and returns the most reproducible clustering result. Out of sample subtyping can be performed using the transformation learned and stored by the main function. A snapshot of the CHIMERA command-line is shown in Figure A.1.

HELIOS is also available as an open-source software, implemented using Python 2.7. It takes a csv file as input containing the subject ID, age at scan, and ROI features. It can also take as input the spatial locations of the ROIs in order to perform the spatial regularization. The program produces clustering memberships as well as global longitudinal trajectories. Multiple RANSAC [45] initializations are performed, and the best fitting is reported as the

¹https://www.nitrc.org/projects/cbica_chimera/

```

CHIMERA--
Clustering heterogenous disease effects via distribution matching of imaging patterns.

*** For best performance, pick optimal -m, -n values based on cross-validation ***

Usage: chimera [OPTIONS]

Required:
[-i <string> ]   File name of data in .csv format with header line, one subject per row
                  Header line can have types of:
                  ID      : subject ID (optional)
                  Group   : subject diagnosis(1: patient; 0: normal control) (required)
                  Set     : subject dataset ID (optional)
                  COVAR   : covariates (optional)
                  IMG     : imaging features (required)
[-r <string> ]   File name of clustering outputs
[-k <int>  ]     Number of sub-groups to find

Options: (See reference for more details)
[-m <float> ]   The value of lambda1, for b (non-negative, default 10)
[-n <float> ]   The value of lambda2, for A (non-negative, default 100)
[-T <string> ]   Transformation to be used. (affine/duo/trans, default affine)
[-e <float> ]   Stopping criterion tolerance (default 0.001)
[-M <int>  ]   Maximum iteration allowed (default 1000)
[-N <int>  ]   Number of Runs of optimization with different initialization (default 50)
[-o <string> ]   File name to save trained model (default not saving)
[-s <float> ]   Weight of covariates distance (default auto)
[-t <float> ]   Weight of set distance (default 10)
[-w <int>  ]   Data normalization. 0: no action; 1: 0-1 normalization; 2: z-score. (default 1)
[-c <int>  ]   Save most reproducible (c=1)/ minimal energy (c=2) result among runs (default 2)
[-v --verbose] Verbose output

[-u --usage ]   Display this message
[-h --help  ]   Display this message
[-V --Version] Display version information

Example:
chimera -i data.csv -r output.txt -k 3 -o model.cpk1 -N 10 -m 20 -v

```

Figure A.1: Command-line interface of CHIMERA.

```

HELIOS--
Parsing the Heterogeneity of Longitudinal Imaging through Integrated Clustering
and Spatiotemporally Regularized Spline Curve Fitting
*** For best performance, pick optimal -m, -l values based on cross-validation ***

Usage: helios [OPTIONS]

Required:
[-i <string> ]   File name of data in .csv format with header line, one subject per row
                  Header line can have types of:
                  ID      : subject ID (required)
                  IMG     : imaging features (required)
                  AGE     : subject age at each scan (required)
[-r <string> ]   File name of clustering outputs
[-k <int>  ]     Number of sub-groups to find

Options: (See reference for more details)
[-s <string> ]   File name of spatial locations of imaging features (if not provided, program
                  will use an identity matrix, equivalent to no spatial regularization)
[-m <float> ]    The value of eta (non-negative, default 0.1)
[-l <float> ]    The value of lambda (non-negative, default 1)
[-e <float> ]    Stopping criterion tolerance (default 0.01)
[-M <int>  ]     Maximum iteration allowed (default 1000)
[-N <int>  ]     Number of Runs of optimization with different initialization (default 10)
[-o <string> ]   File name to save trained model (default not saving)
[-v --verbose ]  Verbose output

[-u --usage ]    Display this message
[-h --help ]     Display this message
[-V --Version ]  Display version information

Example:
helios -i data.csv -r output.txt -k 3 -o model.cpk1 -m 0.01 -v

```

Figure A.2: Command-line interface of HELIOS.

result. A snapshot of the HELIOS command-line is shown in Figure A.2.

Appendix B

List of regions of interest

We have used different sets of multi-atlas segmentation of ROIs [37] in Chapter 3 and Chapter 4. Each set corresponded to a different level of brain organization hierarchy. The names of the ROIs are listed in Table B.1 and Table B.2, respectively.

