BRAIN SEGMENTATION VIA DIFFEOMORPHIC LIKELIHOOD FUSION AND ITS APPLICATIONS TO CLINICAL ANALYSES

by

Xiaoying Tang

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

May, 2014

© Xiaoying Tang 2014
All rights reserved
Abstract

The human brain is composed of a variety of structures, or regions of interest (ROIs), that are responsible for a range of functions. It is therefore an essential step in quantitative analyses to define these ROIs in the human brain anatomy. Due to its intricate formation and function, this anatomy is highly complex and that makes it challenging to segment the human brain. During the past decade, the development of neuroimaging techniques (Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET)) has facilitated mathematical modeling algorithms for the reconstruction and representation of various human brain structures.

The primary goal of this thesis is to develop automated algorithms to accurately segment the human brain into anatomical regions based on two types of MR images – the T1-weighted image and the diffusion tensor image (DTI). We propose an automated segmentation algorithm in the framework of Bayesian parameter estimation. This algorithm is posed in the augmented orbit of a union of multiple atlases. The estimator of the segmentation label is obtained by maximizing the posterior probability
of the segmentation label $W$ given the observable image $I^D$ of the to-be-segmented subject. We assume that the posterior probability of interpreting a single observable image is shared across multiple different atlases, the anatomical information of which is optimally fused via a diffeomorphic likelihood fusion.

The ultimate goal of designing algorithms for neuroimaging studies is to be applicable to clinical studies. It is, therefore, natural and essential to apply the proposed segmentation algorithm to clinical studies and perform statistical analyses to study the functional mechanism of certain structures in the development of specific disorders. The most widely used metric for analysis is the volume measurement of an anatomical structure. In part, this thesis aims to develop statistical methods to examine additional, manifold-based, metrics in the setting of computational anatomy (CA) that can differentiate and predict disease states. We use the 2-D surface that contours the 3-D subvolume of the structure of interest to construct our manifold-based comparison metrics.

This thesis starts with a derivation of a Bayesian parameter estimation algorithm in the framework of multi-atlas random diffeomorphic orbit model. The maximum a posteriori (MAP) estimator is obtained by maximizing the fused likelihoods from multiple atlases, in an expectation maximization (EM) fashion, which we term multi-atlas likelihood fusion (MALF). We design a two-level hierarchical segmentation pipeline for T1-weighted brain images based on MALF. Validation results on a variety of datasets are presented. Following that, we apply the segmentation pipeline to two
large-scale neuroimaging studies and demonstrate its power in the clinical study of various diseases. We also describe an extension of this algorithm to the segmentation of multi-contrast DTI images. Subsequent to that, we illustrate a statistical shape analysis method to contrast two groups in terms of localized shape area, as well as an extension to longitudinal neuroimaging studies, examining the temporal dynamics of these differences – the local shape change rates. In the final portion of this thesis, we present a deformation-pattern based manifold learning and clustering approach applied to differentiating and predicting neurodegenerative diseases. Specifically, classification and prediction results on dementia of the Alzheimer type, based on the deformation pattern, will be presented.

Dr. Michael I. Miller
Dr. Laurent Younes
Dr. John Goutsias
Dr. Mounya Elhilali
Dr. Trac Tran
Acknowledgments

Four years ago, I was fortunate enough to start working with Dr. Michael I. Miller in the Center for Imaging Science (CIS) at Johns Hopkins University as a Ph.D. student. The past four years have been the most challenging but also the most rewarding time in my life so far. I appreciate all the beautiful memories I have of CIS. I am thankful to have studied and worked in such a warm environment. The support, encouragement and caring that the people of CIS gave to me made my Ph.D. life enjoyable regardless of the occasional frustration in graduate school.

I am most grateful to my Ph.D. advisor, Dr. Michael I. Miller, for his support, guidance, and encouragement, not only in research but also in life. Dr. Miller has been teaching me great lessons in many aspects of life. He is a friend, a mentor, and also a father. I would like to express my sincere thanks to Dr. Susumu Mori for teaching me how to research with steadiness and encouraging me at difficult times. The lessons that Dr. Mori taught me about digging into research and how to research in the manner of fishing will be immeasurably valuable across my entire life.

I would also like to thank Dr. Laurent Younes for bringing me into the field of
shape and diffeomorphism as well as statistical analysis on manifolds, one of the most important topics of my Ph.D. research. During the past four years, I have taken five different classes with Dr. Younes, which laid the mathematical foundations of my research. I appreciate Dr. Younes’s insights and suggestions in my research and his great sense of humor. I would like to thank J. Tilak Ratnanather for his support and valuable suggestions in functional MRI analyses. My thanks go to Dr. Andreia V. Faria, who helped me tremendously in understanding the human brain anatomy when I first entered the field of medical image analysis. I also owe my gratitude to Dr. Stewart Mostofsky for his guidance in studies on Autism and ADHD. I am thankful to Dr. John Goutsias for teaching me the basics on image processing and analysis. My thanks also go to Dr. Trac Tran and Dr. Mounya Elhilali for serving as my committee members.

I would like to thank the people who co-worked with me, Dr. Kenichi Oishi, Dr. Can Ceritoglu, Dr. Shoko Yoshida, Deana Crocetti, John Hsu for their support and collaboration on different research topics. I would like to thank my labmates, Daniel Tward, Kwame Kutten, Jianqiao Feng, Valentina Staneva, Saurabh Jain, Siamak Ardekani for their discussions on geometry, diffeomorphism, manifold learning and clustering. I would like to specifically thank the staffs at CIS, Mike Bowers, Anthony Kolasny, Dawn Kilheffer, Maura Vonasek, Timothy Brown, Erika Lance, and Elizabeth Postell for making CIS a sweet home.

I would like to extend my thanks to all my friends for being supportive in my
personal life. Special thanks to my parents, Ping Tang and Taoe Zou, for their infinite and enduring support and love.

The work reported here was supported by grants: National Institutes of Health (NIH) P41 EB015909, RO1 AG020012, R01 EB000975, R01 MH084803, and S10 RR025053.
Contents

Abstract ii

Acknowledgments v

List of Tables xiv

List of Figures xvii

1 Introduction 1

1.1 Previous Work ........................................ 4

1.1.1 Automated Segmentation of Subvolumes .......... 4

1.1.2 Metric Comparisons of Manifolds ................. 6

1.1.3 Statistical Inference on Manifolds ................. 7

1.2 Thesis Contribution ................................... 7

1.2.1 Bayesian Parameter Estimation and Multi-atlas Likelihood Fu-

sion .......................................................... 10

1.2.2 Multi-contrast MALF based Segmentation of DTI .. 12
1.2.3 Surface Based Morphometric Analysis ........................................ 13
1.2.4 Discrimination and Prediction of Disease States Based on Shape
Deformation Markers ................................................................. 14

2 Bayesian Parameter Estimation and Segmentation in the Multi-atlas

Random Orbit Model .................................................................. 16

2.1 Introduction ......................................................................... 16

2.2 Method Derivation and Description ....................................... 21

2.2.1 The Random Deformable Atlas Orbit Model ...................... 21

2.2.2 Multi-atlas Likelihood Fusion ............................................. 24

2.2.3 Probability Model .............................................................. 26

2.2.4 MALF via Expectation Maximization ................................. 29

2.2.5 Implementation of MALF for Brain Segmentation Label Esti-
mation ....................................................................................... 35

2.3 Application .......................................................................... 41

2.3.1 Two-level Hierarchical Brain Segmentation ...................... 42

2.4 Validation Results ............................................................... 46

2.4.1 Validation of MALF in Skull-stripping ............................... 47

2.4.1.1 Pediatric healthy control/Autism/ADHD Dataset ........... 47

2.4.1.2 BIOCARD Dataset ....................................................... 50

2.4.2 Validation of MALF in Segmenting Subcortical and Ventricular
Structures ................................................................................. 54
2.4.2.1 Pediatric healthy control/Autism/ADHD Dataset (3T MPRAGE) .............................. 54
2.4.2.2 Elderly healthy control/dementia/PPA Dataset (3T MPRAGE) ........................... 59
2.4.2.3 Young healthy adult Dataset (3T MPRAGE) .................................................. 65
2.4.2.4 Elderly normal/preclinical AD Dataset (1.5T SPGR) ...................................... 71

2.5 Conclusion .............................................................................................................. 74

3 Volumetric Analysis of MALF in Case-control Clinical Studies ............................. 77

3.1 Introduction ............................................................................................................. 77
3.2 Motivation ............................................................................................................... 78
  3.2.1 ADHD ............................................................................................................... 78
    3.2.1.1 Basal-ganglia circuitry .............................................................................. 78
    3.2.1.2 Amygdala-hippocampus circuitry ......................................................... 79
  3.2.2 Autism ............................................................................................................... 80
    3.2.2.1 Basal-ganglia circuitry .............................................................................. 80
    3.2.2.2 Amygdala-hippocampus circuitry ......................................................... 80
3.3 Method .................................................................................................................... 82
3.4 Results ...................................................................................................................... 83
  3.4.1 ADHD ............................................................................................................... 83
    3.4.1.1 Basal-ganglia circuitry .............................................................................. 83
    3.4.1.2 Amygdala-hippocampus circuitry ......................................................... 87
3.4.2  Autism ......................................................... 87
   3.4.2.1  Basal-ganglia circuitry .............................. 87
   3.4.2.2  Amygdala-hippocampus circuitry .................. 90
3.5  Conclusion ..................................................... 90

4  Whole Brain Segmentation of Diffusion Tensor Imaging via MALF  92
   4.1  Introduction .................................................. 92
   4.2  Multi-contrast MALF ......................................... 94
   4.3  Validation Analysis ......................................... 98
      4.3.1  Patient populations .................................. 98
      4.3.2  MRI scans .............................................. 100
      4.3.3  DTI processing ........................................ 102
      4.3.4  Manual delineation for accuracy measurement .... 102
   4.4  Results ....................................................... 104
   4.5  Conclusion ................................................... 112

5  Surface-based Morphometric Analysis ........................ 114
   5.1  Introduction ................................................ 114
   5.2  Method ....................................................... 116
      5.2.1  Template Surface Generation ....................... 116
      5.2.2  LDDMM-surface Registration ....................... 118
      5.2.3  Vertex-based Statistical Analysis ................. 120
5.3 Simulation Study ................................................. 122
  5.3.1 Mathematical Modeling .................................. 122
  5.3.2 Experimental Results .................................... 123
5.4 Clinical Study ................................................... 125
  5.4.1 Alzheimer’s Disease Neuroimaging Initiative ......... 125
  5.4.2 Participants ............................................... 126
  5.4.3 Image protocol and volumetric segmentation ........... 127
  5.4.4 Surface generation ....................................... 128
  5.4.5 Results .................................................. 130
5.5 Conclusion ....................................................... 140

6 Statistical Shape Analysis of Longitudinal Neuroimage Data 143
  6.1 Introduction ................................................. 143
  6.2 Method .................................................... 145
    6.2.1 Longitudinal SBM Analysis ............................ 146
    6.2.2 Statistical Model in Examining Shape Change Rates . 151
  6.3 Results ................................................... 152
    6.3.1 Three-structure Analysis ................................ 152
    6.3.2 Subregion Analysis ...................................... 154
  6.4 Conclusion .................................................. 155

7 Disease Detection and Prediction using Shape Deformation Pat-
7.1 Introduction ........................................... 159

7.2 Discrimination between HC and AD via Shape Deformation Patterns 161
    7.2.1 Principal Component Analysis ..................... 161
    7.2.2 Linear Discriminant Analysis ...................... 164

7.3 Predicting MCI-to-AD Conversion ........................ 171
    7.3.1 Background ......................................... 171
    7.3.2 Participant ......................................... 173
    7.3.3 Generation of Shape Deformation Features .......... 174
    7.3.4 Selecting the Optimal Predictor .................... 174
    7.3.5 MCI-to-AD Prediction Results ...................... 176

7.4 Conclusion ............................................ 178

Bibliography ............................................. 182
List of Tables

2.1 Mean and standard deviations of the dice score coefficients (DSCs) of the skull-stripping results obtained from the three automated approaches – MALF, HWA, and BET, when compared with the corresponding manually stripped slices. .................................................. 49

2.2 Pearson’s linear correlation coefficients between the manual and three automated structure volume measurements. The automated structures are obtained respectively from MALF, Freesurfer, and FSL. Abbreviations: lcaud – left caudate; lputa – left putamen; lpall – left globus pallidus; lthal – left thalamus; lamyg – left amygdala; lhipp – left hippocampus; rcaud – right caudate; rputa – right putamen; rpall – right globus pallidus; rthal – right thalamus; ramyg – right amygdala; rhipp – right hippocampus .................................................. 56

2.3 The average DSC values between the manual and automated segmentations for each structure of the fourteen for Group 1 subjects, thus comparing STAPLE, Spatial STAPLE, and MALF. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significant in being greater than that of other methods ($p < 0.05$). .................................................. 66

2.4 Mean and standard deviations of DSCs computed across the fifteen subjects in Group 2 for each structure for comparisons of STAPLE, Spatial STAPLE, and MALF. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significantly higher than that of other methods ($p < 0.05$). .................. 67

2.5 Mean and standard deviations of the DSC values obtained respectively from STAPLE, Spatial STAPLE, and MALF segmentations of the 16 structures of the six Group 3 subjects. Bold typesetting indicates that the dice overlap obtained from the corresponding is statistically significantly higher than that of other methods ($p < 0.05$). .................. 68
2.6 Mean and standard deviations of dice overlaps obtained from MALF, Spatial STAPLE, and STAPLE in segmenting the 10 structures of the 16 scans from BIOCARD. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significant higher than that of other methods ($p < 0.05$). 74

3.1 Mean and standard deviations of volume measurements ($mm^3$) of the left and right amygdala and hippocampus, in six groups – the entire ADHD group, the entire healthy control group, the ADHD male-only group, the healthy control male-only group, the ADHD female-only group, as well as the healthy control female-only group. 88

4.1 Anatomical changes in the sixteen subjects used as the multiple DTI atlases. 99

4.2 Anatomical changes in the 25 subjects used for testing the multi-contrast MALF segmentation accuracy. 101

4.3 Pearson’s correlations between the manual and four different automated ROI area measures, with bold typesetting indicating that the correlation between the automated and the manual measures is statistically significant compared to other methods ($p < 0.025$). 106

4.4 Pearson’s correlations between the mean FA value within the manual ROI and the automated segmentations, with bold typesetting suggesting that the correlation between the automated measure and the manual one is statistically significant compared to the other methods ($p < 0.025$). 110

4.5 Pearson’s correlations between the mean diffusivity (MD) value within the manual ROI and the automated segmentations with bold typesetting indicating that the correlation between the automated measure and the manual one is statistically significant compared to the other methods ($p < 0.025$). 112

5.1 Demographic information of the baseline dataset in the ADNI study. 127

5.2 Average volume measurements ($mm^3$) of the original segmentations from FreeSurfer and the ones “filtered” from the LDDMM-based pipeline as well as their mean differences. Abbreviations: lvent - left lateral ventricle, ltha - left thalamus, lcaud - left caudate, lputa - left putamen, lpall - left globus pallidus, lhpp - left hippocampus, lamyg - left amygdala, rvent - right lateral ventricle, rtha - right thalamus, rcaud - right caudate, rputa - right putamen, rpall - right globus pallidus, rhipp - right hippocampus, ramyg - right amygdala. 131
7.1 A comparison of the upper bounds of the confidence intervals for p-values obtained from the volume analysis and the PCA based shape analysis for each structure. Abbreviations: lvent - left lateral ventricle, ltha - left thalamus, lcaud - left caudate, lputa - left putamen, lpall - left globus pallidus, lhipp - left hippocampus, lamyg - left amygdala, rvent - right lateral ventricle, rtha - right thalamus, rcaud - right caudate, rputa - right putamen, rpall - right globus pallidus, rhipp - right hippocampus, ramyg - right amygdala. ........................................ 165

7.2 A comparison of the specificity and the sensitivity obtained from the LDA classifiers built respectively from the volume information and the shape information for each structure. .............................. 170
List of Figures

1.1 Demonstration of the three primary contributions in this thesis: panel (1) represents the construction of subvolumes of subcortical structures from T1-weighted images; panel (2) illustrates metric comparisons of the 2-D shapes bounding the 3-D subvolumes obtained from (1); panel (3) shows a classification of different disease states via shape manifold learning and clustering. ............................... 3

2.1 Figure shows the population of atlases and the to-be-segmented target at the center. Each atlas consists of the deformable template image and the segmentation labels, represented via the pair $(I_a, W_a)$. .................. 27

2.2 Depiction of two local charts and the associated diffeomorphisms chosen to illustrate the weighted interpretation. The local charts are related to each other via diffeomorphic coordinate transformations in a way illustrated in the figure. Two points $X, Y$ that respectively belong to the hippocampus chart and the amygdala chart are compared using the forward and inverse mappings. ................................. 34

2.3 Schematic of the proposed two-level hierarchical segmentation pipeline. Abbreviation: MALF – multi-atlas likelihood fusion. ...................... 46

2.4 Standard locations of the six pre-selected sagittal slices (demonstrated on a coronal image) that have been manually skull stripped for validation analysis. From left to right: slice1, slice2, slice3, slice4, slice5, and slice6. ......................................................... 48

2.5 Panel A) and Panel B) respectively show the automated skull-stripping results of two subjects. The boundaries of the skull-stripping results are superimposed on the corresponding T1-weighted images. From left to right: results obtained from MALF, HWA, and BET. .................. 51
2.6 The top, middle, and bottom panels respectively show the automated skull-stripping results of three scans from the BIOCARD study. The outlines of the masks are superimposed on the corresponding T1-weighted images. From left to right, results are obtained from MALF, HWA, and BET.

2.7 The mean dice score coefficients (DSCs) and the standard deviations of the 12 structures of interest (left and right caudate, globus pallidus, putamen, thalamus, amygdala, as well as hippocampus) obtained from MALF (blue), Freesurfer (red), and FSL (yellow) relative to the manual delineations.

2.8 The top and bottom panel respectively shows the segmentation results of the six basal ganglia structures (left and right caudate, putamen, and globus pallidus) and the bilateral thalamus in two subjects. From left to right, segmentations are obtained from manual delineation, MALF, Freesurfer, and FSL.