#	Name	#	Name
1	3rd Ventricle	41	Parietal Lobe WM L
2	4th Ventricle	42	Temporal Lobe WM R
3	Accumbens Area R	43	Temporal Lobe WM L
4	Accumbens Area L	44	Fornix R
5	Amygdala R	45	Fornix L
6	Amygdala L	46	Anterior Limb Intern. Capsule R
7	Brain Stem	47	Anterior Limb Intern. Capsule L
8	Caudate R	48	Posterior Limb Intern. Capsule R
9	Caudate L	49	Posterior Limb Intern. Capsule L
10	Cerebellum Exterior R	50	Corpus Callosum
11	Cerebellum Exterior L	51	Frontal Inferior GM L
12	Cerebellum WM R	52	Frontal Insular GM L
13	Cerebellum WM L	53	Frontal Lateral GM L
14	CSF	54	Frontal Medial GM L
15	Hippocampus R	55	Frontal Opercular GM L
16	Hippocampus L	56	Limbic Cingulate GM L
17	Inf Lat Vent R	57	Limbic Medialtemporal GM L
18	Inf Lat Vent L	58	Occipital Inferior GM L
19	Lateral Ventricle R	59	Occipital Lateral GM L
20	Lateral Ventricle L	60	Occipital Medial GM L
21	Pallidum R	61	Parietal Lateral GM L
22	Pallidum L	62	Parietal Medial GM L
23	Putamen R	63	Temporal Inferior GM L
24	Putamen L	64	Temporal Lateral GM L
25	Thalamus Proper R	65	Temporal Supratemporal GM L
26	Thalamus Proper L	66	Frontal Inferior GM R
27	Ventral DC R	67	Frontal Insular GM R
28	Ventral DC L	68	Frontal Lateral GM R
29	Vessel R	69	Frontal Medial GM R
30	Vessel L	70	Frontal Opercular GM R
31	Cere. Vermal Lob. 1-5	71	Limbic Cingulate GM R
32	Cere. Vermal Lob. 6-7	72	Limbic Medialtemporal GM R
33	Cere. Vermal Lob. 8-10	73	Occipital Inferior GM R
34	Basal Forebrain L	74	Occipital Lateral GM R
35	Basal Forebrain R	75	Occipital Medial GM R
36	Frontal Lobe WM R	76	Parietal Lateral GM R
37	Frontal Lobe WM L	77	Parietal Medial GM R
38	Occipital Lobe WM R	78	Temporal Inferior GM R
39	Occipital Lobe WM L	79	Temporal Lateral GM R
40	Parietal Lobe WM R	80	Temporal Supratemporal GM R

GM: gray matter; WM: white matter; R: right; L: left.

Table B.1: Names of 80 ROIs used in Chapter 3.

#	Name	#	Name
1	Basal Ganglia L	18	Limbic GM R
2	Deep GM L	19	Occipital GM R
3	Deep WM L	20	Occipital WM R
4	Frontal GM L	21	Parietal GM R
5	Frontal WM L	22	Parietal WM R
6	Limbic GM L	23	Temporal GM R
7	Occipital GM L	24	Temporal WM R
8	Occipital WM L	25	Amygdala R
9	Parietal GM L	26	Amygdala L
10	Parietal WM L	27	Hippocampus R
11	Temporal GM L	28	Hippocampus L
12	Temporal WM L	29	Basal Forebrain R
13	Basal Ganglia R	30	Basal Forebrain L
14	Deep GM R	31	Cerebellum R
15	Deep WM R	32	Cerebellum L
16	Frontal GM R	33	Ventricle R
17	Frontal WM R	34	Ventricle L

GM: gray matter; WM: white matter; R: right; L: left.

Table B.2: Names of 34 ROIs used in Chapter 4.

Appendix C

BLSA cognitive data

The BLSA cognitive data we have used in Chapter 4 are listed below:

1. California Verbal Learning Task: 1) total of 4 list A trials (CVLtca), 2) short delay free recall (CVLfrs), and 3) long delay free recall (CVLfrl) are used to assess verbal learning and memory. Higher values indicate better cognitive performance.
2. Benton Visual Retention Test (BVRTot) quantifies figural memory and visuo constructional ability. Lower values indicate better cognitive performance.
3. CARD Rotation Test (CRDRot) measures the ability to mentally manipulate figures. Higher values indicate better cognitive performance.
4. Category Fluency (FLUCat) measures semantic fluency. Higher values indicate better cognitive performance.
5. Letter Fluency (FLULet) measures phonemic fluency. Higher values indicate better cognitive performance.

Bibliography

- [1] Abdi, H., Williams, L.J.: Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2(4), 433–459 (2010)
- [2] Abraham, C., Cornillon, P.A., Matzner-Løber, E., Molinari, N.: Unsupervised curve clustering using B-splines. *Scandinavian journal of statistics* 30(3), 581–595 (2003)
- [3] Acevedo, A., Loewenstein, D.A., Barker, W.W., Harwood, D.G., Luis, C., Bravo, M., Hurwitz, D.A., Agüero, H., Greenfield, L., Duara, R.: Category fluency test: normative data for English-and Spanish-speaking elderly. *Journal of the International Neuropsychological Society* 6(7), 760–769 (2000)
- [4] Alladi, S., Xuereb, J., Bak, T., Nestor, P., Knibb, J., Patterson, K., Hodges, J.: Focal cortical presentations of Alzheimer’s disease. *Brain* 130(10), 2636–2645 (2007)
- [5] Arvanitakis, Z., Leurgans, S.E., Barnes, L.L., Bennett, D.A., Schneider, J.A.: Microinfarct pathology, dementia, and cognitive systems. *Stroke* 42(3), 722–727 (2011)
- [6] Ashburner, J., Friston, K.J.: Voxel-based morphometrythe methods. *Neuroimage* 11(6), 805–821 (2000)

- [7] Ashburner, J., Klöppel, S.: Multivariate models of inter-subject anatomical variability. *Neuroimage* 56(2), 422–439 (2011)
- [8] Assaf, Y., Pasternak, O.: Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *Journal of molecular neuroscience* 34(1), 51–61 (2008)
- [9] Association, A., et al.: 2017 Alzheimer’s disease facts and figures. *Alzheimer’s & Dementia* 13(4), 325–373 (2017)
- [10] Baron, J., Chetelat, G., Desgranges, B., Perchey, G., Landeau, B., De La Sayette, V., Eustache, F.: In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer’s disease. *Neuroimage* 14(2), 298–309 (2001)
- [11] Beckmann, C.F., Smith, S.M.: Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE transactions on medical imaging* 23(2), 137–152 (2004)
- [12] Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R., Initiative, A.D.N., et al.: Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage* 66, 249–260 (2013)
- [13] Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., Sabuncu, M.R., Initiative, A.D.N., et al.: Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data. *Neuroimage* 81, 358–370 (2013)

- [14] Bilgel, M., Prince, J.L., Wong, D.F., Resnick, S.M., Jernigan, B.M.: A multivariate non-linear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage* 134, 658–670 (2016)
- [15] Bishop, C.M.: *Pattern recognition and machine learning*. springer (2006)
- [16] Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H., Del Tredici, K.: Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta neuropathologica* 112(4), 389–404 (2006)
- [17] Brickman, A.M., Habeck, C., Zarahn, E., Flynn, J., Stern, Y.: Structural MRI covariance patterns associated with normal aging and neuropsychological functioning. *Neurobiology of aging* 28(2), 284–295 (2007)
- [18] Busatto, G.F., Diniz, B.S., Zanetti, M.V.: Voxel-based morphometry in Alzheimers disease. *Expert review of neurotherapeutics* 8(11), 1691–1702 (2008)
- [19] Chetelat, G., Landeau, B., Eustache, F., Mezenge, F., Viader, F., De La Sayette, V., Desgranges, B., Baron, J.C.: Using voxel-based morphometry to map the structural changes associated with rapid conversion in mci: a longitudinal mri study. *Neuroimage* 27(4), 934–946 (2005)
- [20] Chung, M., Worsley, K., Paus, T., Cherif, C., Collins, D., Giedd, J., Rapoport, J., Evans, A.: A unified statistical approach to deformation-based morphometry. *NeuroImage* 14(3), 595–606 (2001)
- [21] Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G.a., Roses, A., Haines, J., Pericak-Vance, M.A.: Gene dose of apolipoprotein E type 4

- allele and the risk of Alzheimers disease in late onset families. *Science* 261(5123), 921–923 (1993)
- [22] Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* 20(3), 273–297 (1995)
- [23] Cox, R.W.: AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research* 29(3), 162–173 (1996)
- [24] Crane, P.K., Carle, A., Gibbons, L.E., Insel, P., Mackin, R.S., Gross, A., Jones, R.N., Mukherjee, S., Curtis, S.M., Harvey, D., et al.: Development and assessment of a composite score for memory in the Alzheimers Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior* 6(4), 502–516 (2012)
- [25] Crutch, S.J., Lehmann, M., Schott, J.M., Rabinovici, G.D., Rossor, M.N., Fox, N.C.: Posterior cortical atrophy. *The Lancet Neurology* 11(2), 170–178 (2012)
- [26] Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehericy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N., 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 (2011)
- [27] Da, X., Toledo, J.B., Zee, J., Wolk, D.A., Xie, S.X., Ou, Y., Shacklett, A., Parmpi, P., Shaw, L., Trojanowski, J.Q., et al.: Integration and relative value of biomarkers for prediction of MCI to AD progression: Spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. *NeuroImage: Clinical* 4, 164–173 (2014)

- [28] Davatzikos, C.: Mapping image data to stereotaxic spaces: applications to brain mapping. *Human Brain Mapping* 6(5-6), 334–338 (1998)
- [29] Davatzikos, C., Genc, A., Xu, D., Resnick, S.M.: Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. *NeuroImage* 14(6), 1361–1369 (2001)
- [30] Davatzikos, C., Resnick, S.M.: Degenerative age changes in white matter connectivity visualized in vivo using magnetic resonance imaging. *Cerebral cortex* 12(7), 767–771 (2002)
- [31] Davatzikos, C., Xu, F., An, Y., Fan, Y., Resnick, S.M.: Longitudinal progression of Alzheimer’s-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain* 132(8), 2026–2035 (2009)
- [32] DeCarli, C., Fletcher, E., Ramey, V., Harvey, D., Jagust, W.J.: Anatomical mapping of white matter hyperintensities (WMH) exploring the relationships between periventricular WMH, deep WMH, and total WMH burden. *Stroke* 36(1), 50–55 (2005)
- [33] Delis, D.C.: CVLT-II: California verbal learning test: adult version. Psychological Corporation (2000)
- [34] Delis, D.C., Massman, P.J., Butters, N., Salmon, D.P., Cermak, L.S., Kramer, J.H.: Profiles of demented and amnesic patients on the California Verbal Learning Test: Implications for the assessment of memory disorders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology* 3(1), 19 (1991)

- [35] Dice, L.R.: Measures of the amount of ecologic association between species. *Ecology* 26(3), 297–302 (1945)
- [36] Dong, A., Honnorat, N., Gaonkar, B., Davatzikos, C.: CHIMERA: clustering of heterogeneous disease effects via distribution matching of imaging patterns. *IEEE transactions on medical imaging* 35(2), 612–621 (2016)
- [37] Doshi, J., Erus, G., Ou, Y., Gaonkar, B., Davatzikos, C.: Multi-atlas skull-stripping. *Academic radiology* 20(12), 1566–1576 (2013)
- [38] Doshi, J., Erus, G., Ou, Y., Resnick, S.M., Gur, R.C., Gur, R.E., Satterthwaite, T.D., Furth, S., Davatzikos, C., Initiative, A.N., et al.: MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage* 127, 186–195 (2016)
- [39] Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T.: Efficient projections onto the l_1 -ball for learning in high dimensions. In: *Proceedings of the 25th international conference on Machine learning*. pp. 272–279. ACM (2008)
- [40] Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P., et al.: The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer’s disease. *International Psychogeriatrics* 21(4), 672–687 (2009)
- [41] Engl, H.W., Hanke, M., Neubauer, A.: *Regularization of inverse problems*, vol. 375. Springer Science & Business Media (1996)