2.9 Panels A), B) respectively demonstrate the segmentation results of the amygdala and the hippocampus in two representative subjects. From left to right, segmentations are obtained from manual delineation, MALF, Freesurfer, and FSL.

2.10 A comparison of segmentation accuracy between single-atlas and multi-atlas segmentations. Panels a), b), c) show the mean and the standard deviations of the DSCs for the sixteen structures obtained from single-atlas LDDMM (red) and MALF (green) for the three different groups.

2.11 Depiction of the variability within single atlases. This figure shows the scatter plots of DSC values for 16 structures in one subject, between the automated segmentations from 6 different atlases via single-atlas LDDMM and the corresponding manual segmentations.

2.12 Examples of subcortical segmentations from single- and multi-atlas approaches. Panel A shows the automated segmentation results of two subjects using the single-atlas, while panel B shows the segmentation results for the same subjects using the multi-atlas approach.

2.13 A comparison of MALF, STAPLE, and Spatial STAPLE, demonstrated via three representative 2-D slices of three structures near medial temporal regions – the amygdala, the hippocampus, and the ventricle in both hemispheres obtained respectively from manual delineation (top row), MALF (2nd row), STAPLE (3rd row), and Spatial STAPLE (bottom row).
2.14 Bar plots showing the mean DSC values, with the standard deviations plotted as the error bars, for 12 segmented subcortical structures and lateral ventricles of 8 young healthy adults, from MALF (blue), FreeSurfer (red), and FSL (green). Keys: left-amyg – left amygdala; right-amyg – right amygdala; left-caud – left caudate; right-caud – right caudate; left-pall – left globus pallidus; right-pall – right globus pallidus; left-hipp – left hippocampus; right-hipp – right hippocampus; left-puta – left putamen; right-puta – right putamen; left-thal – left thalamus; right-thal – right thalamus; left-vent – left lateral ventricle; right-vent – right lateral ventricle.

2.15 Bar plots showing the mean dice overlaps, with the standard deviations plotted as the error bars, for a comparison of MALF (blue), Freesurfer (red), and FSL (yellow) in segmenting the 10 structures of interest in the 16 BIOCARD scans.

2.16 Demonstration of the bilateral putamen and globus pallidus segmentations for representative subjects. From left to right: manual, MALF, Freesurfer, and FSL.

2.17 Visual comparisons of the segmentations of the hippocampus and the amygdala in two representative subjects. The segmentations come from manual delineation, MALF, Freesurfer, and FSL.

3.1 Representation of the volume measurements of the left and right caudate in subjects from the typically developing group (blue diamond) and the ADHD group (red circle). Bar means the averaged value in each group.

3.2 Figure showing the volume measurements of the putamen, in both hemispheres, from the TD group (blue diamond) and the ADHD group (red circle). Results from the entire group, the male only group, and the female only group are tabulated, with bars representing the mean values in each group.

3.3 Demonstration of globus pallidus volumes in both sides for TD (blue diamond) and ADHD (red circle) populations. Measurements of the entire group, the male group, as well as the female group, are listed with mean values plotted as bars.

3.4 Volume measurements of basal ganglia structures in subjects from the age-matched TD group (blue diamond) and the ASD group (red circle). Bar means the averaged value in each group.
4.1 The mean Dice overlaps and the standard deviations of the eighteen ROIs obtained from automated segmentations based on five contrasts (red), the single FA contrast (green), the single MD contrast (blue), the combined EV contrast (yellow), as well as the inter-rate (patterned). The mean values are calculated across fourteen different subjects. Star marks indicate significant difference among the four sets of Dice results by ANOVA ($p \ll 0.05$). Abbreviations are: GCC – genu of corpus callosum; BCC – body of corpus callosum; Caud – caudate; Put – putamen; ALIC – anterior limb of internal capsule; PLIC – posterior limb of internal capsule; CG – cingulate gyrus; MCP – middle cerebellar peduncle; SLF – superior longitudinal fasciculus; CST – corticospinal tract.

4.2 A plot of the correlation between the automated and the manual measurements of the size of the caudate in both hemispheres in square millimeters. Results from the four automated segmentation methods are compared: 5-contrast (red), FA-only (green), MD-only (blue), and EV-only (yellow).

4.3 A correlation plot between the automated and manual measurements of the sizes of left and right corticospinal tracts (CST). Results from the four automated segmentation methods are compared: 5-contrast (red), FA-only (green), MD-only (blue), and EV-only (yellow).

4.4 Demonstration of the segmentation accuracy of the CST in three representative cases with different degrees of anatomical abnormalities. Results from four different approaches are compared. From left to right: the color-coded map without CST definition superimposed, the color-coded map with CST from the 5-contrast MALF superimposed, the color-coded map with CST from the FA-only single-contrast MALF superimposed, the color-coded map with CST from the MD-only single-contrast MALF superimposed, and the color-coded map with CST from the EV-only 3-contrast MALF superimposed.

4.5 Results of the whole brain segmentations into 159 structures in three representative cases with large anatomical variability. The segmentation results are superimposed on color (upper row) and MD (bottom row) images.

5.1 Shape differences between two sets of artificial shapes that vary mainly in length rather than thickness. Warm color suggests compression while cool color suggests expansion in the group with shorter length when compared with the other group. The scale value measures the ratio of the local volume of the group with larger length to that of the group with shorter length in logarithmic scale.
5.2 This figure shows the mean and the standard deviations of the Kappa overlaps between the segmentations from the denoising pipeline and the original FreeSurfer segmented volumes. Black and white bars respectively denote left and right structures. Key: vent – Ventricles, thal – Thalamus, caud – Caudate, puta – Putamen, pall – Pallidum, hipp – Hippocampus, amyg – Amygdala.

5.3 Panels a)-b) respectively show the group shape differences between HC and AD, HC and MCI, as well as MCI and AD measured in the left and right amygdala. Both superior (left) and inferior (right) views are displayed for each group comparison. Negative color scale values indicate surface expansion in the latter group, and positive values indicate atrophy. The scale value quantifies the ratio of the local volume of the former group at a particular location to that of the latter group in logarithmic scale.

5.4 Panels a), b) demonstrate the group shape differences between HC and AD, HC and MCI, MCI and AD of the hippocampus in both hemispheres. Both superior (left) and inferior (right) views are displayed for each group comparison. Negative color scale values indicate surface expansion in the latter group, and positive values indicate atrophy. The scale value measures the ratio of the local volume of the former group to that of the latter group in logarithmic scale.

5.5 Group shape differences detected in the lateral ventricles in both hemispheres. Colors corresponding to negative values indicate local expansion of surfaces in the latter group whereas positive values imply atrophy, compared with the former group. The more negative the value, the more prominent the expansion. The scale value measures the ratio of the local volume of the former group to that of the latter group in logarithmic scale.

5.6 Surface deformation differences between HC and AD, HC and MCI, as well as MCI and AD in basal ganglia regions and the thalamus. Colors corresponding to negative values indicate local expansion of surfaces in the latter group, compared with the former group, whereas positive indicates atrophy in the latter group. The scale value quantifies the ratio of the local volume of the former group to that of the latter group in logarithmic scale.
5.7 Group differences, between every two of the three groups, measured in the surfaces of all the seven structures in the left hemisphere. Warm color denotes regions where the corresponding structure has significant atrophy in the latter group when compared with the former group. Cool color suggests local expansion of the structure in the latter group when compared with the former group. Key: am – amygdala, hi – hippocampus, vent – lateral ventricle, thal – thalamus, puta – putamen, pal – globus pallidus, caud – caudate. The scale value measures the ratio of the local volume around a particular vertex of the former group to that of the latter group in logarithmic scale.

5.8 Group shape variations of all the seven structures in the right hemisphere. Colors corresponding to positive values indicate regions on the structure where there is significant atrophy in the latter group when compared with the former group. Colors corresponding to the negative values suggest local expansion on the structure in the latter group, compared with the former one. Key: am – amygdala, hi – hippocampus, vent – lateral ventricle, thal – thalamus, puta – putamen, pal – globus pallidus, caud – caudate. The scale value measures the ratio of the local volume of the former group to that of the latter group in the logarithmic scale.

5.9 Panels a)-b) display group differences, detected in the left and right amygdala, between the two subtypes: MCI-stable and MCI-AD. Positive values indicate atrophy in MCI-AD while negative indicates expansion as compared with MCI-stable. The scale value measures the ratio of the local volume around each vertex on the template surface of the MCI-stable group to that of the MCI-AD group in logarithmic scale.

5.10 Hippocampal shape differences between MCI-stable and MCI-AD. Warm color indicates regions on the hippocampus where significant atrophy has been detected in MCI-AD when compared with MCI-stable, whereas cool color indicates expansion. The scale value measures the ratio of the local volume of the hippocampus in MCI-stable to that in MCI-AD in the logarithmic scale.

5.11 Shape differences between MCI-stable and MCI-AD detected on lateral ventricles in both hemispheres. The color corresponding to value 0 suggests no regional shape variation. Colors corresponding to negative values suggest expanding regions in MCI-AD when compared with MCI-stable, while those corresponding to positive values indicate atrophy in MCI-AD. The scale value measures the ratio of the local volume of the lateral ventricle in MCI-stable to that in MCI-AD in logarithmic scale.
6.1 Demonstration of the longitudinal shape analysis pipeline. All statistical analyses are performed in a common registered coordinate system.

6.2 Illustrations of the regional shape change rate differences, located on the three structures (amygdala, hippocampus, and ventricle), in both hemispheres, between every two of the three groups (HC, MCI, and AD). For the amygdala and the hippocampus, colors with positive values represent larger atrophy rates whereas colors with negative values represent smaller atrophy rates in the latter group relative to the former. For the ventricle, colors with negative values indicate larger expansion rates while those with positive values indicate smaller expansion rates in the latter group when compared with the former group.

6.3 Spatial maps of the atrophy rate differences of the left hippocampus, between every group pair (out of HC, MCI, and AD), for the entire surface (the first column from left) and the restrictions to the CA1 zone (the second column from left), the CA2 zone (the middle column), the CA3/dentate gyrus zone (the fourth column from left), and the subiculum zone (the fifth column from left). Colors with positive values indicate greater atrophy in the latter group relative to the former group.

6.4 Spatial maps of the atrophy rate differences of the right hippocampus, between every group pair (out of HC, MCI, and AD), for the entire surface (the first column from left) and the restrictions to the CA1 zone (the second column from left), the CA2 zone (the middle column), the CA3/dentate gyrus zone (the fourth column from left), and the subiculum zone (the fifth column from left). Colors with positive values indicate greater atrophy in the latter group relative to the former group.

6.5 Significant differences between every group pair (out of HC, MCI, and AD), in terms of the atrophy rates of the entire left amygdala (the first column from left), the left basolateral amygdala (the second column from left), the left basomedial amygdala (the third column from left), the left centromedial amygdala (the fourth column from left), and the left lateral nucleus of the amygdala (the fifth column from left). Colors with positive values indicate greater atrophy in the latter group relative to the former group.

7.1 PCA of the initial momentum matrices of the bilateral amygdala and hippocampus. Each sub-figure shows permutation test results for a structure’s group comparison based on the first $M$ PCs that account for 95% of the total variability. Shown are: 1) value (solid blue line) of each group comparison (total of three comparisons); 2) $p = 0.0001$ (red dotted line), $p = 0.001$ (blue dotted line), and $p = 0.05$ (black dotted line) for reference; 3) p-value derived from the 40000 permutation tests (solid green line).
7.2 This figure summarizes the procedure of selecting the optimal LDA classifier based on the PCs from all the fourteen structures, comprised of three steps. Step 1 is to create all possible combinations of PCs, resulting in a total of 127. Step 2 is to test the mean classification rate for each set of PCs, based on leave-one-out LDA. Step 3 is to select the LDA classifier with the highest mean correct classification rate in Step 2. Key: hi - Hippocampus, am - Amygdala, vl - Lateral Ventricle, th - Thalamus, pu - Putamen, pa - Pallidum, ca - Caudate. 

7.3 This figure summarizes the leave-one-out cross-validation procedure of testing the true classification rate that we would be able to yield using our procedure.

7.4 Flowchart demonstrating the procedure of selecting the optimal linear discriminant analysis (LDA) classifier in classifying MCI-C from MCI-NC.

7.5 Flowchart demonstrating the procedure of performing leave-one-out (LOO) cross-validation without being biased by the subject removed at the beginning. This process is repeated for each MCI subject.

7.6 Bar graph showing the total number of MCI subjects converted under different conversion times and the number of MCI-C subjects correctly classified under each corresponding conversion time.
Chapter 1

Introduction

Comprising of 100 billion neurons, the brain is the most complex part of the human body. It is the origin of our intelligence and all senses, the initiator of our bodily movement, as well as the controller of our behavior. The human brain, in some sense, defines all the qualities of humanity. For centuries, scientists from many disciplines have been trying to unveil the secrets of this part of our anatomy. During the past decade, there has been an acceleration in the pace of research on the brain due to the development of various imaging techniques such as magnetic resonance imaging (MRI), computed tomography (CT), magnetoencephalography (MEG), positron emission tomography (PET), and single-photon emission computed tomography (SPECT). These high-performance imaging techniques enable doctors and researchers to understand the brain with greater depth and comprehension.

With the development of imaging techniques and imaging scanners, comes an ever-
growing quantity of volumetric image data from subjects all over the world. Clinical decisions made from the medical images rely on doctor’s qualitative evaluation, which is still subjective. With this in mind, the benefits of quantitative, objective, and automated clinical judgements, in the field of medical image analysis, are clear.

Along with the development of neuroimaging techniques, the emerging discipline of computational anatomy (CA) [1] enables precise mathematical study of human brain anatomy. There are three main components in the field of CA [2]: (1) automated construction/representation of anatomical manifolds such as points, curves, surfaces, and subvolumes; (2) quantitative metric comparison of such anatomical manifolds in association with various disorders; (3) statistical judgements of disease states based on those anatomical manifolds. This thesis addresses the three aspects of CA. Towards (1), we describe an approach to automatically construct 3-dimensional (3D) subvolume manifolds - namely, the automated segmentation of regions of interest (ROIs) in the human brain. Towards (2), this thesis details a statistical method for studying anatomical shapes via metric comparisons. Finally, towards (3), we propose a manifold learning and clustering method to differentiate and predict functional stages of neurodegenerative diseases via a supervised learning of the anatomical shape manifolds.

Figure 1.1 illustrates the three main contributions of this thesis: (i) automatically segmenting the structures of interest (for example, the subcortical structures) from the T1-weighted images (as well as diffusion tensor images) of subjects with different
disease states (healthy control, mildly impaired, or severely impaired) in terms of a diverse range of disorders; (ii) studying the group differences via metric comparisons of manifolds (for example, the 2-D surfaces contouring each of the subvolumes of subcortical structures); (iii) differentiating the diseased population (for example, the Alzheimer’s disease (AD) population) from a healthy control population as well predicting a disease status (for example, conversion of mild cognitive impairment (MCI) to AD) via a learning and clustering of the manifold features obtained in stage (ii).

Figure 1.1: Demonstration of the three primary contributions in this thesis: panel (1) represents the construction of subvolumes of subcortical structures from T1-weighted images; panel (2) illustrates metric comparisons of the 2-D shapes bounding the 3-D subvolumes obtained from (1); panel (3) shows a classification of different disease states via shape manifold learning and clustering.
1.1 Previous Work

1.1.1 Automated Segmentation of Subvolumes

To quantitatively analyze the function of brain structures or ROIs via medical image analysis techniques, one of the first essential steps is to extract the ROIs from the medical images. There are many types of automated and manual approaches that have been proposed to define ROIs in the brain, based on spatial and photometric features of the structures of interest. These methods often require a priori knowledge as a form of atlas. For manual ROI drawing, an atlas could be a simple pictorial representation of the structure of interest, which guides operators to define the same structure in target images. The manual delineation, while often used as a gold standard, is a time-consuming approach. Various types of automated segmentation tools have been proposed, aiming to automatically define the boundary of anatomical structures based on image contrast information [3–6]. Some of the advanced tools incorporate a priori knowledge about the spatial information of the target structures as a form of probabilistic atlas [7–9]. This spatial constraint prevents the contrast-based boundary definition from leaking into unlikely regions.

To create a probabilistic atlas for a specific ROI, which here we will call $S$, multiple atlas images with that specific ROI $S$ manually delineated in each atlas will be used. The standard way is to align all the atlas images to a common image space and then bring the ROI $S$ of each atlas to that common space as well, based on image
registration techniques. Then the probability for any voxel $x$, in the domain of the common image space, to be in $S$ is computed as the ratio of the number of the deformed atlas ROIs that have non-zero value at $x$. One benefit of the probability atlas approach is that it incorporates the anatomical information of images with different phenotypes and is therefore able to segment the human brain with a wide range of anatomical features. This process can be regarded as a pre-fusion of the anatomical features from multiple atlases because the probabilistic atlas is created before hand. In other words, it is not target-dependent. For different to-be-segmented subjects, the probabilistic atlas will be the same.

In recent years, segmentation via a post-fusion of the anatomical information of multiple atlases has become popular. Literally, the atlas information is fused in a manner that is target-dependent. The fusion results may vary from target to target. Each atlas is warped to the to-be-segmented subject image space, and the definition (label) of the ROI $S$ in each atlas is cast to the subject space as well, which can then be fused (aka “label fusion”) based on pre-defined algorithms such as those proposed by [10–14]. It has been shown that simple label fusion techniques via majority voting yield robust segmentations [10,15,16]. More recently, weighted majority voting strategies with an incorporation of the intensity information have demonstrated significant improvement in the segmentation accuracy. A variety of weighting approaches, based on intensity similarity metrics, have been proposed – global [11], local [17,18], semi-local [18], and non-local [19].
1.1.2 Metric Comparisons of Manifolds

Due to the unique and complex structural and functional variability in the human brain anatomy, scalar morphological comparison is usually not sufficient for detecting mild abnormalities in disease populations. Metric or vector comparison allows for a more powerful characterization of the complex biological variability in human neuroanatomy. The key idea of metric comparison is that we assume each individual anatomy is a deformed version of a canonical representation of the anatomy – the deformable template. In this deformable template model, vector mapping techniques are required to build spatial connections between two neuroanatomical spaces. Considering the highly convoluted nature of brain anatomy, this is a challenge. During the past two decades, there has been tremendous development in vector based brain mapping techniques [20–24]. More recently, diffeomorphic mapping has become prevalent by allowing researchers to examine biological shape variations in a mechanism akin to the idea of fluid mechanics [25–30].