- [42] Fan, Y., Resnick, S.M., Wu, X., Davatzikos, C.: Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage* 41(2), 277–285 (2008)
- [43] Fenton, W.S., McGlashan, T.H., Victor, B.J., Blyler, C.R.: Symptoms, subtype, and suicidality in patients with schizophrenia spectrum disorders. *American journal of psychiatry* 154(2), 199–204 (1997)
- [44] Ferreira, L.K., Diniz, B.S., Forlenza, O.V., Busatto, G.F., Zanetti, M.V.: Neurostructural predictors of Alzheimer's disease: a meta-analysis of VBM studies. *Neurobiology of aging* 32(10), 1733–1741 (2011)
- [45] Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), 381–395 (1981)
- [46] van der Flier, W.M., Pijnenburg, Y.A., Fox, N.C., Scheltens, P.: Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ϵ 4 allele. *The Lancet Neurology* 10(3), 280–288 (2011)
- [47] Fotenos, A.F., Snyder, A., Girton, L., Morris, J., Buckner, R.: Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology* 64(6), 1032–1039 (2005)
- [48] Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J., Rossor, M.N.: Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer dis-

- ease: power calculations and estimates of sample size to detect treatment effects. *Archives of Neurology* 57(3), 339–344 (2000)
- [49] Friston, K., Frith, C., Liddle, P., Frackowiak, R.: Functional connectivity: the principal-component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow & Metabolism* 13(1), 5–14 (1993)
- [50] Fusar-Poli, P., Radua, J., McGuire, P., Borgwardt, S.: Neuroanatomical maps of psychosis onset: voxel-wise meta-analysis of antipsychotic-naïve VBM studies. *Schizophrenia bulletin* 38(6), 1297–1307 (2011)
- [51] Gaffney, S., Smyth, P.: Trajectory clustering with mixtures of regression models. In: *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. pp. 63–72. ACM (1999)
- [52] Gaffney, S.J., Smyth, P.: Joint probabilistic curve clustering and alignment. In: *Advances in neural information processing systems*. pp. 473–480 (2005)
- [53] Gaonkar, B., Pohl, K., Davatzikos, C.: Pattern based morphometry. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 459–466. Springer (2011)
- [54] Garcia-Closas, M., Hall, P., Nevanlinna, H., Pooley, K., Morrison, J., Richesson, D.A., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Arias, J.I., et al.: Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS genetics* 4(4), e1000054 (2008)

- [55] Gaser, C., Volz, H.P., Kiebel, S., Riehemann, S., Sauer, H.: Detecting structural changes in whole brain based on nonlinear deformationsapplication to schizophrenia research. *Neuroimage* 10(2), 107–113 (1999)
- [56] Gibbons, L.E., Carle, A.C., Mackin, R.S., Harvey, D., Mukherjee, S., Insel, P., Curtis, S.M., Mungas, D., Crane, P.K., Initiative, A.D.N., et al.: A composite score for executive functioning, validated in Alzheimers Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain imaging and behavior* 6(4), 517–527 (2012)
- [57] Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L.: Brain development during childhood and adolescence: a longitudinal MRI study. *Nature neuroscience* 2(10), 861–863 (1999)
- [58] Goldszal, A.F., Davatzikos, C., Pham, D.L., Yan, M.X., Bryan, R.N., Resnick, S.M.: An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of computer assisted tomography* 22(5), 827–837 (1998)
- [59] Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N., Friston, K.J., Frackowiak, R.S.: A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *NeuroImage* 14(1), 21–36 (2001)
- [60] Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. *NeuroImage* 45(1), S133–S142 (2009)

- [61] Gordon, E., Rohrer, J., Kim, L., Omar, R., Rossor, M., Fox, N., Warren, J.: Measuring disease progression in frontotemporal lobar degeneration A clinical and MRI study. *Neurology* 74(8), 666–673 (2010)
- [62] Gorno-Tempini, M., Brambati, S., Ginex, V., Ogar, J., Dronkers, N., Marcone, A., Perani, D., Garibotto, V., Cappa, S., Miller, B.: The logopenic/phonological variant of primary progressive aphasia. *Neurology* 71(16), 1227–1234 (2008)
- [63] Habes, M., Erus, G., Toledo, J.B., Zhang, T., Bryan, N., Launer, L.J., Rosseel, Y., Janowitz, D., Doshi, J., Van der Auwera, S., et al.: White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain* p. aww008 (2016)
- [64] Hansen, L.K., Larsen, J., Nielsen, F.Å., Strother, S.C., Rostrup, E., Savoy, R., Lange, N., Sidtis, J., Svarer, C., Paulson, O.B.: Generalizable patterns in neuroimaging: How many principal components? *NeuroImage* 9(5), 534–544 (1999)
- [65] Hebert, L.E., Weuve, J., Scherr, P.A., Evans, D.A.: Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology* 80(19), 1778–1783 (2013)
- [66] Ho, B.C., Andreasen, N.C., Nopoulos, P., Arndt, S., Magnotta, V., Flaum, M.: Progressive structural brain abnormalities and their relationship to clinical outcome: a longitudinal magnetic resonance imaging study early in schizophrenia. *Archives of general Psychiatry* 60(6), 585–594 (2003)

- [67] Honea, R., Crow, T.J., Passingham, D., Mackay, C.E.: Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *American Journal of Psychiatry* 162(12), 2233–2245 (2005)
- [68] Hubert, L., Arabie, P.: Comparing partitions. *Journal of classification* 2(1), 193–218 (1985)
- [69] Hurd, M.D., Martorell, P., Delavande, A., Mullen, K.J., Langa, K.M.: Monetary costs of dementia in the United States. *New England Journal of Medicine* 368(14), 1326–1334 (2013)
- [70] Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis, vol. 46. John Wiley & Sons (2004)
- [71] Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Feldman, H.H., Frisoni, G.B., Hampel, H., Jagust, W.J., Johnson, K.A., Knopman, D.S., et al.: A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology* 87(5), 539–547 (2016)
- [72] Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L Whitwell, J., Ward, C., et al.: The Alzheimer’s disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* 27(4), 685–691 (2008)
- [73] Jack, C.R., Petersen, R.C., Grundman, M., Jin, S., Gamst, A., Ward, C.P., Sencakova, D., Doody, R.S., Thal, L.J., (ADCS, A.D.C.S., et al.: Longitudinal mri findings from

- the vitamin e and donepezil treatment study for mci. *Neurobiology of aging* 29(9), 1285–1295 (2008)
- [74] Jack, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E.: Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* 49(3), 786–794 (1997)
- [75] Jack, C.R., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease. *Neurology* 51(4), 993–999 (1998)
- [76] Jack, C.R., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Lowe, V., Kantarci, K., Bernstein, M.A., Senjem, M.L., Gunter, J.L., et al.: Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. *Archives of neurology* 69(7), 856–867 (2012)
- [77] Jack, C.R., Wiste, H.J., Weigand, S.D., Knopman, D.S., Mielke, M.M., Vemuri, P., Lowe, V., Senjem, M.L., Gunter, J.L., Reyes, D., et al.: Different definitions of neurodegeneration produce similar amyloid/neurodegeneration biomarker group findings. *Brain* 138(12), 3747–3759 (2015)
- [78] Jeste, S.S., Geschwind, D.H.: Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nature Reviews Neurology* 10(2), 74–81 (2014)
- [79] Job, D.E., Whalley, H.C., McConnell, S., Glabus, M., Johnstone, E.C., Lawrie, S.M.: Structural gray matter differences between first-episode schizophrenics and normal controls using voxel-based morphometry. *Neuroimage* 17(2), 880–889 (2002)
- [80] Johnson, S.C.: Hierarchical clustering schemes. *Psychometrika* 32(3), 241–254 (1967)

- [81] Karas, G., Burton, E., Rombouts, S., Van Schijndel, R., O'Brien, J., Scheltens, P., McKeith, I., Williams, D., Ballard, C., Barkhof, F.: A comprehensive study of gray matter loss in patients with Alzheimers disease using optimized voxel-based morphometry. *Neuroimage* 18(4), 895–907 (2003)
- [82] Karas, G., Scheltens, P., Rombouts, S., Visser, P., Van Schijndel, R., Fox, N., Barkhof, F.: Global and local gray matter loss in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 23(2), 708–716 (2004)
- [83] Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourão-Miranda, J., Vemuri, P.: Diagnostic neuroimaging across diseases. *Neuroimage* 61(2), 457–463 (2012)
- [84] Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.: Automatic classification of MR scans in Alzheimer's disease. *Brain* 131(3), 681–689 (2008)
- [85] Knickmeyer, R.C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J.K., Hamer, R.M., Lin, W., Gerig, G., Gilmore, J.H.: A structural MRI study of human brain development from birth to 2 years. *Journal of Neuroscience* 28(47), 12176–12182 (2008)
- [86] Koutsouleris, N., Gaser, C., Jäger, M., Bottlender, R., Frodl, T., Holzinger, S., Schmitt, G.J., Zetzsche, T., Burgermeister, B., Scheuerecker, J., et al.: Structural correlates of psychopathological symptom dimensions in schizophrenia: a voxel-based morphometric study. *Neuroimage* 39(4), 1600–1612 (2008)
- [87] Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., et al.: Use of neu-

- roanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Archives of general psychiatry* 66(7), 700–712 (2009)
- [88] Kraemer, H.C., Yesavage, J.A., Taylor, J.L., Kupfer, D.: How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry* 157(2), 163–171 (2000)
- [89] Laakso, M., Soininen, H., Partanen, K., Helkala, E.L., Hartikainen, P., Vainio, P., Hallikainen, M., Hänninen, T., Riekkinen Sr, P.: Volumes of hippocampus, amygdala and frontal lobes in the MRI-based diagnosis of early Alzheimer’s disease: correlation with memory functions. *Journal of Neural Transmission-Parkinson’s Disease and Dementia Section* 9(1), 73–86 (1995)
- [90] Lam, B., Masellis, M., Freedman, M., Stuss, D.T., Black, S.E.: Clinical, imaging, and pathological heterogeneity of the Alzheimers disease syndrome. *Alzheimers Res Ther* 5(1), 1 (2013)
- [91] Lanckriet, G.R., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I.: Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research* 5, 27–72 (2004)
- [92] Lao, Z., Shen, D., Liu, D., Jawad, A.F., Melhem, E.R., Launer, L.J., Bryan, R.N., Davatzikos, C.: Computer-assisted segmentation of white matter lesions in 3D MR images using support vector machine. *Academic radiology* 15(3), 300–313 (2008)