Our efforts in metric comparisons of biological shapes and forms have largely relied on the method of large deformation diffeomorphic metric mapping (LDDMM) [31–36], in which bijective smooth transformations are created. This technique reduces the human brain anatomy space to a metrizable space with a metric induced by the geodesic lengths of the diffeomorphic flows that connect the anatomies. By doing this, the morphological comparison of anatomical systems is simplified to the metric comparison of the diffeomorphisms created in LDDMM, which we term diffeomorphometry [37]. Dif-
feomorphometry has been widely applied to studying 3-D subvolumes as well as 2-D surfaces in a variety of neurodegenerative diseases in the framework of CA [38–46].

1.1.3 Statistical Inference on Manifolds

The topic of statistical inference on manifolds has developed rapidly along with the advent of numerous automated algorithms for manifold construction: diffeomorphic atlas building in the framework of CA [47,48], growth modeling of the brain anatomy [49,50], as well as machine learning on diffeomorphisms [51]. The idea of performing disease-related statistical inference, based on anatomical manifolds, has been widely utilized in neuroimaging studies on a diverse range of neurodegenerative diseases [52], including but not exclusive to: dementia of the Alzheimer type [53–60], schizophrenia [61–69], attention-deficit hyperactivity disorder (ADHD) [70–73], Autism [74–77] and comparisons across multiple diseases [78].

1.2 Thesis Contribution

The primary goal of this thesis is to develop a fully-automated medical image analysis pipeline to facilitate the study of biological shapes (e.g. 3-D subvolumes, 2-D surfaces, 1-D curves) of brain structures in populations with different anatomical states. The image modalities we work with are T1-weighted images and diffusion tensor images (DTIs) of the human brain.
To analyze specific brain structures of interest, the first thing we do is to segment their 3-D subvolumes from T1-weighted images or DTIs. To address this requirement, automated segmentation methods, based on both T1-weighted images and DTIs, will be demonstrated. We shall start, in Chapter 2, with the theoretical formulations of the automated segmentation method, multi-atlas likelihood fusion (MALF). Following on from that, we describe a hierarchical segmentation pipeline for T1-weighted images, based on a two-level MALF. Two stages of segmentations are presented – skull-stripping and segmentation of subcortical and ventricular structures. Validation analysis of the pipeline, in terms of the segmentation accuracy of those two stages, is performed on T1-weighted images of subjects from a wide range of populations: 30 pediatric subjects with different disease states (healthy control, ADHD, and Autism); 8 healthy young adults; 35 elderly subjects with different disease states (normal aging, AD, and primary progressive aphasia (PPA)). The above 3 datasets were all acquired from 3 Tesla scanners. We also evaluate the accuracy of the pipeline in segmenting 16 elderly subjects (normal aging and preclinical AD) obtained from 1.5 Tesla scanners. With the validity established, we apply the segmentation pipeline, in Chapter 3, to two case-control clinical studies on ADHD and Autism, investigating the volumetric abnormalities in basal ganglia structures (the caudate, putamen, and globus pallidus), the amygdala, and the hippocampus in populations with those two disorders. Chapter 4 seeks to extend MALF to whole brain segmentations based on multi-contrast DTIs. Validation in segmenting a DTI of the human brain into a total
of 159 anatomical regions is presented based on two datasets – 25 pediatric subjects with different disease states (normal, mild abnormal, and severe abnormal) and 16 healthy young subjects.

The next step, after automatically constructing the anatomical manifolds of interest, is to morphologically analyze the physiological signals (e.g. cortical thickness, local surface area, subvolume size) present in the anatomical manifolds. For this, and our focus of Chapter 5, we illustrate a cross-sectional analysis approach to statistically compare the local surface features amongst different groups. This surface based morphometric (SBM) analysis pipeline is first validated on a simulation study and then applied to one large-scale publicly available clinical study – the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). With this dataset, we analyze the surface variations of fourteen subcortical and ventricular structures (the left and right caudate, putamen, globus pallidus, thalamus, hippocampus, amygdala, and ventricle) in MCI as well as AD, compared to age-matched normal aging. Following that, Chapter 6 extends the statistical model to accommodate longitudinal SBM analyses. Temporal dynamics of the fourteen structures, quantified by the vertex-based shape change rates, are examined in ADNI subjects. Spatial patterns of the temporal change dynamics of the amygdala, the hippocampus, and the ventricle in MCI and AD, relative to the normal aging population are investigated.

In the final chapter of this thesis, we introduce a pipeline for the discrimination of subjects with AD and the prediction of conversion from MCI to AD, through
supervised learning of the shape deformation patterns of the fourteen subcortical and ventricular structures in ADNI.

In the following subsections, we will give a more detailed outline of this thesis. Moreover to give an alternative viewpoint to its structure, we group the chapters akin to the three challenges of CA and contributions of this thesis we outlined earlier. Section 1.2.1, Bayesian Parameter Estimation and Multi-atlas Likelihood Fusion, relates to point (1) and will be expounded upon in Chapters 2 and 3. Section 1.2.2, Multi-contrast MALF based Segmentation of DTI, finishes our address of (1) with an outline of Chapter 4. Point (2) comes to the fore in the summary of Chapter 5 and 6, Section 1.2.3, Surface Based Morphometric Analysis. Finally, point (3) is touched upon by Chapter 7 and its summary, Section 1.2.4, Discrimination and Prediction of Disease States Based on Shape Deformation Markers.

### 1.2.1 Bayesian Parameter Estimation and Multi-atlas Likelihood Fusion

We present a parameter estimation approach in the framework of CA and apply it to the segmentation of brain structures in T1-weighted images. The estimation problem is approached in the setting of the multi-atlas random diffeomorphic orbit model, which is a natural extension of the random orbit model that was first established in [1]. We assume that there exist multiple MR images with the ROIs pre-delineated
in each of them. In our model, each MR image is a deformable atlas, each containing a
collection of locally-defined charts generated via the manually-defined ROIs. The esti-
mator is obtained via maximum a posteriori (MAP) estimation. The to-be-segmented
MR image is modeled as conditionally a Gaussian or a Gaussian mixture, conditioned
on the randomly selected atlas and the diffeomorphic change of coordinates of each chart in the selected atlas. The randomly selected atlas is unknown and viewed as
the latent variable. We demonstrate that the expectation-maximization (EM) algo-
rithm arises naturally, yielding the multi-atlas likelihood fusion algorithm which the
a posteriori estimator maximizes. For each atlas, the locally optimized diffeomorphic
correspondence of each chart is introduced into the E-step of the EM algorithm via
mode approximation. The likelihoods being fused are modeled as conditionally Gauss-
ian random fields with mean fields being functions of each atlas-specific chart under
its diffeomorphic change of coordinates. The conditional-mean in the EM algorithm
specifies the convex weights with which the chart-specific likelihoods are fused.

We derive and implement this Bayesian parameter estimation algorithm, and then
illustrate its application in two medical image processing procedures: skull-stripping
and segmentation of the subcortical structures in T1-weighted images. Segmentation
results and validations will be shown. In addition, we demonstrate the application of
this segmentation algorithm in several clinical studies.
1.2.2 Multi-contrast MALF based Segmentation of DTI

The T1-weighted image is a good modality for the segmentation of deep gray matter structures such as the caudate and the putamen. However, for the segmentation of more challenging brain structures, like various fiber tracts and cortical regions, we need DTIs. DTI can generate multiple quantitative maps with markedly different qualities of anatomical contrasts. The mean diffusivity (MD) contrast provides a clear distinction between the tissue (generally within the range of 0.6 - 0.9) and the CSF (approximately 3.0 - 3.5), providing a strong constraint to define the ventricles and the brain surface. The fractional anisotropy (FA) contrast provides a sharp distinction between the gray (typically FA < 0.15-0.25) and white matter structures (FA > 0.15-0.25). The eigenvector can differentiate intra-white matter structures based on their characteristic orientations.

We extend the MALF algorithm from single-contrast T1-weighted images to multi-contrast DTIs, with the purpose of segmenting the human brain into a total of 159 anatomical regions. To do this, we model the DTI as a high dimensional field, with each voxel exhibiting a vector-valued feature comprised of MD, FA, and fiber angle. For each structure, the probability distribution of each element in the feature vector is modeled as a mixture of Gaussians, the parameters of which are estimated from the labeled atlases. The structure-specific feature vector is then used to segment the test...
image. For each atlas, a likelihood is iteratively computed based on the structure-specific feature vector. The likelihoods from multiple atlases are then fused. The updating and fusion of the likelihoods are again accomplished in the EM framework, which we employed for the segmentation of T1-weighted images as well.

With regards to validation, we first demonstrate the performance of the algorithm by examining the segmentation accuracy of 18 gray and white matter structures from 25 subjects with a varying degree of structural abnormality. Then, we present the scan-rescan reproducibility of this multi-contrast MALF based segmentation algorithm using another dataset of 16 DTIs, each subject of which has been scanned twice.

1.2.3 Surface Based Morphometric Analysis

After extracting the various anatomical structures from the MR images in the aforementioned way, we move to the second component of CA – metric comparisons of manifolds. We describe and evaluate a morphometric comparison method based on the 2-D surfaces contouring the 3-D subvolume of the structures of interest. The surface based morphometric (SBM) model was originally proposed in [45], in which the metric is quantified as a scalar field indexed at each vertex of the common template surface – the determinant of the Jacobian matrix $D\phi$ of the diffeomorphism $\phi$ that maps the template surface to the subject surface. For each subject, we shall call this scalar field the “deformation marker” $J_k(\cdot)$ which is indexed at each vertex $k$ of the
template surface. In our statistical analysis, we employ a linear mixed-effects (LME) model by modeling the noise structure as a sum of two Gaussian processes. The LME model has been shown to be powerful in various MRI analyses [45, 79].

In the corresponding chapters, we will first describe the whole pipeline of the SBM analysis for analyzing subcortical and ventricular structures. Synthetic surfaces are then simulated to quantify the breadth of accuracy for this statistical analysis technique. Towards the end of the section, we present applications of this method to the morphological comparisons of the local surface volumes of fourteen subcortical and ventricular structures (left and right caudate, putamen, globus pallidus, thalamus, amygdala, hippocampus, and ventricle) in MCI and AD compared with normal aging. In the end, statistical group comparisons of the temporal evolution dynamics, in terms of the localized surface change rates along time, are illustrated using a similar LME model.

1.2.4 Discrimination and Prediction of Disease States Based on Shape Deformation Markers

After evaluating the metric comparison methods for anatomical manifolds (in our case, two types – the volumetric measurement and the surface deformation marker), it is natural to apply it to clinical diagnostics. The ultimate goal of medical image analysis is to be capable of automatically providing accurate diagnostic and prognos-
tic information about disease populations. It is essential to develop computer-aided methods to facilitate the predictive prognosis of which individuals are likely to deteriorate to disease status. The ability to identify an individual’s risk of developing a disorder is crucial for clinical decision-making.

In our approach to the third challenge of CA, we develop a manifold learning and clustering mechanism by applying the aforementioned surface morphometric features to the discrimination of AD and the prediction of conversion from MCI to AD. We work with the initial momentum vectors $\alpha_0$ in the setting of geodesic shooting [80]. Similar to the deformation marker $J_k(\cdot)$, $\alpha_0$ is also indexed at each vertex of the template surface. These vectors form a very high dimensional space. For each subject, $\alpha_0$ is of dimension $N \times 3$, where $N$ denotes the number of vertices on the template surface. Kernel based principal component analysis (PCA) is used to reduce the dimension of the original high-dimensional feature space. Then an automated feature selection procedure is employed to select the optimal classifier/predictor via linear discriminant analysis (LDA). In addition to differentiating 175 AD subjects from 210 normal aging ones, we validate the approach in predicting the MCI-to-AD conversion in 135 MCI subjects who deteriorated to AD within a follow-up of 3 years and 87 MCI subjects who did not convert.
Chapter 2

Bayesian Parameter Estimation and Segmentation in the Multi-atlas Random Orbit Model

2.1 Introduction

In CA, there are two principle objects of study, one being the anatomical manifolds (e.g. 1-D curves, 2-D surfaces, and 3-D subvolumes) that encode various biological signals of brain structures, the other being the dense imagery characterizing those structures (scalar or tensor). In the deformable template framework, the automated construction of those anatomical manifolds relies on a deformable template, in which the anatomical manifolds of interest have been pre-delineated, and this is the foun-
dation of the atlas-based approaches. Literally, a brain image atlas is a format that contains information on various brain structures so as to guide the delineation of those structures in other brain images. The simplest format is a pictorial representation of these brain structures, guiding anatomists in defining the boundaries between neighboring structures. In CA, a deformable atlas is defined by the pairing of a deformable template with its anatomical structure definitions.

Implicit in the notion of a deformable atlas is the existence of transformations that will connect the coordinates of the deformable templates to those of given targets. The transformations we use are diffeomorphisms \( \varphi \in G \), where \( G \) represents a group of bijective (invertible) smooth transformations, with smooth inverses, on the background space \( \mathbb{R}^3 \) where the imagery is defined, with law of composition \( \varphi \circ \phi(\cdot) = (\varphi(\phi(\cdot))) \) and inverse \( \varphi^{-1} \). Connections built by such transformations ensure that the target inherits the smoothness and the topological properties of the atlas. Within this deformable template and diffeomorphism framework, we are particularly interested in the automated construction of 3-D subvolumes representing various brain structures based on T1-weighted images. In our context, the deformable atlas is a collection of pre-segmented single-subject, population-averaged, or multiple-subject brain coordinate systems, including: whole brain coordinate systems [81–83], white matter coordinate systems [84–87], and surface coordinate systems [88–90], in which one can transfer information across different anatomical coordinate systems via diffeomorphisms.

The problem focused on in this chapter is to extend the random diffeomorphic orbit
model that has been extensively used in the automated segmentation of subvolumes based on a single atlas [1,28,91] to the multi-atlas setting. In the multi-atlas random diffeomorphic orbit model, not only are the diffeomorphic changes of coordinates unknown but also the jointly measurable parameters are hidden; such as those arising in: (1) atlas disease labels corresponding to disease inference; (2) structure parameters such as the volumetric measurement of a specific structure; (3) dense label field associated with segmentations of subvolumes. All three examples are posed in the augmented orbit of unions of multiple atlases. This makes our generative Bayesian probability model multi-modal, in which the posterior probability of interpreting a single image is shared across multiple atlases. The parameters, to be estimated, are not “isolated” from the simultaneous acquisition of the global shape phenotype, which is encoded via the structural definitions of each single atlas and the associated deformations.

Since the atlases used for interpreting the target image as well as the associated deformations are not known, the conditional-mean technology of the expectation-maximization (EM) algorithm [92] underlies the problem. As we will show, the conditional-mean explicates the weight with which each single atlas contributes to the interpretation of the target image in the multi-modal representation. In this setting, for every possible value of the parameter, there is a likelihood indexed over each atlas, which is then combined via superposition to generate the single a posteriori distribution that the Bayes MAP estimator optimizes. The superposed weights are
the conditional expectations of the latent variables determining the amount that each
atlas-specific likelihood is factored into the single a posteriori likelihood. We name
this method multi-atlas likelihood fusion (MALF).

A significant extension of the multi-atlas random orbit model is to add to the
global deformable template the notion of locality, which is usually associated to the
local approaches from differential geometry. Here an atlas is defined through a collec-
tion of local charts linked to each other via diffeomorphic transformations of coordi-
nates. Our anatomical model constructs the atlas via charts mapping open subsets of
the anatomical manifold to the subcortical and cortical segmentation labeled regions
that have been delineated in varying anatomical coordinate systems. In our case, we
focus on subcortical structures and the ventricles in T1-weighted images. Segmen-
tation via the MAP estimation assigns a single label to each voxel of the target image
with the maximum posterior probability, a mixture of the chart-specific conditional
a posteriori probabilities. Given that for any voxel, a chart from any of the atlases
could be the generator of its mean field and the associated conditional a posteriori
probability, the conditional-mean of the latent variable on chart selection is calculated
voxelwise, thus providing locality as part of the global model.

The essential aim of this chapter is to develop an automated segmentation method
for brain structures using the MALF algorithm. Segmenting brain structures using
multiple pre-segmented atlases has become popular recently. It has been shown that
simple label fusion techniques based on majority voting yield robust segmentations
More recently, weighted majority voting strategies, incorporating intensity information, demonstrated significant improvement in the segmentation accuracy. A variety of weighting approaches, based on intensity similarity metrics, have been proposed - global [11], local [17,18], semi-local [18], and non-local [19]. In addition to voting, a statistical fusion technique (i.e. Simultaneous Truth and Performance Level Estimation, STAPLE [14]) and a collection of its variants [93–95] have been proposed, in which a stochastic model of rater behavior has been incorporated in the estimation process. Compared with voting techniques, the main limitation of statistical fusion strategies is that the decision rule is independent of the image intensity, while the major advantage is its underlying elegant mathematical theory. Initial attempts to incorporate the intensity information into the STAPLE framework have relied on a priori similarity measures or estimating the voxelwise correspondence between the registered rater and the subject using intensity information [94].

Unlike previous label fusion algorithms, MALF does not fuse a set of binary label maps obtained from the atlas-to-subject propagations. MALF poses the segmentation problem in the multi-atlas random orbit model, estimating the maximizing segmentation labels given the observable image intensity. The MAP estimation is handled within the class of generative models by representing the observable imagery as a conditionally Gaussian mixture random field, conditioned on the randomly selected atlas. The selection of the atlas is unknown and viewed as the latent variable in the EM algorithm. The atlas-specific diffeomorphic change of coordinates is incorporated
into the probability model via mode approximation when computing the expected value of complete-data log likelihood. Locality was introduced by the fact that each deformable atlas consists of multiple local charts and, as a result, for a given atlas to interpret voxels in the target image, the local optimal diffeomorphism varies from label to label. The MAP estimation is solved by iterating between fixing the local optimal diffeomorphisms and obtaining the maximizing segmentation labels, and then locally optimizing the local diffeomorphisms for the fixed segmentation, in an EM fashion. The atlas-dependent structure-specific local diffeomorphisms are estimated in the M-step in the EM algorithm.

2.2 Method Derivation and Description

The goal of this section is to describe the working mechanism of Bayesian parameter estimation in the multi-atlas random diffeomorphic orbit model as well as to give a detailed description of the methodology in estimating the subcortical and ventricular structures in T1-weighted images.

2.2.1 The Random Deformable Atlas Orbit Model

We first give some background knowledge for the random diffeomorphic orbit model that the segmentation method is based on.

The Algebraic Orbit Model
CA examines the interplay between imaged anatomical structures, indexed implicitly or explicitly with a coordinate system, and their group of transformations. A transformation $\varphi \in G$ is applied to an anatomical structure or form $I \in \mathcal{I}$ lying in some set $\mathcal{I}$ called the anatomical orbit; the deformable template model of CA views these transformations as group actions. The transformation group $G$ generates the orbit of images through group actions.

The algebraic orbit model stipulates that an element $I \in \mathcal{I}$ and a transformation $\varphi \in G$ interact as a pair $(\varphi, I)$ via group action, denoted algebraically as $(\varphi, I) \mapsto \varphi \cdot I$. The actions of the group $G$ are introduced on the background space $G : X \rightarrow X$.

For our purpose, the transformations in $G$ are diffeomorphisms, bijective smooth coordinate changes with smooth inverse denoted by $\varphi^{-1}$, $\forall \varphi \in G$. The anatomical orbit $\mathcal{I}$ is a homogeneous space under the action of $G$ so that for all $\varphi \in G$, $\phi \in G$, $(\varphi \circ \phi) \cdot I = \varphi \cdot (\phi \cdot I)$, where $\circ$ denotes the group operation, and that for all $I \in \mathcal{I}$, $J \in \mathcal{I}$, there exists $\varphi \in G$ such that $J = \varphi \cdot I$.

Define the set of images $\mathcal{I}$ on $X \subset \mathbb{R}^d$, the background space, as $\mathcal{I} = \{I : X \rightarrow V\}$ with $V$ being the space of image values. A group action on the images, defined by the group of diffeomorphisms $G$, is as follows. The transformation $\varphi$ acts on the right via its inverse, $(\varphi, I) \mapsto \varphi \cdot I \doteq I \circ \varphi^{-1}$.

The Data Channel Model

In CA, observations are made of the underlying coordinates. Generally, the data
is observed in multiple forms, including measurements of both material properties of tissue, such as scalar and vector valued imagery derived from imaging systems, and geometric properties of objects associated with submanifolds, including curves, surfaces, and subvolumes.

We model the observed image $I^D \in \mathcal{I}^D$ as a random vector field, resulting from some distortion or noise process $D \in \mathcal{D}$ operating on an element $I$ in the anatomical orbit $\mathcal{I}$. The distortion or noise is introduced by the MRI, CT, nuclear emission, and optical and acoustic scanners. We define a distance measure $C$ between the underlying source or anatomical structure giving rise to the measurements and the observable measurement as $C : \mathcal{I} \times \mathcal{I}^D \to \mathbb{R}^+$. We will generally study this distortion process using a Gaussian random field model, in which the distance function arises as the potential associated with such a model. For this, we model the observable $I^D$ as a Gaussian random field with mean field $I \in \mathcal{I}$. Then the distance takes the form

$$C = \frac{1}{2\sigma^2} \| I^D - I \|_2^2,$$  \hspace{1cm} (2.1)

where $\|\cdot\|_2$ is the $L^2$ vector norm.

For the norm squared cost in Eq. (2.1), the stochastic model then corresponds to the additive white noise model, $I^D = I + \varepsilon$, with $\varepsilon$ being a white noise of variance 1.

**The Deformable Template Model**

The deformable template $\mathcal{I}_a$ corresponds to the orbit under the group $\mathcal{G}$ of one
selected and fixed image, which is termed the template image \( I_a \in \mathcal{I} \), such that

\[
\mathcal{I}_a = \mathcal{G} \cdot I_a = \{ I \in \mathcal{I} : I = I_a \circ \varphi^{-1}, \varphi \in \mathcal{G} \}. \tag{2.2}
\]

In the random diffeomorphic orbit model, when the deformable template \( I_a \) is known, the underlying intact image \( I \) of the observable \( I^D \) is assumed to be an element of the anatomical orbit \( \mathcal{I}_a \), a random deformation of the deformable template under the group action of \( \mathcal{G} \), denoted by \( I \in \mathcal{I}_a \) with \( \mathcal{I}_a = \mathcal{G} \cdot I_a \). The deformable template \( I_a \) serves as the starting point of the evolution process.

Due to the sheer complexity and biological variability of the human brain anatomy, it is insufficient and impractical to model the complete anatomy as the anatomical orbit of a single deformable template. Instead, the complete anatomy, \( \mathcal{I} \), is taken to be the union of the anatomical orbits evolved from multiple deformable templates,

\[
\mathcal{I} = \bigcup I_a = \bigcup \mathcal{G} \cdot I_a.
\]

Each element in a specific anatomical orbit \( \mathcal{I}_a \) is evolved from the initial deformable template \( I_a \) under the group of transformations \( \mathcal{G} \). Within this framework, we have added to our model a randomness in the deformable template image that generates the intact image \( I \).

### 2.2.2 Multi-atlas Likelihood Fusion

Literally, a brain atlas is a format that incorporates information on various brain structures so as to guide the delineation of those structures in other brain images. In CA, the definition of a deformable atlas is formulated by adding to the deformable
template image $I_a$ its anatomical structure definitions $W_a$, and is thus the pair $(I_a, W_a)$.

Assume that there exist $N$ atlas T1-weighted images with pre-delineated segmentation labels $\{(I_{a1}, W_{a1}), (I_{a2}, W_{a2}), \ldots, (I_{a2}, W_{a2})\}$, where $I_a$ denotes the T1-weighted image of atlas $a$ and $W_a : \Omega \rightarrow \{0, 1, 2, \ldots, n_{\text{Struct}}\}$ denotes the segmentation label image of atlas $a$ where $\Omega$ denotes the image domain, a bounded subset of the $\mathbb{R}^3$ space. The function $W_a$ maps to a subset of the non-negative integers; $W_a(x) = 0$ for voxel $x$ belonging to the unlabeled background, and $W_a(x) = k$, $k \in \{1, 2, 3, \ldots, n_{\text{Struct}}\}$ for voxel $x$ labeled as the $k$-th structure such as the left caudate, the right putamen, and so on. The segmentation label images of the atlases are usually pre-defined by neuroanatomists via manual delineation.

Our goal is to obtain the “optimal” estimator of the segmentation label $W$ based on the observable image $I^D$. We approach this problem via Bayesian estimation, due to the fact that it provides a natural and principled way to combine the prior information and the data. In addition, the Bayesian theorem ensures a straightforward incorporation of information from the set of deformable atlases $\{(I_a, W_a)\}$. We aim to maximize the conditional probability of the parameter $W$, conditioned on the observed image $I^D$. Mathematically,

$$\hat{W} = \arg \max_W p(W | I^D). \quad (2.3)$$

**Statement 2.1** The MAP estimator $\hat{W}$, defined in Eq. (2.3), can be equivalently
obtained via

\[ \hat{W} = \arg \max_W p(I^D, W), \quad (2.4) \]

where \( p(I^D, W) \) is computed via a fusion of likelihoods from multiple deformable atlases, as

\[ p(I^D, W) = \sum_a p(I^D, W|a)p_A(a), \quad (2.5) \]

where \( p_A(a) \) is the prior of atlas \( a \). The prior is selected to be some known probability distribution to limit the complexity of the family of models over all atlas classes.

Eq. (2.4) comes from the fact that \( p(I^D, W) = p(W|I^D)p(I^D) \) and that \( p(I^D) \) is constant with respect to the parameter \( W \). With multiple atlases generating the observed image, the fusion of the likelihood functions yields the multi-modal mixture model with the prior averaging over models. This is the generative model with which we score each atlas and the essence of the multi-atlas likelihood fusion (MALF) algorithm. Shown in Figure 2.1 is a depiction of brain segmentation in the multi-atlas framework.

### 2.2.3 Probability Model

Suppose that the set of deformable template images \( \{I_a\} \), the starting points of the evolution process, are given, the process of acquiring the observable image \( I^D \) can be mathematically formulated as \( I^D = I_A \circ \varphi^{-1} + \varepsilon \), where \( I_A \in \{I_a\}, \varphi \in \mathcal{G}, \) and \( \varepsilon \sim Gaussian \). Given this explicit mathematical connection between the observable \( I^D \)
Figure 2.1: Figure shows the population of atlases and the to-be-segmented target at the center. Each atlas consists of the deformable template image and the segmentation labels, represented via the pair \((I_a, W_a)\).
and the deformable template images \( \{I_a\} \), statistical estimation can be performed on any type of functional signals (random variables) that come along with the deformable template images. These functional signals can take several forms – the disease type of the subject associated to the observed image, the functional activation responding to a certain stimulus, the fiber tracking connectivity between different ROIs, or the segmentation labeling of the observed image.

When the application is to segment brain structures from MR images, the random variable encoding the functional signal on the deformable template is a mapping defined on the image domain, denoted here as \( W_a \). Statistical estimation of the segmentation label \( W \) in the observed image \( I^D \) can be made via learning the \( \{(I_a, W_a)\} \) and their mathematical connections to \( I^D \), leading to various atlas based segmentation methods.

Under the assumption that the source intact image \( I \) of the observable data comes from a random deformation (randomness of the diffeomorphism \( \varphi \)) of a random atlas (randomness of the selected atlas \( A \)), our probability model can thus be described as

\[
p(W, I^D, A = a, \varphi) = p(I^D|W, A = a, \varphi)p(W|\varphi, A = a)\pi(\varphi, A = a),
\]

where \( p(I^D|W, A = a, \varphi) \) is computed using a splitting model; conditioned on the segmentation label \( W \) and the selected atlas \( A = a \), the joint measurement of \( I^D \) and \( \varphi \) are conditionally independent. Namely, \( p(I^D|W, A = a, \varphi) = p(I^D|W, a) \). The quantity \( p(I^D|W, a) \) is modeled as a pre-selected probability density function. In our case, we use either a single Gaussian or a mixture of Gaussians depending on
the structure of interest. The parameters of the probability density functions are estimated from the selected atlas \( a \). The term \( p(W|\varphi, A = a) \) describes the prior information on the segmentation label \( W \), for which we use \( \text{Dice}(W, W_a \circ \varphi^{-1}) \times (W_a \circ \varphi^{-1})_{\text{TL}} \). The quantity \( \text{Dice}(W, W_a \circ \varphi^{-1}) \) is a scalar, measuring the distance between \( W \) and \( W_a \circ \varphi^{-1} \), and \( (W_a \circ \varphi^{-1})_{\text{TL}} \) is a 3-D matrix, deforming the atlas label \( W_a \) with the diffeomorphism \( \varphi \) under trilinear interpolation (TL). More details about these two terms will be covered in section 2.2.5. Prior information on the joint measurement of the atlas \( a \) and the diffeomorphism \( \varphi \), \( \pi(\varphi, A = a) \), will be described in detail in the subsequent section 2.2.5.

### 2.2.4 MALF via Expectation Maximization

The MAP problem defined in Eq. (2.4) can be solved iteratively via the expectation-maximization (EM) algorithm in which the observed image \( I^D \) is the incomplete data, the selected atlas \( A \) and the diffeomorphism \( \varphi \) together are the latent variable, making \((I^D, A, \varphi)\) the complete data, and \( W \) the parameter we would like to estimate.

**Statement 2.2** The EM iterations \( W^{\text{old}} \leftarrow W^{\text{new}} \) given by

\[
W^{\text{new}} = \arg \max_W Q(W; W^{\text{old}}) \\
= \arg \max_W E_{p(A, \varphi|W^{\text{old}}, I^D)} \{ \log p(W, I^D, A, \varphi)|W^{\text{old}}, I^D \} \\
= \arg \max_W \int \log p(W, I^D, A, \varphi) dp(A, \varphi|W^{\text{old}}, I^D)
\]  

(2.7)
is monotonically increasing in the log-likelihood as

$$\log p(I^D, W^{new}) \geq \log p(I^D, W^{old}).$$  \hspace{1cm} (2.8)$$

The proof of **Statement 2.2** can be found in the Appendix.

To simplify the Q-function

$$Q(W; W^{old}) = \int \log p(W, I^D, A, \varphi) dp(A, \varphi|W^{old}, I^D)$$

$$= \sum_a \int \log p(W, I^D, a, \varphi) dp(A = a, \varphi|W^{old}, I^D)$$ \hspace{1cm} (2.9)$$

we use mode approximation; the probability measure $dp(A = a, \varphi|W^{old}, I^D)$ is approximated with a dirac measure such that

$$dp(A = a, \varphi|W^{old}, I^D) = p(A = a, \varphi|W^{old}, I^D) d\delta_{\hat{\varphi}_a}(\varphi).$$  \hspace{1cm} (2.10)$$

Under this approximation and the aforementioned splitting model, the Q-function is thus computed as

$$Q(W; W^{old}) = \sum_a \hat{p}(A = a, \hat{\varphi}_a|W^{old}, I^D) \log p(W, I^D, a, \hat{\varphi}_a)$$

$$= \sum_a \hat{p}(A = a, \hat{\varphi}_a|W^{old}, I^D) \log p(I^D|W, a)p(W, a, \hat{\varphi}_a)$$ \hspace{1cm} (2.11)$$

where

$$\hat{\varphi}_a = \arg \max_{\varphi} \ p(a, \varphi|W^{old}, I^D)$$

$$= \arg \max_{\varphi} \ p(I^D|W^{old}, a, \varphi)p(W^{old}, a, \varphi)$$

$$= \arg \max_{\varphi} \ p(W^{old}|a, \varphi)\pi(\varphi|a).$$  \hspace{1cm} (2.12)$$
The term $\hat{p}(A = a, \hat{\varphi}_a | W^{old}, I^D)$ is termed as the “atlas-selector” function, measuring how much each atlas contributes to interpreting the observable image $I^D$. This “atlas-selector” function is computed as

$$
\hat{p}(A = a, \hat{\varphi}_a | W^{old}, I^D) = \frac{p(a, \hat{\varphi}_a | W^{old}, I^D)}{\sum_{a'} p(a', \hat{\varphi}_{a'} | W^{old}, I^D)} = \frac{p(I^D | W^{old}, a)p(W^{old}, a, \hat{\varphi}_a)}{\sum_{a'} p(I^D | W^{old}, a')p(W^{old}, a', \hat{\varphi}_{a'}).}
$$

We now present the steps involved in each iteration of the EM algorithm in Algorithm 1.

**Algorithm 1 (The EM algorithm for the MAP estimation via MALF)**

We initialize $W^{old}, \hat{\varphi}_a^{old}$ for each atlas $a$.

1. Compute the “atlas selector” function $\hat{p}(A = a, \hat{\varphi}_a^{old} | W^{old}, I^D)$ as in Eq. (2.13).

2. Compute the Q-function as in Eq. (2.11).

3. Obtain a new estimator of the parameter via $W^{new} = \arg \max_W Q(W; W^{old})$.

4. Update $W^{old} \leftarrow W^{new}$, and for each atlas $a$, compute a new optimal diffeomorphism $\hat{\varphi}_a^{new}$ as in Eq. (2.12).

5. Update $\hat{\varphi}_a^{old} \leftarrow \hat{\varphi}_a^{new}$, go to step 2.

The above iteration is continued until convergence of the estimator or the number of performed iterations is bigger than a pre-specified threshold.

Given the smoothness of the brain anatomy, we model it as a $C^\infty$ smooth manifold. By definition of a smooth manifold, there exists a family of smooth local coordinate
charts covering the manifold. For each smooth local coordinate chart \((O_\alpha, \phi_\alpha)\), \(\phi_\alpha\) is a diffeomorphism from an open neighborhood \(O_\alpha\) of the manifold to an open set in the Euclidean space. The charts \(\{(O_\alpha, \phi_\alpha)\}\) are pairwise compatible so that \(O_\alpha \cap O_\beta \neq \emptyset\) implies that the map \(\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(O_\alpha \cap O_\beta) \rightarrow \phi_\alpha(O_\alpha \cap O_\beta)\) gives a smooth change of coordinates.

With this in mind, we introduce the locally optimized diffeomorphism for each open set covering the submanifold representing each brain structure of interest. In other words, we assume that for a given atlas \(a\), the locally optimized diffeomorphism \(\hat{\phi}_a(x)\) may vary depending on the anatomical label of the spatial location \(x \in \Omega\), i.e. \(\hat{\phi}_a(x) = \hat{\phi}_a(y)\) when \(W(x) = W(y)\).

The local charts for different submanifolds are related to each other via diffeomorphic coordinate transformations. As described in Figure 2.2, two points \(X\) and \(Y\) belonging to the submanifolds that respectively correspond to the hippocampus and the amygdala are related to each other via forward and inverse mappings as \(Y = \varphi_a^{-1} \circ \varphi_h(X)\) and \(X = \varphi_h^{-1} \circ \varphi_a(Y)\), where \(\varphi_a\) and \(\varphi_h\) respectively denote diffeomorphisms transforming the local coordinates embedding the amygdala and the hippocampus. This overlap resulting from multiple local charts allows for a weighted interpretation from different atlases, given that all “mediation” of errors occurs at the boundary of neighboring structures. At one boundary of the hippocampus, for example, are portions of the lateral ventricle while at another are regions belonging to the amygdala. Weighted interpretation of those boundary voxels is supported by
multiple local charts which may overlap and thus offer alternative contributions.