- [93] Larner, A.: Getting it wrong: the clinical misdiagnosis of Alzheimer's disease. *International journal of clinical practice* 58(11), 1092–1094 (2004)
- [94] Lee, J.G., Han, J., Whang, K.Y.: Trajectory clustering: a partition-and-group framework. In: *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. pp. 593–604. ACM (2007)
- [95] Li, C., Gore, J.C., Davatzikos, C.: Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation. *Magnetic resonance imaging* 32(7), 913–923 (2014)
- [96] Liu, W., Miller, B.L., Kramer, J.H., Rankin, K., Wyss-Coray, C., Gearhart, R., Phengrasamy, L., Weiner, M., Rosen, H.J.: Behavioral disorders in the frontal and temporal variants of frontotemporal dementia. *Neurology* 62(5), 742–748 (2004)
- [97] Lloyd, S.: Least squares quantization in PCM. *IEEE transactions on information theory* 28(2), 129–137 (1982)
- [98] Marusyk, A., Polyak, K.: Tumor heterogeneity: causes and consequences. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1805(1), 105–117 (2010)
- [99] Mathalon, D.H., Sullivan, E.V., Lim, K.O., Pfefferbaum, A.: Progressive brain volume changes and the clinical course of schizophrenia in men: a longitudinal magnetic resonance imaging study. *Archives of general psychiatry* 58(2), 148–157 (2001)
- [100] McCullagh, P.: Generalized linear models. *European Journal of Operational Research* 16(3), 285–292 (1984)

- [101] McIntosh, A.R., Mišić, B.: Multivariate statistical analyses for neuroimaging data. *Annual review of psychology* 64, 499–525 (2013)
- [102] McKeith, I.G., Fairbairn, A.F., Perry, R.H., Thompson, P.: The clinical diagnosis and misdiagnosis of senile dementia of Lewy body type (SDLT). *The British Journal of Psychiatry* 165(3), 324–332 (1994)
- [103] McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M.: Clinical diagnosis of Alzheimer’s disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease. *Neurology* 34(7), 939–939 (1984)
- [104] Ment, L.R., Kesler, S., Vohr, B., Katz, K.H., Baumgartner, H., Schneider, K.C., Delancy, S., Silbereis, J., Duncan, C.C., Constable, R.T., et al.: Longitudinal brain volume changes in preterm and term control subjects during late childhood and adolescence. *Pediatrics* 123(2), 503–511 (2009)
- [105] Miller, B.J., Buckley, P., Seabolt, W., Mellor, A., Kirkpatrick, B.: Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biological psychiatry* 70(7), 663–671 (2011)
- [106] Misra, C., Fan, Y., Davatzikos, C.: Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* 44(4), 1415–1422 (2009)
- [107] Montagne, A., Barnes, S.R., Sweeney, M.D., Halliday, M.R., Sagare, A.P., Zhao, Z., Toga, A.W., Jacobs, R.E., Liu, C.Y., Amezcua, L., et al.: Blood-brain barrier break-

- down in the aging human hippocampus. *Neuron* 85(2), 296–302 (2015)
- [108] Moon, T.K.: The expectation-maximization algorithm. *IEEE Signal processing magazine* 13(6), 47–60 (1996)
- [109] Mugler, J.P., Brookeman, J.R.: Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine* 15(1), 152–157 (1990)
- [110] Murray, M.E., Graff-Radford, N.R., Ross, O.A., Petersen, R.C., Duara, R., Dickson, D.W.: Neuropathologically defined subtypes of Alzheimer’s disease with distinct clinical characteristics: a retrospective study. *The Lancet Neurology* 10(9), 785–796 (2011)
- [111] Myronenko, A., Song, X.: Point set registration: Coherent point drift. *PAMI* 32, 2262–2275 (2010)
- [112] Neal, R.M., Hinton, G.E.: A view of the EM algorithm that justifies incremental, sparse, and other variants. In: *Learning in graphical models*, pp. 355–368. Springer (1998)
- [113] Nenadic, I., Gaser, C., Sauer, H.: Heterogeneity of brain structural variation and the structural imaging endophenotypes in schizophrenia. *Neuropsychobiology* 66(1), 44–49 (2012)
- [114] Nettiksimmons, J., DeCarli, C., Landau, S., Beckett, L., Initiative, A.D.N., et al.: Biological heterogeneity in ADNI amnesic mild cognitive impairment. *Alzheimer’s & Dementia* 10(5), 511–521 (2014)

- [115] Noh, Y., Jeon, S., Lee, J.M., Seo, S.W., Kim, G.H., Cho, H., Ye, B.S., Yoon, C.W., Kim, H.J., Chin, J., et al.: Anatomical heterogeneity of Alzheimer disease Based on cortical thickness on MRIs. *Neurology* 83(21), 1936–1944 (2014)
- [116] Oakes, T.R., Fox, A.S., Johnstone, T., Chung, M.K., Kalin, N., Davidson, R.J.: Integrating VBM into the General Linear Model with voxelwise anatomical covariates. *Neuroimage* 34(2), 500–508 (2007)
- [117] Ossenkoppele, R., Pijnenburg, Y.A., Perry, D.C., Cohn-Sheehy, B.I., Scheltens, N.M., Vogel, J.W., Kramer, J.H., van der Vlies, A.E., La Joie, R., Rosen, H.J., et al.: The behavioural/dysexecutive variant of Alzheimers disease: clinical, neuroimaging and pathological features. *Brain* 138(9), 2732–2749 (2015)
- [118] Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C.: DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Medical image analysis* 15(4), 622–639 (2011)
- [119] Perrin, R.J., Fagan, A.M., Holtzman, D.M.: Multimodal techniques for diagnosis and prognosis of Alzheimer’s disease. *Nature* 461(7266), 916–922 (2009)
- [120] Petersen, R., Aisen, P., Beckett, L., Donohue, M., Gamst, A., Harvey, D., Jack, C., Jagust, W., Shaw, L., Toga, A., et al.: Alzheimer’s disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology* 74(3), 201–209 (2010)
- [121] Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology* 56(3), 303–308 (1999)