After introducing the locally optimized diffeomorphisms, the “atlas selector”
\[ \hat{p}(A = a, \hat{\varphi}_a(x) | W^{\text{old}}(x), I^D(x)) \]
becomes localized as well, representing the conditional probability that atlas \( a \) is selected to interpret voxel \( x \) in the observed image, conditioned on the observed image and the segmentation label. Similar to Eq. (2.13), we have

\[ \hat{p}(a, \hat{\varphi}_a(x) | W^{\text{old}}(x), I^D(x)) = \frac{p(I^D(x) | W^{\text{old}}(x), a)p(W^{\text{old}}(x), a, \hat{\varphi}_a(x))}{\sum_{a'} p(I^D(x) | W^{\text{old}}(x), a')p(W^{\text{old}}(x), a', \hat{\varphi}_{a'}(x))}. \] (2.14)

With locality introduced, each step of the EM iteration is summarized in Algorithm 2.

**Algorithm 2 (The EM algorithm for the MAP estimation via MALF, with an incorporation of locally optimized diffeomorphisms)**

We initialize \( W^{\text{old}}(x), \hat{\varphi}_{a}^{\text{old}}(x), \forall x \in \Omega \) for each atlas \( a \).

1. Compute the “atlas selector” function \( \hat{p}(a, \hat{\varphi}_a^{\text{old}}(x) | W^{\text{old}}(x), I^D(x)) \) as in Eq. (2.14).

2. Compute the quantity \( Q(W; W^{\text{old}}) \) as

\[ \sum_{x \in \Omega} \sum_a \hat{p}(a, \hat{\varphi}_a^{\text{old}}(x) | W^{\text{old}}(x), I^D(x)) \log p(I^D(x) | W(x), a)p(W(x), a, \hat{\varphi}_a^{\text{old}}(x)). \] (2.15)

3. Obtain a new estimator of the parameter via \( W^{\text{new}} = \arg \max_W Q(W; W^{\text{old}}) \).
Figure 2.2: Depiction of two local charts and the associated diffeomorphisms chosen to illustrate the weighted interpretation. The local charts are related to each other via diffeomorphic coordinate transformations in a way illustrated in the figure. Two points $X$, $Y$ that respectively belong to the hippocampus chart and the amygdala chart are compared using the forward and inverse mappings.
4. For each atlas $a$, update the locally optimized diffeomorphisms via

$$
\hat{\varphi}^\text{new}_a(x) = \hat{\varphi}^\text{new}_k, \ \forall x : W^\text{new}(x) = k, \tag{2.16}
$$

where

$$
\hat{\varphi}^\text{new}_k = \arg \max_{\varphi} p(W^\text{new}_k | a, \varphi) \pi(\varphi | a). \tag{2.17}
$$

5. Update $W^\text{old} \leftarrow W^\text{new}$, $\hat{\varphi}^\text{old}_a(x) \leftarrow \hat{\varphi}^\text{new}_a(x), \ \forall x \in \Omega$, go to step 2.

The above iteration is continued until convergence of the parameter estimator or the number of performed iterations is bigger than a pre-specified threshold.

## 2.2.5 Implementation of MALF for Brain Segmentation Label Estimation

In this section, we give details on the implementation of the MALF algorithm for estimating the brain segmentation labels. To be specific, we introduce the computation of each step in Algorithm 2.

- **Initialization**

  For every atlas $a$, initialize the local optimal diffeomorphism to be the same everywhere on the image domain $\hat{\varphi}^0_a(x) = \hat{\varphi}^0_a, \ \forall x \in \Omega$. Given our diffeomorphic orbit framework, we obtain this initial global optimal diffeomorphism from the
LDDMM algorithm for images, which was originally proposed in [33]. In LDDMM, the diffeomorphism that connects two coordinate systems is constructed from a flow of ordinary differential equations (ODEs).

Given the target image $I^D$ and the atlas image $I_a$, LDDMM computes a diffeomorphic deformation $\varphi$ between the two images such that $I^D = I_a \circ \varphi^{-1}$.

This diffeomorphism is assumed to be generated as the end point, $\varphi = \phi_v^T$, of the flow of a smooth time-dependent vector field, $v_t \in V, t \in [0, 1]$, via the ordinary differential equation $\frac{d\phi_v^T}{dt} = v_t(\phi_v^T), t \in [0, 1]$, where $\phi_0$ is the identity transformation. To ensure that these ODEs generate diffeomorphisms, the set of time-indexed vector fields $v_t$ must be sufficiently spatially smooth. We thus require $V$ be a Hilbert space of smooth and compactly supported vector fields on the background space [91] with norm $\| \cdot \|_V$. One choice is to construct $V$ as the completion of a space of smooth, compactly-supported vector fields with the inner-product defined through a differential operator $L$ (with adjoint denoted by $L^\dagger$), given by:

$$\langle f, g \rangle_V \doteq \langle Lf, Lg \rangle_{L^2} = \langle L^\dagger Lf, g \rangle_{L^2}, \quad (2.18)$$

with $L = -\alpha \nabla^{2p} + \gamma$, where $\nabla^{2p}$ is the Laplacian operator with power $p \geq 1$.

In MALF, $p = 1$, $\gamma = 1$ and $\alpha$ is selected according to the cascading method, described in [36] as $0.01 - 0.005 - 0.002$.

For a fixed atlas $a$, considering our data channel model $I^D = I_a \circ \varphi^{-1} + \varepsilon$, with
\( \varepsilon \sim \text{Gaussian} \), we define a distance measure between the underlying source \( I \) giving rise to the measurements, and the observable measurement \( I^D \) as
\[
C = \frac{1}{2\sigma^2} \| I^D - I_a \circ \varphi^{-1} \|^2_{L^2},
\]
where \( \| \cdot \|_{L^2} \) is the \( L^2 \) vector norm. Given the observed \( I^D \) with mean field \( I_a \circ \varphi^{-1} \) and matching cost \( C \), the goal of LDDMM is to find the diffeomorphism \( \varphi \) that minimizes \( C \). As with many minimum-norm variational problems, there are many possible paths \( t \to \phi^v_t \) that can generate a diffeomorphism minimizing the matching cost \( C \). Therefore, we choose the ones that also minimize the length (which can also be viewed as the energy) of the path \( t \to \phi^v_t \) connecting the two diffeomorphisms \( \phi^v_0 \) and \( \phi^v_1 \), which is quantified by \( E(v) = \int_0^1 \| v_t \|^2_V \, dt \), in addition to minimizing the matching cost \( C \). The quantity \( E(v) \) measures the amount of deformation induced on the whole space. It also denotes the square of the total length of the path \( t \to \phi^v_t \).

To be specific, the optimal diffeomorphic deformation is generated by integrating the vector field that solves
\[
\dot{v}_t = \arg \min_{v_t: \frac{\partial \phi^v_t}{\partial t} = v_t(\phi^v_t)} \left( \int_0^1 \| v_t \|^2_V \, dt + \frac{1}{\sigma^2} \| I_a \circ \varphi^{-1} - I^D \|^2_{L^2} \right), \tag{2.19}
\]
where \( \sigma \) controls the weighting of the image-matching term \( \| I_a \circ \varphi^{-1} - I^D \|^2_{L^2} \) relative to the smoothness regularization term \( \int_0^1 \| v_t \|^2_V \, dt \). We choose \( \sigma = 1 \).

A steepest gradient descent approach is employed in minimizing Eq. (2.19), in which at iteration \( k \) the velocity vector field \( v_t \) is updated via
\[
v_t^{k+1} = v_t^k - \varepsilon \left( \nabla_{v_t^k} E_t \right),
\]
where \( \nabla_{v_t^k} E_t \) is the gradient of the cost function shown in Eq. (2.19).
with respect to $v_t$, computed as:

$$\nabla_{v_t} E_t = 2v_t - \left(L^\dagger L\right)^{-1} * \left(\frac{2}{\sigma^2} |D\phi_t|^r \nabla J^0_t (J^0_t - J^1_t)\right). \tag{2.20}$$

In Eq. (2.20), $\phi_{s,t} = \phi_t \circ \phi_s^{-1}$, $J^0_t = I_a \circ \phi_{t,0}$, $J^1_t = I^D \circ \phi_{t,1}$, $|D\phi_t|^r$ is the determinant of the Jacobian matrix, and $*$ is the convolution operation. In the numerical implementation, the time parameter $t$ is discretized with a fixed number of time-steps $T$, the default value of which is 10 in LDDMM-image mapping. The smaller the time-step parameter $T$ is, the more computationally efficient the mapping is, but at the cost of a larger final mismatch error $\|I_a \circ (\phi^0_1)^{-1} - I^D\|_L^2$ between the registered atlas image and the target image.

For the initialization of the segmentation label $W^{(0)}$, we use the propagation of the segmentation label of atlas $a_1$ under the initial optimal global diffeomorphism $\varphi^0_{a_1}$, $W^0 = W_{a_1} \circ (\varphi^0_{a_1})^{-1}$.

**Computation of $p(I^D(x)|W^{odd}(x), a)$**

In calculating $p(I^D(x)|W^{odd}(x), a)$, we model it as the probability density function of a single Gaussian distribution, the parameters of which are computed from the selected atlas. To be specific, for atlas $a$, the distribution of the image intensity $I_a(x)$ in each ROI $k$ is estimated as a Gaussian,

$$p(I_a(x)|W_a(x) = k) = \frac{1}{\sqrt{2\pi\sigma_{ak}}} \exp\left\{ -\frac{(I_a(x) - \mu_{ak})^2}{2(\sigma_{ak})^2} \right\}, \tag{2.21}$$
and
\[
p(I^D(x)|W^{old}(x) = k, a) = \frac{1}{\sqrt{2\pi\sigma_{ak}}} \exp\left\{-\frac{(I^D(x) - \mu_{ak})^2}{2(\sigma_{ak})^2}\right\}.
\] (2.22)

The parameter $\mu_{ak}$ is estimated as the empirical mean and $(\sigma_{ak})^2$ the empirical variance.

- **Computation of $p(W(x), a, \hat{\varphi}_a(x))$**

Another quantity we need to compute is $p(W(x), a, \hat{\varphi}_a(x))$. According to Bayes’ rule, we have
\[
p(W(x), a, \hat{\varphi}_a(x)) = p(W(x)|a, \hat{\varphi}_a(x))\pi(a, \hat{\varphi}_a(x)).
\] (2.23)

The term $p(W(x)|a, \hat{\varphi}_a(x))$ computes the prior information on the segmentation label $W(x)$. Based on empirical experience and a need for computational efficiency, we compute this as the multiplication of two terms,
\[
p(W(x)|a, \hat{\varphi}_a(x)) = Dice(W_k, W_{ak} \circ \hat{\varphi}_{ak}^{-1}) \times \left(W_{ak} \circ (\hat{\varphi}_a^0)^{-1}\right)(x)_{TL},
\] (2.24)
for all $x$ such that $W(x) = k$, where $k$ denotes different brain structures. The term $Dice(W_k, W_{ak} \circ \hat{\varphi}_{ak}^{-1})$ computes the Dice overlap measurement [96] between the $k$-th structure of $W$ and that of $W_{a} \circ \hat{\varphi}_{a}^{-1}$, and $\left(W_{ak} \circ (\hat{\varphi}_a^0)^{-1}\right)(x)_{TL}$ denotes the result of deforming the $k$-th structure of the atlas segmentation label $W_a$ with the initial global diffeomorphism $\varphi_a^0$ under trilinear interpolation (TL). The calculation of $\pi(a, \hat{\varphi}_a(x))$ will be described while updating the locally optimized diffeomorphisms.
• **Update of local diffeomorphisms** \( \hat{\varphi}_{ak} \)

To find the optimized local diffeomorphism for each ROI as in Eq. (2.17), we assume that the optimal local diffeomorphism comes from a composition of the optimal global diffeomorphism obtained in the initialization and a locally optimized 12-parameter affine transformation, \( \hat{\varphi}_{ak} = \hat{\varphi}^0_a \circ \hat{\alpha}_{ak} \). \( \hat{\varphi}^0_a \) is the optimal global diffeomorphism computed in the initialization step and \( \hat{\alpha}_{ak} \) is a locally optimized 12-parameter affine transformation for the \( k \)-th structure of interest.

According to Eq. (2.24) and Eq. (2.17), we have

\[
\hat{\varphi}_{ak}^{new} = \arg \max_{\varphi} p(W_{\text{new}}^k | a, \varphi) \pi(\varphi | a)
= \arg \max_{\varphi} \left( \text{Dice} \left( W_{\text{new}}^k, W_{ak} \circ \varphi^{-1} \right) \times \left( W_{\text{new}}^k \circ (\varphi^0_a)^{-1} \right)_{\text{TL}} \times \pi(\varphi | a) \right)
= \arg \max_{\varphi} \left( \text{Dice} \left( W_{\text{new}}^k, W_{ak} \circ \varphi^{-1} \right) \pi(\varphi | a) \right).
\] (2.25)

For atlas \( a \), the prior distribution of the transformation \( \pi(\varphi | a) \) is estimated as the multiplication of two terms

\[
\pi(\varphi | a) = \pi(\varphi^0_a | a) \pi(\alpha | a, \varphi^0_a).
\] (2.26)

Given Eq. (2.25) and Eq. (2.26), we therefore obtain the locally optimized 12-parameter affine transformation matrix via

\[
\hat{\alpha}_{ak} = \arg \max_{\alpha} \text{Dice} \left( W_{\text{new}}^k, W_{ak} \circ (\varphi^0_a \circ \alpha)^{-1} \right) \pi(\alpha | a, \varphi^0_a).
\] (2.27)

In computing the prior terms in Eq. (2.26), \( \pi(\varphi^0_a | a) \) is estimated via one over the metric distance [28] in diffeomorphism space given by the exponential of
the geodesic length, which is computed from the LDDMM-image mapping. The prior on the 12-parameter affine transformation \( \pi(\alpha | a, \varphi_0^a) \) is modeled as a multivariate Gaussian (i.e. \( \pi(\alpha | a, \varphi_0^a) = N(M_\alpha, C_\alpha) \)), similar to the strategy adopted in [97]. In our approach, we use \( M_\alpha = [1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0]^T \) and \( C_\alpha = \text{diag}(\{1e^{-2}, 1e^{-2}, ..., 1e^{-2}, 1e^2, 1e^2, 1e^2\}) \). We assume that all parameters are mutually independent, and thus the covariance matrix is diagonal. Since the first 9 parameters in \( \alpha \) represent the affine matrix, their variances should be small, for which we assign 0.01. The last 3 parameters respectively represent the translation in the \( x, y, z \) directions, and therefore their variances should be big, for which we use 100.

To sum up, the MAP estimation problem is solved in an EM approach. We iterate between fixing the local optimal diffeomorphism for each structure in each atlas and obtaining the maximizing segmentation label of the observed image, and then locally optimizing the diffeomorphisms associated to each local label in each atlas given the fixed segmentation.

### 2.3 Application

To perform image registration between paired T1-weighted brain images, a prerequisite is often to isolate the brain from other “non-brain” regions, a process usually referred to as “skull-stripping”. To be feasible for large-scale neuroimaging studies,
it is desirable that skull-stripping methods be accurate and automated as much as possible. A number of automated approaches for skull-stripping have been proposed and are widely used in the neuroimaging community [98–103]. Besides its original proposal for the segmentation of internal brain structures, the MALF algorithm can be naturally applied to removing the skull from T1-weighted brain images. We will now present an automated pipeline for skull-stripping and segmenting brain structures from T1-weighted images in a unified manner. This pipeline is built on a two-level hierarchical MALF; the first level strips the skull from the input T1-weighted image while the second takes this output image and aims to segment various internal brain structures.

2.3.1 Two-level Hierarchical Brain Segmentation

The pipeline of the two-level hierarchical brain segmentation is illustrated in Figure 2.3. After obtaining the T1-weighted image of the to-be-segmented subject from the scanner, we first re-orient the image and correct the inhomogeneity of the image intensity. We then perform the first-level segmentation to obtain the skull-stripped T1-weighted image of the subject, followed by the second-level segmentation to obtain the segmentations of brain structures of interest.

Since there are two stages of segmentation, there are correspondingly two sets of pre-delineated segmentation labels for the atlas dataset. The first set, used for skull-stripping, consists of six global labels – the whole-brain gray matter (GM), the whole-
brain white matter (WM), the whole-brain cerebrospinal fluid (CSF), the lateral ventricle (LV), the skull, and the background of the entire image. The T1-weighted image of each atlas is first manually skull-stripped by neuroanatomists to separate brain tissue from the non-brain regions. The manual skull-stripping is usually performed using the software RoiEditor (http://www.MriStudio.org, Kennedy Krieger Institute and Johns Hopkins University, X. Li, H. Jiang, and S. Mori). Two global labels, the skull and the background of the image, are then created by thresholding the non-brain regions in the T1-weighted images. The lateral ventricle labels are obtained from manual tracing. The whole-brain GM/WM/CSF labels are created by performing unified segmentation on the manually skull-stripped T1-weighted images using the tissue segmentation algorithm [97], which is incorporated in the Statistical Parametric Mapping (SPM) software. Given that the whole-brain GM/WM/CSF labels come from an automated tissue segmentation algorithm, we would expect inaccuracy of these three labels to some degree. However, since we are only concerned with the skull-stripping result, the mild inaccuracy of the internal GM/WM/CSF label definitions is acceptable. We expect that the extraction of the brain mask would not be affected by the inaccuracy of the three GM/WM/CSF labels as long as the boundary separating the brain from the “non-brain” regions is sufficiently accurate. Regarding the second-level of segmentation, labeling of the brain structures of interest for each atlas is obtained from manual delineations.

Each step of the hierarchical segmentation pipeline is detailed as follows:
1. **Preprocessing:** After the T1-weighted image of the subject is acquired from the scanner, its orientation is adjusted so that it matches the orientation of the atlas images. We use axial view as our standard orientation. The inhomogeneity of the intensity in the T1-weighted image of the subject is then corrected using a nonparametric non-uniform intensity normalization method N3 [104]. This inhomogeneity correction method is selected because of its applicability to T1-weighted images of the human brain with skull and its outstanding performance compared with other methods [105].

2. **Skull-stripping:** To obtain the skull-stripped T1-weighted image of the input subject, we first perform a fast version of MALF to estimate the six global labels (LV, GM, WM, CSF, skull, and the background) of the subject. The term “fast version” is applied for two reasons. First of all, the global 6-label estimation is performed on down-sampled T1-weighted images. We down-sample both the subject’s T1-weighted image and the atlas T1-weighted images from the original resolution of $1mm \times 1mm \times 1mm$ to $2mm \times 2mm \times 2mm$, and therefore decrease the image size by half. Secondly, when we estimate the initial global diffeomorphisms using LDDMM, we use small deformations. The time-step parameter, which is used to discretize the time parameter $t$ in Eq. (2.20) in LDDMM-image mapping, is selected to be $T = 2$ to approximate small deformations. The down-sampling of the images and the initial small deformations makes the entire likelihood-fusion procedure significantly faster.
The four labels, LV/GM/WM/CSF, are grouped to create the brain mask which is then morphologically post-processed using a two-step process. First, holes in the mask are filled using an iterative voting process. If the majority of the 26 neighbors of a voxel are foreground, the voxel is marked as foreground. Otherwise it is marked as background. Likewise, if the majority of the neighbors are background, then the voxel is marked as background. Otherwise it is marked as foreground. If the votes for foreground and background are equal, we label the voxel as background. This process is repeated until a maximum number of ten iterations is reached or until the brain mask remains unchanged. Second, morphological erosion with a 1 voxel radius ball is applied to remove small objects and excess foreground voxels introduced by the previous hole-filling process.

3. Brain structure identification: After skull-stripping the T1-weighted image of the input subject, we then perform the regular version of MALF to estimate the internal structures of interest. By regular, we mean that the input skull-stripped T1-weighted images, for both the to-be-segmented subject and the atlas dataset, are of original 1mm × 1mm × 1mm resolution, and that the initial global optimal diffeomorphisms are obtained from the LDDMM image mapping using the default time-step parameter $T = 10$. 
2.4 Validation Results

In our experiments, we validate the MALF algorithm in skull-stripping as well as segmentation of subcortical and ventricular structures in T1-weighted images. The validation analysis on skull-stripping is performed on two datasets: (1) a mixture of normal aging and dementia subjects acquired from a 1.5 Tesla Spoiled Gradient Echo (SPGR) scanner; (2) a mixture of cognitively normal pediatric subjects, children with Autism, as well as children with ADHD, the T1-weighted images of which were all acquired from a 3.0 Tesla Magnetization Prepared Rapid Gradient Recalled Echo (MPRAGE) scanner. In terms of the segmentation of subcortical structures and ventricles, the validation is performed with respect to a variety of datasets, including:
images of subjects with different age ranges (pediatric, young adults, and elderly adults), images of subjects with different disease states (healthy, Autism, ADHD, PPA, as well as dementia), and images of subjects obtained from different types of scanners (1.5 Tesla SPGR and 3.0 Tesla MPRAGE).

2.4.1 Validation of MALF in Skull-stripping

2.4.1.1 Pediatric healthy control/Autism/ADHD Dataset

For the first dataset used for skull-stripping validation, whole brain, high resolution T1-weighted 3D-volume MPRAGE images (matrix size=256×256, echo time=3.76 ms, repetition time=7.99 ms, field of view=256 × 256 mm, slice thickness=1.0 mm) of 30 subjects were acquired from a 3T Philips Gyroscan NT scanner (Royal Philips Electronics, Amsterdam, The Netherlands). This dataset includes 13 healthy control subjects (mean age: 10.42 years old; 5 males and 8 females), 6 male subjects with ASD (mean age: 9.74 years old) and 11 subjects diagnosed with ADHD (mean age: 10.2 years old; 4 males and 7 females).

In order to quantitatively assess the performance of the skull-stripping procedure, we compare the automated skull-stripping results with the manual delineations on six pre-selected sagittal slices (Figure 2.4) in all 30 subjects, using the Dice score coefficient (DSC). The Dice score coefficient is calculated as: $D = \frac{2TP}{2TP + FP + FN}$, where $TP$ is the area of the region in which the automated result overlaps with the manual
one, $FP$ is the area of the region that belongs to the automated result but not the manual, and $FN$ is the area of the region that belongs to the manual but not the automated result. The selection of the six sagittal slices follows the procedure suggested in [106].

![Figure 2.4: Standard locations of the six pre-selected sagittal slices (demonstrated on a coronal image) that have been manually skull stripped for validation analysis. From left to right: slice1, slice2, slice3, slice4, slice5, and slice6.](image)

The performance of MALF, in terms of skull-stripping, is compared, both quantitatively and qualitatively, with two of the most widely used skull-stripping methods.
Table 2.1: Mean and standard deviations of the dice score coefficients (DSCs) of the skull-stripping results obtained from the three automated approaches – MALF, HWA, and BET, when compared with the corresponding manually stripped slices.

<table>
<thead>
<tr>
<th></th>
<th>slice1</th>
<th>slice2</th>
<th>slice3</th>
<th>slice4</th>
<th>slice5</th>
<th>slice6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALF</td>
<td>0.962 (0.013)</td>
<td>0.981 (0.007)</td>
<td>0.979 (0.006)</td>
<td>0.979 (0.006)</td>
<td>0.982 (0.005)</td>
<td>0.955 (0.022)</td>
</tr>
<tr>
<td>HWA</td>
<td>0.965 (0.012)</td>
<td>0.979 (0.016)</td>
<td>0.969 (0.026)</td>
<td>0.974 (0.021)</td>
<td>0.979 (0.015)</td>
<td>0.952 (0.017)</td>
</tr>
<tr>
<td>BET</td>
<td>0.944 (0.019)</td>
<td>0.946 (0.023)</td>
<td>0.919 (0.025)</td>
<td>0.918 (0.024)</td>
<td>0.943 (0.024)</td>
<td>0.888 (0.039)</td>
</tr>
</tbody>
</table>

– HWA [99] incorporated in the software Freesurfer (version 5.2.0) and BET [101] incorporated in the software FSL (version 5.0). Table 2.1 shows the mean dice score coefficients and the standard deviations, computed across all the 30 subjects, for each of the six slices. According to a paired Student’s t-test, the DSCs between the skull-stripping results from MALF and those from the manual delineations are statistically equivalent to the DSCs between the skull-stripping results from HWA and the manual approach for all slices ($p > 0.06$). The skull-stripping results from both MALF and HWA are statistically superior to that from BET in terms of DSCs ($p < 1e^{-5}$) for each slice.

Although the automated skull-stripping results, of the six pre-selected sagittal slices, obtained from HWA are quantitatively comparable to those obtained from MALF, qualitative visual reviews of all individual subjects reveal that MALF is superior to HWA in several aspects. First, HWA sometimes misses cortical and cerebellar tissue. In two cases, it completely excludes the entire cerebellum region. Second, HWA acts aggressively in regions close to the external dura, tending to include all
the external dura. Lastly, the brain contour of the skull-stripped images from HWA is not smooth in a number of cases. Those three types of errors were not found in the MALF results. Visual examination also suggests that BET consistently under-estimates brain regions. In Figure 2.5, we demonstrate the skull-stripping results, obtained from the three methods, for two representative subjects. It is clear that for subject A) both HWA and BET miss some cortical brain tissue. For subject B), HWA tends to include more “non-brain” tissue while BET misses some brain tissue. As illustrated by the yellow contour lines shown in Figure 2.5, the boundaries of the skull-stripped brain images from both HWA and BET are not as smooth as those from MALF.

2.4.1.2 BIOCARD Dataset

The second dataset we use to examine the performance of MALF in skull-stripping came from a longitudinal study characterizing the structural MRI abnormalities in subjects with dementia of the Alzheimer type (known as the BIOCARD study). All subjects were cognitively normal when they were recruited. The mean age of the BIOCARD subjects at baseline was 57.1 years. Scans were acquired during the period 1995 - 2005. The participants have been followed for up to 17 years.

MRI scans were obtained on 335 participants at baseline. An additional 470 scans were obtained in subsequent years, resulting in a total of 805 scans. The mean interval between scan acquisitions was 2.02 years. The MRI scans acquired at the NIH were
Figure 2.5: Panel A) and Panel B) respectively show the automated skull-stripping results of two subjects. The boundaries of the skull-stripping results are superimposed on the corresponding T1-weighted images. From left to right: results obtained from MALF, HWA, and BET.
obtained using a standard multi-modal protocol using GE 1.5T scanner. The scanning protocol included localizer scans, Axial FSE (Fast Spin Echo) sequence (TR=4250, TE=108, FOV=512×512, thickness/gap=5.0/0.0 mm, flip angle=90, 28 slices), Axial Flair sequence (TR=9002, TE=157.5, FOV=256×256, thickness/gap=5.0/0.0 mm, flip angle=90, 28 slices), Coronal SPGR (Spoiled Gradient Echo) sequence (TR=24, TE=2, FOV=256×256, thickness/gap=2.0/0.0 mm, flip angle=20, 124 slices), Sagittal SPGR (Spoiled Gradient Echo) sequence (TR=24, TE=3, FOV=256×256, thickness/gap=1.5/0.0 mm, flip angle=45, 124 slices).

Skull-stripping via MALF was performed on a total of 400 scans and qualitative evaluation was conducted by several neuroanatomists. Figure 2.6 shows the skull-stripping results, of three representative scans, from MALF, HWA, and BET. It is apparent that MALF works much better than both HWA and BET in removing the skull from T1-weighted images, especially those with poor image quality.
Figure 2.6: The top, middle, and bottom panels respectively show the automated skull-stripping results of three scans from the BIOCARD study. The outlines of the masks are superimposed on the corresponding T1-weighted images. From left to right, results are obtained from MALF, HWA, and BET.
2.4.2 Validation of MALF in Segmenting Subcortical and Ventricular Structures

2.4.2.1 Pediatric healthy control/Autism/ADHD Dataset (3T MPRAGE)

The pediatric dataset that consists of a mixture of healthy control subjects, subjects with Autism, and subjects with ADHD (the dataset described in section 2.4.1.1) is also used in validating the application of MALF to the segmentation of subcortical structures – the bilateral caudate, putamen, globus pallidus, thalamus, amygdala, and hippocampus. Those 12 structures were manually delineated by two anatomists using MIPAV (Medical Image Processing, Analysis, and Visualization) in the T1-weighted images of the 30 atlas subjects. We perform leave-one-out (LOO) analysis on the 30 atlas subjects; one subject is treated as the to-be-processed subject and the other 29 serve as the atlas set to segment the excluded one.

The automated subcortical segmentations from MALF are compared with the manual segmentations using two comparison metrics – the DSC and the correlation coefficient between the volumetric measurement of the automated segmentation and that of the manual segmentation. The segmentation accuracy of the hierarchical segmentation pipeline (shown in Figure 2.3) is compared with that of the segmentation tools included in Freesurfer [5] and FSL [3]. These two algorithms are chosen for comparison because they both provide state-of-the-art segmentation accuracy for
subcortical structures compared with other segmentation algorithms.

The mean and the standard deviations of the DSCs, for all 12 structures produced by the three methods, are listed in Figure 2.7. Two-sample t-tests reveal that the 12 structure segmentations obtained from MALF are statistically superior to those obtained from either Freesurfer or FSL ($p < 1e^{-5}$). Comparing Freesurfer and FSL, FSL is better than Freesurfer in segmenting the right caudate, the bilateral putamen, the bilateral globus pallidus, the right thalamus, the left amygdala, and the right hippocampus ($p < 5e^{-3}$), in terms of overlap analysis. The Pearson’s linear correlation coefficients, between the size of the manual segmentations and that of the corresponding automated segmentations, are listed in Table 2.2. A strong linear association between the automated volume obtained from MALF and the manual volume can be observed for a majority of the structures of interest, especially for the caudate, the putamen, the thalamus, and the hippocampus. In Figure 2.8, we illustrate two examples of the basal ganglia and thalamus segmentations, while Figure 2.9 gives those of the amygdala and the hippocampus, from manual delineations as well as the three automated methods. The structure definitions are superimposed on the corresponding T1-weighted images.
Figure 2.7: The mean dice score coefficients (DSCs) and the standard deviations of the 12 structures of interest (left and right caudate, globus pallidus, putamen, thalamus, amygdala, as well as hippocampus) obtained from MALF (blue), Freesurfer (red), and FSL (yellow) relative to the manual delineations.

Table 2.2: Pearson’s linear correlation coefficients between the manual and three automated structure volume measurements. The automated structures are obtained respectively from MALF, Freesurfer, and FSL. Abbreviations: lcaud – left caudate; lputa – left putamen; lpall – left globus pallidus; ltha – left thalamus; lamyg – left amygdala; lhipp – left hippocampus; rcaud – right caudate; rputa – right putamen; rpall – right globus pallidus; rtha – right thalamus; ramyg – right amygdala; rhipp – right hippocampus

<table>
<thead>
<tr>
<th></th>
<th>lcaud</th>
<th>lputa</th>
<th>lpall</th>
<th>ltha</th>
<th>lamyg</th>
<th>lhipp</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALF</td>
<td>0.959</td>
<td>0.896</td>
<td>0.706</td>
<td>0.985</td>
<td>0.483</td>
<td>0.886</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.94</td>
<td>0.882</td>
<td>0.665</td>
<td>0.92</td>
<td>0.296</td>
<td>0.832</td>
</tr>
<tr>
<td>FSL</td>
<td>0.857</td>
<td>0.839</td>
<td>0.557</td>
<td>0.926</td>
<td>0.418</td>
<td>0.773</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>rcaud</th>
<th>rputa</th>
<th>rpall</th>
<th>rtha</th>
<th>ramyg</th>
<th>rhipp</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALF</td>
<td>0.963</td>
<td>0.912</td>
<td>0.736</td>
<td>0.983</td>
<td>0.223</td>
<td>0.911</td>
</tr>
<tr>
<td>Freesurfer</td>
<td>0.957</td>
<td>0.881</td>
<td>0.725</td>
<td>0.903</td>
<td>0.521</td>
<td>0.794</td>
</tr>
<tr>
<td>FSL</td>
<td>0.849</td>
<td>0.807</td>
<td>0.638</td>
<td>0.946</td>
<td>0.212</td>
<td>0.612</td>
</tr>
</tbody>
</table>
Figure 2.8: The top and bottom panel respectively shows the segmentation results of the six basal ganglia structures (left and right caudate, putamen, and globus pallidus) and the bilateral thalamus in two subjects. From left to right, segmentations are obtained from manual delineation, MALF, Freesurfer, and FSL.
Figure 2.9: Panels A), B) respectively demonstrate the segmentation results of the amygdala and the hippocampus in two representative subjects. From left to right, segmentations are obtained from manual delineation, MALF, Freesurfer, and FSL.
2.4.2.2 Elderly healthy control/dementia/PPA Dataset (3T MPRAGE)

In constructing the second dataset for the evaluation of the segmentation accuracy of MALF, MPRAGE T1-weighted images (TR/TE = 8.4/3.9 ms) were acquired from 35 subjects from three groups (Group 1: 14 normal aging subjects; Group 2: 15 subjects with dementia of Alzheimer type; Group 3: 6 with PPA) using 3T whole-body MRI scanners (Philips Medical Systems, Best, The Netherlands), with an axial orientation and an image matrix of 256 $\times$ 256. Participants were scanned with two slightly different protocols: one used a field of view (FOV) of 230 $\times$ 230 mm and 120 slices of 1mm thickness; and the other used an FOV of 240 $\times$ 240 mm and 140 slices of 1.2 mm thickness. These images were then manually segmented into sixteen structures – left and right hippocampus, amygdala, caudate, putamen, pallidum, lateral ventricle, thalamus, the 3rd ventricle, and the 4th ventricle.

LOO cross-validation is employed again on the datasets of Group 1 and Group 2. For Group 3, we use datasets from Group 1 and Group 2 as the atlases for segmentation. Manual segmentations are regarded as the gold standard. The segmentation accuracy is measured through the use of the DSC.

For this 3-group dataset, we first examine the level of accuracy we can achieve using a single-atlas LDDMM approach and the degree of improvement by incorporating the multi-atlas approach. In addition, we compare the segmentation accuracy of MALF with that of two label-fusion based segmentation methods.
Single-atlas VS. Multi-atlas

It is clear that having multiple atlases increases the computational complexity. We would like to be able to quantify the advantages of supporting multiple atlas anatomies in the solution. For this, we perform multiple experiments. The first experiment examines the variability in the segmentation performance of single-atlas LDDMM using a LOO technique in which a subject is chosen as the single atlas and all the other subjects in the group are segmented via the LDDMM-image mapping [33] procedure. For this purpose, the data in Group 1 and Group 2 are combined; one of the 29 subjects is used as the atlas and the other 28 images are segmented. This process is repeated for 29 different atlases, implying that each subject is segmented 28 times using 28 different atlases. For segmenting subjects in Group 3, the subjects from Group 1 and Group 2 are used as the atlases to avoid the potential bias of the LOO approach. The mean and standard deviations of the DSCs for the automated segmentations of various structures from single-atlas LDDMM are shown in Figure 2.10.

As demonstrated in Figure 2.10, the single-atlas approach performs relatively poorly in segmenting several of the structures for all 3 groups, especially the amygdala and the hippocampus. These two structures are adjacent to the inferior horn of the lateral ventricles, which tend to have poor segmentation results due to a large topological variability and resultant mapping inaccuracy in these areas.

Figure 2.11 shows the results for six representative single atlases segmenting six-
Figure 2.10: A comparison of segmentation accuracy between single-atlas and multi-atlas segmentations. Panels a), b), c) show the mean and the standard deviations of the DSCs for the sixteen structures obtained from single-atlas LDDMM (red) and MALF (green) for the three different groups.
teen different structures in one subject. This figure suggests that the best atlas varies depending on the structure; there is no single atlas that outperforms all the others in segmenting all sixteen structures. For example, for the segmentation of the right putamen and the thalamus in both hemispheres, atlas2 outperforms the other atlases, whereas, for the third ventricle, atlas2 yields the lowest segmentation accuracy in terms of the DSC value.

Figure 2.11: Depiction of the variability within single atlases. This figure shows the scatter plots of DSC values for 16 structures in one subject, between the automated segmentations from 6 different atlases via single-atlas LDDMM and the corresponding manual segmentations.