- [122] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2016)
- [123] Resnick, S., Sojkova, J., Zhou, Y., An, Y., Ye, W., Holt, D., Dannals, R., Mathis, C., Klunk, W., Ferrucci, L., et al.: Longitudinal cognitive decline is associated with fibrillar amyloid-beta measured by [11C] PiB. *Neurology* 74(10), 807–815 (2010)
- [124] Resnick, S.M., Pham, D.L., Kraut, M.A., Zonderman, A.B., Davatzikos, C.: Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *Journal of Neuroscience* 23(8), 3295–3301 (2003)
- [125] Reuter, M., Rosas, H.D., Fischl, B.: Highly accurate inverse consistent registration: a robust approach. *Neuroimage* 53(4), 1181–1196 (2010)
- [126] Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B.: Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61(4), 1402–1418 (2012)
- [127] Riddle, W.R., Li, R., Fitzpatrick, J.M., DonLevy, S.C., Dawant, B.M., Price, R.R.: Characterizing changes in MR images with color-coded Jacobians. *Magnetic resonance imaging* 22(6), 769–777 (2004)
- [128] Sabuncu, M.R., Desikan, R.S., Sepulcre, J., Yeo, B.T.T., Liu, H., Schmansky, N.J., Reuter, M., Weiner, M.W., Buckner, R.L., Sperling, R.A., et al.: The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Archives of neurology* 68(8), 1040–1048 (2011)

- [129] Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C.: A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of neurology* 60(7), 989–994 (2003)
- [130] Scahill, R., Fox, N.: Longitudinal imaging in dementia. *The British journal of radiology* (2014)
- [131] Schneider, J.A., Arvanitakis, Z., Bang, W., Bennett, D.A.: Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* 69(24), 2197–2204 (2007)
- [132] Schölkopf, B., Smola, A.J.: Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press (2002)
- [133] Schumann, C.M., Bloss, C.S., Barnes, C.C., Wideman, G.M., Carper, R.A., Akshoomoff, N., Pierce, K., Hagler, D., Schork, N., Lord, C., et al.: Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. *Journal of Neuroscience* 30(12), 4419–4427 (2010)
- [134] Schwarz, C., Fletcher, E., DeCarli, C., Carmichael, O.: Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR. In: *Information processing in medical imaging*. pp. 239–251. Springer (2009)
- [135] Searcey, T., Bierer, L., Davis, K.L.: A longitudinal study of alzheimers disease: measurement, rate, and predictors of cognitive deterioration. *Am J Psychiatry* 1, 51 (1994)

- [136] Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., et al.: Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology* 65(4), 403–413 (2009)
- [137] Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Figurski, M., Coart, E., Blennow, K., Soares, H., Simon, A.J., Lewczuk, P., Dean, R.A., et al.: Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. *Acta neuropathologica* 121(5), 597–609 (2011)
- [138] Sheline, Y.I., Price, J.L., Yan, Z., Mintun, M.A.: Resting-state functional MRI in depression unmasks increased connectivity between networks via the dorsal nexus. *Proceedings of the National Academy of Sciences* 107(24), 11020–11025 (2010)
- [139] Shiino, A., Watanabe, T., Maeda, K., Kotani, E., Akiguchi, I., Matsuda, M.: Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. *Neuroimage* 33(1), 17–26 (2006)
- [140] Shock, N.W., et al.: Normal human aging: The Baltimore longitudinal study of aging. (1984)
- [141] Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging* 17(1), 87–97 (1998)
- [142] Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E., Toga, A.W.: Longitudinal mapping of cortical thickness and brain growth in normal children.

Journal of Neuroscience 24(38), 8223–8231 (2004)

- [143] Sowell, E.R., Trauner, D.A., Gamst, A., Jernigan, T.L.: Development of cortical and subcortical brain structures in childhood and adolescence: a structural MRI study. *Developmental Medicine & Child Neurology* 44(1), 4–16 (2002)
- [144] Steinbrink, C., Vogt, K., Kastrup, A., Müller, H.P., Juengling, F., Kassubek, J., Riecker, A.: The contribution of white and gray matter differences to developmental dyslexia: insights from DTI and VBM at 3.0 T. *Neuropsychologia* 46(13), 3170–3178 (2008)
- [145] Tager-Flusberg, H., Joseph, R.M.: Identifying neurocognitive phenotypes in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences* 358(1430), 303–314 (2003)
- [146] Thal, D.R., Rüb, U., Orantes, M., Braak, H.: Phases of A β -deposition in the human brain and its relevance for the development of AD. *Neurology* 58(12), 1791–1800 (2002)
- [147] Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J.L., Resnick, S.M.: Longitudinal changes in cortical thickness associated with normal aging. *Neuroimage* 52(4), 1215–1223 (2010)
- [148] Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.B.: Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage* 35(1), 105–120 (2007)
- [149] Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. *The journal of machine learning research* 1, 211–244 (2001)