Figure 2.12 shows the segmentation results of two subjects for a comparison between the single-atlas and the multi-atlas approach. The DSC values resulting from MALF are also demonstrated in Figure 2.10 for a direct comparison with those from the averaged single-atlas LDDMMs. Because of the possibility that the LOO analysis, when using data with an identical image protocol (data from Group 1 and Group 2), may not represent the real-world performance of the proposed approach, the method
is applied to the Group 3 data, which were acquired with a different scanner and imaging parameters. The MRIs from Groups 1 and 2 are taken as the atlases. The DSC values for the segmentations of Group 3 using the single-atlas and multi-atlas approach are also illustrated in Figure 3, demonstrating a comparable level of DSC from MALF as those obtained in Groups 1 and 2.

**Label-fusion VS. MALF**

The generative probability model on which MALF is based averages likelihoods, generating a single maximizing label for each voxel in the image domain. It is natural to compare MALF with competitive methods that average segmentation labels via label-fusion techniques. To do this, we compare MALF with two representative label-fusion techniques, STAPLE [14] and Spatial STAPLE [93]. One might expect that while label-fusion should be more robust when applied to images for which the generative model is not accurate, likelihood fusion should provide benefits in circumstances when the generative model is valid. Tables 2.3 - 2.5 tabulate the mean values and standard deviations of the dice overlaps for the three methods computed across subjects in the three groups (Group 1, Group 2, and Group 3). The performance of Spatial STAPLE and MALF are statistically identical for a majority of structures segmented in the control group (Table 2.3), providing superior performance relative to STAPLE. MALF outperforms Spatial STAPLE in segmenting the third and the fourth ventricle in control subjects (Table 2.3). For the brain imagery from disease populations, significant improvement by MALF over Spatial STAPLE is observed for
Figure 2.12: Examples of subcortical segmentations from single- and multi-atlas approaches. Panel A shows the automated segmentation results of two subjects using the single-atlas, while panel B shows the segmentation results for the same subjects using the multi-atlas approach.
9 structures in the AD (Table 2.4) and 3 structures in the PPA populations (Table 2.5). One of the most notable improvements is found in the area around the inferior and posterior horns of the lateral ventricles, where the ventricle anatomy has a substantial amount of anatomical variability (Figure 2.13). This benefit must arise from the fact that, even though these anatomies are diseased, we are able to adequately model the generative probability and therefore the atlas selector function effectively averages the proper likelihoods to fit the anatomies.

2.4.2.3 Young healthy adult Dataset (3T MPRAGE)

For the third dataset, we collected high-resolution structural MR images from 8 neurologically normal adults using a 3 Tesla scanner. All 8 scans are of young adults (256 × 256 (1 × 1 mm) in-plane resolution, 120 1-mm slices without gaps; mean age, 23 years; age range, 19-26 years). Manual delineations of fourteen subcortical and ventricular structures (the left and right caudate, putamen, globus pallidus, thalamus, hippocampus, amygdala, as well as lateral ventricle) were performed by two neuroanatomists with a high inter-rater segmentation accuracy.

The accuracy of MALF in segmenting this dataset is again compared with that of the segmentation module in FSL [3] and Freesurfer [89]. The results are given in Figure 2.14, illustrating the comparison of the three methods in terms of the dice score coefficient.
Table 2.3: The average DSC values between the manual and automated segmentations for each structure of the fourteen for Group 1 subjects, thus comparing STAPLE, Spatial STAPLE, and MALF. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significant in being greater than that of other methods ($p < 0.05$).

<table>
<thead>
<tr>
<th>Structure</th>
<th>STAPLE</th>
<th>Spatial STAPLE</th>
<th>MALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. putamen</td>
<td>0.878 (0.0250)</td>
<td>0.908 (0.0154)</td>
<td>0.908 (0.0148)</td>
</tr>
<tr>
<td>L. putamen</td>
<td>0.857 (0.0362)</td>
<td>0.891 (0.0350)</td>
<td>0.892 (0.0329)</td>
</tr>
<tr>
<td>R. caudate</td>
<td>0.836 (0.0409)</td>
<td>0.867 (0.0355)</td>
<td>0.865 (0.0446)</td>
</tr>
<tr>
<td>L. caudate</td>
<td>0.812 (0.0506)</td>
<td>0.852 (0.0365)</td>
<td>0.851 (0.0465)</td>
</tr>
<tr>
<td>R. pallidum</td>
<td>0.776 (0.0357)</td>
<td>0.840 (0.0291)</td>
<td>0.836 (0.0378)</td>
</tr>
<tr>
<td>L. pallidum</td>
<td>0.730 (0.0543)</td>
<td>0.817 (0.0501)</td>
<td>0.822 (0.0463)</td>
</tr>
<tr>
<td>R. thalamus</td>
<td>0.907 (0.0224)</td>
<td>0.911 (0.0183)</td>
<td>0.911 (0.0185)</td>
</tr>
<tr>
<td>L. thalamus</td>
<td>0.883 (0.0370)</td>
<td>0.906 (0.0246)</td>
<td>0.898 (0.0199)</td>
</tr>
<tr>
<td>R. amygdala</td>
<td>0.786 (0.0401)</td>
<td>0.862 (0.0245)</td>
<td>0.861 (0.0188)</td>
</tr>
<tr>
<td>L. amygdala</td>
<td>0.767 (0.0527)</td>
<td>0.822 (0.0393)</td>
<td>0.826 (0.0388)</td>
</tr>
<tr>
<td>R. hippocampus</td>
<td>0.769 (0.0461)</td>
<td>0.838 (0.0337)</td>
<td>0.839 (0.0297)</td>
</tr>
<tr>
<td>L. hippocampus</td>
<td>0.795 (0.0499)</td>
<td>0.846 (0.0278)</td>
<td>0.840 (0.0227)</td>
</tr>
<tr>
<td>R. ventricle</td>
<td>0.874 (0.0465)</td>
<td>0.917 (0.0267)</td>
<td>0.934 (0.0247)</td>
</tr>
<tr>
<td>L. ventricle</td>
<td>0.880 (0.0503)</td>
<td>0.918 (0.0336)</td>
<td>0.937 (0.0259)</td>
</tr>
<tr>
<td>Third ventricle</td>
<td>0.548 (0.1712)</td>
<td>0.776 (0.1214)</td>
<td>0.824 (0.0817)</td>
</tr>
<tr>
<td>Fourth ventricle</td>
<td>0.671 (0.1287)</td>
<td>0.799 (0.0931)</td>
<td>0.854 (0.0776)</td>
</tr>
</tbody>
</table>
Table 2.4: Mean and standard deviations of DSCs computed across the fifteen subjects in Group 2 for each structure for comparisons of STAPLE, Spatial STAPLE, and MALF. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significantly higher than that of other methods ($p < 0.05$).

<table>
<thead>
<tr>
<th>Structure</th>
<th>STAPLE</th>
<th>Spatial STAPLE</th>
<th>MALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. putamen</td>
<td>0.638 (0.0504)</td>
<td>0.878 (0.0351)</td>
<td>0.891 (0.0238)</td>
</tr>
<tr>
<td>L. putamen</td>
<td>0.613 (0.0661)</td>
<td>0.848 (0.0525)</td>
<td>0.854 (0.0444)</td>
</tr>
<tr>
<td>R. caudate</td>
<td>0.776 (0.0558)</td>
<td>0.856 (0.0346)</td>
<td>0.879 (0.0260)</td>
</tr>
<tr>
<td>L. caudate</td>
<td>0.746 (0.0696)</td>
<td>0.828 (0.0454)</td>
<td>0.850 (0.0359)</td>
</tr>
<tr>
<td>R. pallidum</td>
<td>0.784 (0.0547)</td>
<td>0.831 (0.0389)</td>
<td>0.834 (0.0392)</td>
</tr>
<tr>
<td>L. pallidum</td>
<td>0.746 (0.0475)</td>
<td>0.792 (0.0522)</td>
<td>0.791 (0.0668)</td>
</tr>
<tr>
<td>R. thalamus</td>
<td>0.681 (0.0355)</td>
<td>0.900 (0.0313)</td>
<td>0.891 (0.0270)</td>
</tr>
<tr>
<td>L. thalamus</td>
<td>0.655 (0.0541)</td>
<td>0.876 (0.0381)</td>
<td>0.872 (0.0275)</td>
</tr>
<tr>
<td>R. amygdala</td>
<td>0.703 (0.0974)</td>
<td>0.796 (0.0948)</td>
<td>0.803 (0.0816)</td>
</tr>
<tr>
<td>L. amygdala</td>
<td>0.647 (0.0708)</td>
<td>0.785 (0.0647)</td>
<td>0.806 (0.0592)</td>
</tr>
<tr>
<td>R. hippocampus</td>
<td>0.670 (0.0988)</td>
<td>0.783 (0.0759)</td>
<td>0.821 (0.0362)</td>
</tr>
<tr>
<td>L. hippocampus</td>
<td>0.703 (0.0859)</td>
<td>0.799 (0.0556)</td>
<td>0.818 (0.0363)</td>
</tr>
<tr>
<td>R. ventricle</td>
<td>0.904 (0.0481)</td>
<td>0.917 (0.0326)</td>
<td>0.929 (0.0299)</td>
</tr>
<tr>
<td>L. ventricle</td>
<td>0.908 (0.0466)</td>
<td>0.921 (0.0333)</td>
<td>0.932 (0.0278)</td>
</tr>
<tr>
<td>Third ventricle</td>
<td>0.678 (0.1244)</td>
<td>0.830 (0.0426)</td>
<td>0.862 (0.0668)</td>
</tr>
<tr>
<td>Fourth ventricle</td>
<td>0.706 (0.1245)</td>
<td>0.793 (0.0862)</td>
<td>0.839 (0.0741)</td>
</tr>
</tbody>
</table>
Table 2.5: Mean and standard deviations of the DSC values obtained respectively from STAPLE, Spatial STAPLE, and MALF segmentations of the 16 structures of the six Group 3 subjects. Bold typesetting indicates that the dice overlap obtained from the corresponding is statistically significantly higher than that of other methods ($p < 0.05$).

<table>
<thead>
<tr>
<th></th>
<th>STAPLE</th>
<th>Spatial STAPLE</th>
<th>MALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. putamen</td>
<td>0.866 (0.0297)</td>
<td>0.884 (0.0166)</td>
<td>0.878 (0.0260)</td>
</tr>
<tr>
<td>L. putamen</td>
<td>0.856 (0.0392)</td>
<td>0.878 (0.0413)</td>
<td>0.872 (0.0499)</td>
</tr>
<tr>
<td>R. caudate</td>
<td>0.861 (0.0327)</td>
<td>0.854 (0.0253)</td>
<td>0.859 (0.0258)</td>
</tr>
<tr>
<td>L. caudate</td>
<td>0.832 (0.0432)</td>
<td>0.836 (0.0218)</td>
<td>0.847 (0.0322)</td>
</tr>
<tr>
<td>R. pallidum</td>
<td>0.788 (0.0393)</td>
<td>0.817 (0.0261)</td>
<td>0.810 (0.0212)</td>
</tr>
<tr>
<td>L. pallidum</td>
<td>0.763 (0.0520)</td>
<td>0.777 (0.0686)</td>
<td>0.784 (0.0847)</td>
</tr>
<tr>
<td>R. thalamus</td>
<td>0.871 (0.0298)</td>
<td>0.854 (0.0494)</td>
<td>0.867 (0.0482)</td>
</tr>
<tr>
<td>L. thalamus</td>
<td>0.849 (0.0415)</td>
<td>0.828 (0.0546)</td>
<td>0.843 (0.0563)</td>
</tr>
<tr>
<td>R. amygdala</td>
<td>0.707 (0.0568)</td>
<td>0.769 (0.0668)</td>
<td>0.769 (0.0741)</td>
</tr>
<tr>
<td>L. amygdala</td>
<td>0.662 (0.0752)</td>
<td>0.745 (0.0561)</td>
<td>0.770 (0.0580)</td>
</tr>
<tr>
<td>R. hippocampus</td>
<td>0.777 (0.0987)</td>
<td>0.796 (0.0385)</td>
<td>0.825 (0.0432)</td>
</tr>
<tr>
<td>L. hippocampus</td>
<td>0.805 (0.0251)</td>
<td>0.839 (0.0236)</td>
<td>0.858 (0.0174)</td>
</tr>
<tr>
<td>R. ventricle</td>
<td>0.927 (0.0144)</td>
<td>0.924 (0.0192)</td>
<td>0.923 (0.0229)</td>
</tr>
<tr>
<td>L. ventricle</td>
<td>0.927 (0.0246)</td>
<td>0.926 (0.0201)</td>
<td>0.927 (0.0255)</td>
</tr>
<tr>
<td>Third ventricle</td>
<td>0.749 (0.1034)</td>
<td>0.803 (0.0327)</td>
<td>0.808 (0.0511)</td>
</tr>
<tr>
<td>Fourth ventricle</td>
<td>0.738 (0.0487)</td>
<td>0.811 (0.0292)</td>
<td>0.836 (0.0428)</td>
</tr>
</tbody>
</table>
Figure 2.13: A comparison of MALF, STAPLE, and Spatial STAPLE, demonstrated via three representative 2-D slices of three structures near medial temporal regions – the amygdala, the hippocampus, and the ventricle in both hemispheres obtained respectively from manual delineation (top row), MALF (2nd row), STAPLE (3rd row), and Spatial STAPLE (bottom row).
Figure 2.14: Bar plots showing the mean DSC values, with the standard deviations plotted as the error bars, for 12 segmented subcortical structures and lateral ventricles of 8 young healthy adults, from MALF (blue), FreeSurfer (red), and FSL (green).

Keys: left-amyg – left amygdala; right-amyg – right amygdala; left-caud – left caudate; right-caud – right caudate; left-pall – left globus pallidus; right-pall – right globus pallidus; left-hipp – left hippocampus; right-hipp – right hippocampus; left-puta – left putamen; right-puta – right putamen; left-thal – left thalamus; right-thal – right thalamus; left-vent – left lateral ventricle; right-vent – right lateral ventricle.
2.4.2.4 Elderly normal/preclinical AD Dataset (1.5T SPGR)

The last dataset we use for the segmentation validation is a subset of scans from the aforementioned BIOCARD study, including a total of 16 scans (mean age: 52 years old; 8 females and 8 males). For this dataset, 10 structures (left and right putamen, globus pallidus, amygdala, hippocampus, and lateral ventricle) have been manually delineated for the purpose of validation. LOO MALF segmentation is employed to segment each scan; to segment one scan, the other 15 scans play the role of the multi-atlas set. The accuracy of MALF in segmenting this dataset is first compared with FSL and Freesurfer for each of the structures of interest. Quantitative comparisons of MALF, FSL, and Freesurfer, in terms of dice overlaps for each structure, are demonstrated in Figure 2.15. Paired Student’s t-test reveals that MALF is statistically significant better than Freesurfer in segmenting the bilateral ventricles (left: $p = 0.036$, right: $p = 0.042$). Furthermore, MALF is statistically superior to both Freesurfer and FSL in segmenting the other 8 structures except the right globus pallidus (left amygdala: $p < 0.005$, right amygdala: $p < 0.002$, left hippocampus: $p < 2.3e^{-6}$, right hippocampus: $p < 6.8e^{-8}$, left putamen: $p < 1.5e^{-8}$, right putamen: $p < 3e^{-8}$, left globus pallidus: $p < 0.05$, right globus pallidus: $p > 0.05$). Qualitative comparisons of the three methods in segmenting the putamen and the globus pallidus are illustrated in Figure 2.16, and in Figure 2.17 for the comparison of the amygdala and the hippocampus.

In addition to the comparisons with Freesurfer and FSL, we explore the differ-
Figure 2.15: Bar plots showing the mean dice overlaps, with the standard deviations plotted as the error bars, for a comparison of MALF (blue), Freesurfer (red), and FSL (yellow) in segmenting the 10 structures of interest in the 16 BIOCARD scans.

Figure 2.16: Demonstration of the bilateral putamen and globus pallidus segmentations for representative subjects. From left to right: manual, MALF, Freesurfer, and FSL.
Figure 2.17: Visual comparisons of the segmentations of the hippocampus and the amygdala in two representative subjects. The segmentations come from manual delineation, MALF, Freesurfer, and FSL.

ences between MALF and label-fusion techniques (STAPLE and Spatial STAPLE) in segmenting brain structures of this dataset. The main motivation is that we would like to compare those three methods on images with relatively poor image quality (acquired from 1.5T scans of subjects that are aging or with dementia). The mean and the standard deviations of the DSC values of automated segmentations for the 10 structures, obtained from MALF, STAPLE, Spatial STAPLE, are listed in Table 2.6. Student’s t-tests indicate that, for this dataset, MALF is statistically significantly better than either STAPLE or Spatial STAPLE in segmenting all structures of interest (left putamen: $p < 0.001$, right putamen: $p < 3.5e^{-6}$, left globus pallidus: $p < 5.8e^{-5}$, right globus pallidus: $p < 3.7e^{-5}$, left amygdala: $p < 0.005$, right amygdala: $p < 0.005$, left hippocampus: $p < 0.005$, right hippocampus: $p < 0.001$, right
Table 2.6: Mean and standard deviations of dice overlaps obtained from MALF, Spatial STAPLE, and STAPLE in segmenting the 10 structures of the 16 scans from BIOCARD. Bold typesetting indicates that the dice overlap obtained from the corresponding method is statistically significant higher than that of other methods ($p < 0.05$).

<table>
<thead>
<tr>
<th></th>
<th>MALF</th>
<th>Spatial STAPLE</th>
<th>STAPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>left putamen</td>
<td>0.878 ± 0.026</td>
<td>0.861 ± 0.036</td>
<td>0.817 ± 0.047</td>
</tr>
<tr>
<td>right putamen</td>
<td>0.875 ± 0.032</td>
<td>0.851 ± 0.027</td>
<td>0.801 ± 0.042</td>
</tr>
<tr>
<td>left globus pallidus</td>
<td>0.796 ± 0.073</td>
<td>0.751 ± 0.085</td>
<td>0.701 ± 0.086</td>
</tr>
<tr>
<td>right globus pallidus</td>
<td>0.784 ± 0.063</td>
<td>0.743 ± 0.062</td>
<td>0.687 ± 0.072</td>
</tr>
<tr>
<td>left amygdala</td>
<td>0.838 ± 0.038</td>
<td>0.822 ± 0.043</td>
<td>0.827 ± 0.041</td>
</tr>
<tr>
<td>right amygdala</td>
<td>0.843 ± 0.049</td>
<td>0.829 ± 0.041</td>
<td>0.835 ± 0.031</td>
</tr>
<tr>
<td>left hippocampus</td>
<td>0.859 ± 0.031</td>
<td>0.823 ± 0.051</td>
<td>0.782 ± 0.074</td>
</tr>
<tr>
<td>right hippocampus</td>
<td>0.856 ± 0.021</td>
<td>0.831 ± 0.033</td>
<td>0.789 ± 0.056</td>
</tr>
<tr>
<td>left ventricle</td>
<td>0.912 ± 0.024</td>
<td>0.891 ± 0.023</td>
<td>0.867 ± 0.052</td>
</tr>
<tr>
<td>right ventricle</td>
<td>0.924 ± 0.018</td>
<td>0.874 ± 0.029</td>
<td>0.856 ± 0.049</td>
</tr>
</tbody>
</table>

lateral ventricle: $p < 0.02$) except the segmentation of the left lateral ventricle for which MALF and Spatial STAPLE are statistical equivalent ($p = 0.141$) and both superior to STAPLE ($p < 0.05$).