- [150] Toledo, J.B., Arnold, S.E., Raible, K., Brettschneider, J., Xie, S.X., Grossman, M., Monsell, S.E., Kukull, W.A., Trojanowski, J.Q.: Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimers Coordinating Centre. *Brain* 136(9), 2697–2706 (2013)
- [151] Toledo, J.B., Bjerke, M., Chen, K., Rozycki, M., Jack, C.R., Weiner, M.W., Arnold, S.E., Reiman, E.M., Davatzikos, C., Shaw, L.M., et al.: Memory, executive, and multidomain subtle cognitive impairment Clinical and biomarker findings. *Neurology* 85(2), 144–153 (2015)
- [152] Toledo, J.B., Brettschneider, J., Grossman, M., Arnold, S.E., Hu, W.T., Xie, S.X., Lee, V.M.Y., Shaw, L.M., Trojanowski, J.Q.: CSF biomarkers cutoffs: the importance of coincident neuropathological diseases. *Acta neuropathologica* 124(1), 23–35 (2012)
- [153] Toledo, J.B., Cairns, N.J., Da, X., Chen, K., Carter, D., Fleisher, A., Householder, E., Ayutyanont, N., Roontiva, A., Bauer, R.J., et al.: Clinical and multimodal biomarker correlates of ADNI neuropathological findings. *Acta neuropathologica communications* 1(1), 1 (2013)
- [154] Toledo, J.B., Gopal, P., Raible, K., Irwin, D.J., Brettschneider, J., Sedor, S., Waits, K., Boluda, S., Grossman, M., Van Deerlin, V.M., et al.: Pathological α -synuclein distribution in subjects with coincident Alzheimers and Lewy body pathology. *Acta neuropathologica* pp. 1–17 (2015)
- [155] Toledo, J.B., Weiner, M.W., Wolk, D.A., Da, X., Chen, K., Arnold, S.E., Jagust, W., Jack, C., Reiman, E.M., Davatzikos, C., et al.: Neuronal injury biomarkers and prognosis

- in ADNI subjects with normal cognition. *Acta neuropathologica communications* 2(1), 1 (2014)
- [156] Vemuri, P., Gunter, J.L., Senjem, M.L., Whitwell, J.L., Kantarci, K., Knopman, D.S., Boeve, B.F., Petersen, R.C., Jack, C.R.: Alzheimer’s disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage* 39(3), 1186–1197 (2008)
- [157] Verbeke, G., Lesaffre, E.: A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* 91(433), 217–221 (1996)
- [158] Wåhlstedt, C., Thorell, L.B., Bohlin, G.: Heterogeneity in ADHD: neuropsychological pathways, comorbidity and symptom domains. *Journal of abnormal child psychology* 37(4), 551–564 (2009)
- [159] Wang, L., Shi, F., Yap, P.T., Lin, W., Gilmore, J.H., Shen, D.: Longitudinally guided level sets for consistent tissue segmentation of neonates. *Human brain mapping* 34(4), 956–972 (2013)
- [160] Ward Jr, J.H.: Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58(301), 236–244 (1963)
- [161] Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Donohue, M.C., Green, R.C., Harvey, D., Jack, C.R., et al.: Impact of the Alzheimer’s disease neuroimaging initiative, 2004 to 2014. *Alzheimer’s & Dementia* 11(7), 865–884 (2015)

- [162] Whitwell, J.L., Avula, R., Master, A., Vemuri, P., Senjem, M.L., Jones, D.T., Jack, C.R., Josephs, K.A.: Disrupted thalamocortical connectivity in PSP: a resting-state fMRI, DTI, and VBM study. *Parkinsonism & related disorders* 17(8), 599–605 (2011)
- [163] Whitwell, J.L., Dickson, D.W., Murray, M.E., Weigand, S.D., Tosakulwong, N., Senjem, M.L., Knopman, D.S., Boeve, B.F., Parisi, J.E., Petersen, R.C., et al.: Neuroimaging correlates of pathologically defined subtypes of Alzheimer’s disease: a case-control study. *The Lancet Neurology* 11(10), 868–877 (2012)
- [164] Whitwell, J.L., Weigand, S., Gunter, J., Boeve, B.F., Rademakers, R., Baker, M., Knopman, D.S., Wszolek, Z.K., Petersen, R.C., Jack, C., et al.: Trajectories of brain and hippocampal atrophy in FTD with mutations in MAPT or GRN. *Neurology* 77(4), 393–398 (2011)
- [165] Wilson, R.S., Yu, L., Trojanowski, J.Q., Chen, E.Y., Boyle, P.A., Bennett, D.A., Schneider, J.A.: TDP-43 pathology, cognitive decline, and dementia in old age. *JAMA neurology* 70(11), 1418–1424 (2013)
- [166] Wolz, R., Heckemann, R.A., Aljabar, P., Hajnal, J.V., Hammers, A., Lötjönen, J., Rueckert, D., Initiative, A.D.N., et al.: Measurement of hippocampal atrophy using 4D graph-cut segmentation: application to ADNI. *NeuroImage* 52(1), 109–118 (2010)
- [167] Yamasue, H., Kasai, K., Iwanami, A., Ohtani, T., Yamada, H., Abe, O., Kuroki, N., Fukuda, R., Tochigi, M., Furukawa, S., et al.: Voxel-based analysis of MRI reveals anterior cingulate gray-matter volume reduction in posttraumatic stress disorder due

to terrorism. *Proceedings of the National Academy of Sciences* 100(15), 9039–9043 (2003)

[168] Zanetti, M., Ballabio, C., Abbate, C., Cutaia, C., Vergani, C., Bergamaschini, L.: Mild cognitive impairment subtypes and vascular dementia in community-dwelling elderly people: A 3-year follow-up study. *Journal of the American Geriatrics Society* 54(4), 580–586 (2006)

[169] Zhang, T., Davatzikos, C.: ODVBA: optimally-discriminative voxel-based analysis. *IEEE transactions on medical imaging* 30(8), 1441–1454 (2011)