2.5 Conclusion

In this chapter, we have proposed an automated approach for the segmentation of T1-weighted images in the framework of the multi-atlas random diffeomorphic orbit model. Detailed derivations of the underlying mathematical theorems and motivations were presented at the beginning of the chapter. The method described here
fuses the anatomical information of multiple atlases via convex combination of the atlas-specific likelihoods, with the weights in the convex combination given by the conditional-mean formula. We therefore name the algorithm multi-atlas likelihood fusion (MALF) and with this framework established, we proceeded to build a two-level hierarchical brain segmentation pipeline based on MALF. The first level of the pipeline utilizes a fast version of MALF, aiming to remove the skull from the brain region in T1-weighted images. A regular version of MALF is used in the second level of the pipeline for the purpose of segmenting internal brain structures, such as the subcortical structures, as well as ventricles. We presented experimental results and validated the accuracy of the hierarchical segmentation pipeline in terms of both skull-stripping and segmenting deep gray matter structures.

The validation analysis was performed on T1-weighted images of subjects with a wide range of variability in cognitive states (subjects that are cognitively normal as well as subjects with a variety of diseases – Autism, ADHD, dementia, and PPA), age range (subjects that are pediatric, young adult, and elderly adults), as well as imaging parameter (3.0T MPRAGE and 1.5T SPGR). The accuracy of the proposed method was compared with several other state-of-the-art segmentation algorithms, in terms of overlap accuracy with respect to the manual segmentations and correlation between the automated volume size and the volume size of the manual segmentations. Highly reliable and accurate segmentation results have been demonstrated.

With the reliability of the MALF algorithm established, we conclude that it is
applicable to clinical studies as well as a subsequent morphological analysis on segmented structures of interest. In the following few chapters we will turn our attention to what can be achieved through this application. Chapter 3 will focus on MALF in its current form and applications to pediatric subjects with various disorders, while in Chapter 4 we look to the possibility of extending MALF to Diffusion Tensor Imaging and whole brain segmentations for subjects representing a range of anatomical phenotypes.
Chapter 3

Volumetric Analysis of MALF in Case-control Clinical Studies

3.1 Introduction

The main purpose of segmenting brain structures in individual images, via either manual delineation or automated algorithms, is to characterize the neuroanatomical abnormalities in a certain brain structure that are associated with a specific neuropsychiatric disorder such as the ones mentioned in previous chapters—Alzheimer’s disease, PPA, Autism, ADHD, and so on. The power of an automated segmentation method can be revealed by its applicability to detecting abnormalities in brain structures in disease populations. Our interest is to investigate the applicability of the aforementioned segmentation pipeline, built on MALF, to the detection of structure-
specific anatomical abnormalities in various disorders.

In this chapter, we apply the hierarchical segmentation pipeline described in section 2.3.1 to T1-weighted images from two case-control clinical studies on Autism Spectrum Disorders (ASD) and ADHD. We are particularly interested in the basal-ganglia circuitry and the amygdala-hippocampus circuitry in these two populations. To be specific, we focus on the caudate, the putamen, the globus pallidus, the amygdala, and the hippocampus, in both hemispheres. Statistical analyses are performed, with respect to the aforementioned ten structures, for the comparison of the age-matched typically developing (TD) population and the ADHD population, as well as the age-matched TD population and the Autism population.

3.2 Motivation

3.2.1 ADHD

3.2.1.1 Basal-ganglia circuitry

The basal ganglia have long been suspected as contributing to the pathophysiology of ADHD. A number of MRI studies have examined basal ganglia structures in individuals with ADHD, however the findings have been highly inconsistent, especially regarding sex-specific abnormalities. Some studies suggest reduced left caudate in ADHD girls but no change in boys [107,108], whereas some indicate reduced left
caudate in ADHD boys but not in girls [71]. For the putamen, some studies suggest no difference [107, 109] while some suggest reduction [71] in ADHD boys. There has also been marked inconsistency in globus pallidus findings, including: reduced left with no difference in the right in ADHD boys [109]; reduced right with no change in the left in ADHD boys, no change in girls [107, 108]; reduced left and right in ADHD boys but not in girls [71].

3.2.1.2 Amygdala-hippocampus circuitry

ADHD is traditionally defined by symptoms of inattention and hyperactivity or impulsivity. Yet a significant body of research has shown emotional processing and regulation deficits in individuals with ADHD such as emotional face processing deficits, emotional impulsiveness, poor frustration tolerance, and increased levels of depression compared to TD controls [110–113]. Neuroimaging research has begun to examine differences within limbic structures including the hippocampus and amygdala in children with ADHD. While some studies suggest decreased amygdala and increased hippocampus volumes in children with ADHD compared to TD controls, others suggest no group differences [72, 107, 114].
3.2.2 Autism

3.2.2.1 Basal-ganglia circuitry

It has been suggested that abnormalities in basal ganglia structure and function may contribute to the pathophysiology of ASD. A number of MRI studies have examined basal ganglia volume in subjects with ASD, with inconsistent results. For example, one study suggests that children with ASD have enlarged caudates [115], while others have found no differences in basal ganglia volumes [116]. ADHD is one of the most common comorbid conditions in ASD with reported co-occurrence ranging from 40% to 78% [117]. Basal ganglia dysfunction has also been proposed in neurologic models of ADHD and several studies have found smaller basal ganglia volumes in children with ADHD [71,118,119]. In light of these studies, it remains unclear how the presence of ADHD affects basal ganglia development in children with ASD.

3.2.2.2 Amygdala-hippocampus circuitry

The amygdala and hippocampus are two medial temporal lobe regions implicated in emotion processing and psychopathology, and are hypothesized to play a role in the neuropathology of ASD. A number of studies have examined amygdala and hippocampal structure in youth with ASD but the findings have been markedly confounding. For example, some data indicate enlarged amygdala volumes in children with ASD compared to TD control children [120], whereas other data indicate that the changes
in amygdala volume across development are similar to those in control children [121]. Hippocampal data are also inconsistent and show enlarged [122], or reduced [123] volumes in children with ASD compared to control subjects.

Clearly, in both ADHD and Autism, the findings regarding volumetric abnormalities of the ten structures of interest are highly conflicting. These discrepant results may be attributed to small sample sizes, clinical heterogeneity amongst the sample, as well as limitations in the imaging methodologies used to examine the structure volumes. To address these limitations, we examine volumetric differences in the basal ganglia, the amygdala, and the hippocampus in children with ADHD as well as children with Autism, compared to TD control children using a large MRI dataset. Given the growing evidence for significant gender differences in the neurobiological basis of ADHD and Autism, we also examine the impact of sex on these findings.

As mentioned before, ADHD often co-occurs with Autism, we therefore also investigate the impact of co-occurring ADHD on the basal ganglia volumes in children with high functioning ASD. In examining the volumetric abnormalities of the amygdala and the hippocampus in the ADHD population, we also look at the relationship between amygdala and hippocampus volume measurements and measures of clinical symptoms (mood, anxiety symptoms) within the ADHD group.
3.3 Method

A total of 345 pediatric subjects were involved in the study, including: 98 children with ADHD (no comorbidities), ages 8-12 years (mean age=10.12 years, 66 males); 74 children with well-characterized high functioning ASD (mean age=10.5 years, 62 males; 43 with comorbid ASD+ADHD (36 males)); and 173 typically developing control children (mean age=10.37 years, 121 males). The above three groups are well-matched in terms of age, gender, socioeconomic status (SES), and perceptual reasoning index (PRI) from the WISC-IV. For each subject, high-resolution T1-weighted 3D-volume MPRAGE image (matrix size=256 × 256, echo time=3.76 ms, repetition time=7.99 ms, field of view=256 × 256 mm, slice thickness=1.0 mm) was acquired from a 3T Philips Gyroscan NT scanner (Royal Philips Electronics, Amsterdam, The Netherlands).

For each subject, the segmentations of the bilateral caudate, putamen, globus pallidus, amygdala, and hippocampus are automatically extracted using the aforementioned hierarchial segmentation pipeline. Prior to group comparisons, structure volumes are adjusted by the total brain size, computed as the number of voxels inside the brain with skull removed, scaled by image resolution. For case-control group comparisons, two-sample t-tests are performed to evaluate the group difference and the statistical significance of that difference. For ADHD versus TD control as well as ASD versus TD control, comparisons are performed in terms of the volume measurements of each structure, collapsed across sex and within-sex. In the examination
of the basal ganglia in the Autism group, a multivariate general linear model (GLM) is used to test for an overall effect of diagnosis on the basal ganglia volumes followed by a Fisher’s least significant difference (LSD) post-hoc examination of each of the diagnostic categories (ASD only, ASD+ADHD, TD) to examine for an effect of ADHD co-morbidity. To investigate the relationship between regional brain volumes and measures of clinical symptoms, Pearson product-moment correlations are used.

3.4 Results

3.4.1 ADHD

3.4.1.1 Basal-ganglia circuitry

Figures 3.1-3.3 show the scatter plots of the volumetric measurements of the basal ganglia structures as well as the group means. Two-sample t-tests indicate significant bilateral reduction in putamen volumes (left: \( p = 0.008 \); right: \( p = 0.015 \)) and globus pallidus volumes (left: \( p = 0.001 \); right: \( p = 0.003 \)) in ADHD. No significant group difference was detected in the caudate (\( p > 0.19 \)). Separate sex-group analyses revealed no basal ganglia volume differences in ADHD girls. In contrast, reduced bilateral putamen and globus pallidus volumes were observed in ADHD boys (left putamen: \( p = 0.023 \); right putamen: \( p = 0.049 \); left globus pallidus: \( p = 0.003 \); right globus pallidus: \( p = 0.004 \)).
Figure 3.1: Representation of the volume measurements of the left and right caudate in subjects from the typically developing group (blue diamond) and the ADHD group (red circle). Bar means the averaged value in each group.
Figure 3.2: Figure showing the volume measurements of the putamen, in both hemispheres, from the TD group (blue diamond) and the ADHD group (red circle). Results from the entire group, the male only group, and the female only group are tabulated, with bars representing the mean values in each group.
Figure 3.3: Demonstration of globus pallidus volumes in both sides for TD (blue diamond) and ADHD (red circle) populations. Measurements of the entire group, the male group, as well as the female group, are listed with mean values plotted as bars.
3.4.1.2 Amygdala-hippocampus circuitry

The mean values and the standard deviations of amygdala and hippocampus volumes are tabulated in Table 3.1. ADHD and TD participants did not differ in left amygdala, right amygdala, left hippocampus nor right hippocampus volume. Similarly, examination of male participants showed no volumetric differences in any region. In contrast, females with ADHD demonstrated smaller right hippocampus volumes compared to TD females \( p = 0.05 \). No differences were observed within females for the left hippocampus, left amygdala or right amygdala volume.

Examination of brain-behavior correlations for all ADHD participants showed a significant negative relationship between right amygdala size and the Child Behavior Checklist total Affective Problems subscale \( (r = -0.22, p = 0.049) \). For males with ADHD, there was a significant positive relationship between total Screen for Child Anxiety Related Disorders score (anxiety) and left amygdala volume \( (r = 0.27, p = 0.047) \). No brain-behavior correlations were shown for ADHD females.

3.4.2 Autism

3.4.2.1 Basal-ganglia circuitry

Individual volumes as well as group means for the caudate, the putamen, and the globus pallidus (left and right separated) are displayed in Figure 3.4. The GLM revealed an overall main effect of diagnosis on basal ganglia volumes including left
Table 3.1: Mean and standard deviations of volume measurements ($mm^3$) of the left and right amygdala and hippocampus, in six groups – the entire ADHD group, the entire healthy control group, the ADHD male-only group, the healthy control male-only group, the ADHD female-only group, as well as the healthy control female-only group.

<table>
<thead>
<tr>
<th></th>
<th>Left amygdala</th>
<th>Right amygdala</th>
<th>Left hippocampus</th>
<th>Right hippocampus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD (whole population)</td>
<td>1152.38 ± 164.54</td>
<td>1049.75 ± 163.38</td>
<td>3005.55 ± 265.09</td>
<td>3200.32 ± 251.56</td>
</tr>
<tr>
<td>TD (whole population)</td>
<td>1187.54 ± 152.91</td>
<td>1071.04 ± 136.92</td>
<td>3044.46 ± 240.52</td>
<td>3224.09 ± 228.36</td>
</tr>
<tr>
<td>ADHD (male population)</td>
<td>1159.29 ± 154.03</td>
<td>1060.76 ± 153.40</td>
<td>3011.48 ± 269.00</td>
<td>3212.32 ± 260.45</td>
</tr>
<tr>
<td>TD (male population)</td>
<td>1187.69 ± 152.51</td>
<td>1070.19 ± 134.93</td>
<td>3025.31 ± 234.09</td>
<td>3201.17 ± 220.91</td>
</tr>
<tr>
<td>ADHD (female population)</td>
<td>1138.56 ± 185.54</td>
<td>1027.73 ± 182.21</td>
<td>2993.70 ± 260.80</td>
<td>3176.31 ± 234.78</td>
</tr>
<tr>
<td>TD (female population)</td>
<td>1187.20 ± 155.29</td>
<td>1073.02 ± 142.73</td>
<td>3088.91 ± 251.49</td>
<td>3277.29 ± 238.47</td>
</tr>
</tbody>
</table>

and right caudate ($p > 0.001$ and $p = 0.003$, respectively), left and right putamen ($p > 0.001$ and $p > 0.001$, respectively), and left and right globus pallidus ($p > 0.001$ and $p = 0.002$, respectively). For all structures, children with ASD showed larger volumes than TD children. Post-hoc analyses revealed that statistical comparisons of the ASD+ADHD population versus TD revealed more robust differences in the left and right caudate ($p = 0.003$ and $p = 0.010$, respectively) and left and right globus pallidus ($p = 0.002$ and $p = 0.009$, respectively) in contrast to analyses of the ASD-only population versus TD (left and right caudate $p = 0.021$ and $p = 0.063$, respectively; left and right globus pallidus $p = 0.015$ and $p = 0.033$, respectively).
Figure 3.4: Volume measurements of basal ganglia structures in subjects from the age-matched TD group (blue diamond) and the ASD group (red circle). Bar means the averaged value in each group.
3.4.2.2 Amygdala-hippocampus circuitry

Findings from the t-tests indicate that the ASD group had significantly higher mean volumes for the right [ASD: 3349.6 mm$^3$ (SD = 245.7 mm$^3$)] versus TD controls: 3224.1 mm$^3$ (SD = 228.4 mm$^3$), $p = 1.4e^{-4}$] and left [ASD: 3146.9 mm$^3$ (SD = 250.3 mm$^3$)] versus TD controls: 3044.5 mm$^3$ (SD = 240.5 mm$^3$), $p = 0.003$] hippocampi. When comparing gender differences, males had significantly higher mean volumes for the right [ASD: 3352.9 mm$^3$ (SD = 247.8 mm$^3$)] versus TD controls: 3143.6 mm$^3$ (SD = 249.5 mm$^3$), $p = 3.6e^{-5}$] and left [ASD: 3143.6 mm$^3$ (SD = 249.5 mm$^3$)] versus TD controls: 3025.3 mm$^3$ (SD = 234.1 mm$^3$), $p = 0.002$] hippocampi. No significant differences were present in the mean hippocampal volumes for girls. All group differences in mean amygdala volumes failed to reach statistical significance (for all group comparisons, $p > 0.05$).

3.5 Conclusion

According to the automated segmentation results from MALF, children with ADHD show reduced basal ganglia volumes bilaterally, especially in the putamen and the globus pallidus. These abnormalities were present in boys but not in girls with ADHD, suggesting sex-dependent effects on basal ganglia abnormalities in ADHD. As for the amygdala and the hippocampus, while our results suggest no overall group differences in amygdala or hippocampus volume between ADHD and TD youth, when
sex was taken into account, results suggested smaller right hippocampus volumes in ADHD females compared to TD females. While these results should be interpreted with caution, they provide preliminary evidence of atypical limbic system volumes in females with ADHD.

Children with high-functioning Autism show larger basal ganglia volumes when compared to TD children. This is observed diffusely with increased volumes bilaterally in the caudate, putamen, and the globus pallidus. This suggests that altered basal ganglia development may contribute to a range of dysfunctions that characterize Autism, including impaired regulation of sensory, motor, emotional and social-communicative behavior. Furthermore, differences in the caudate and globus pallidus appear more robust when ADHD comorbidity is present suggesting that it may be important to consider associated psychiatric diagnoses in examining neural contributions to the pathophysiology of Autism. For the memory-related amygdala-hippocampus circuitry, the enlarged hippocampal findings may suggest abnormal neuronal organization and connectivity within this structure and may provide insights into the learning and emotional difficulties in Autism.
Chapter 4

Whole Brain Segmentation of Diffusion Tensor Imaging via MALF

4.1 Introduction

The broad spectrum of the MR contrast mechanism makes MRI one of the most powerful imaging tools. The aforementioned T1-weighted contrast provides a reasonable format for the delineation of deep gray matter structures and ventricles. However, for certain brain structures, such as peripheral gray matter and white matter, T1-weighted contrast is not sufficient, in which case we need to utilize other types of MRI contrast. Diffusion tensor imaging (DTI) is a promising method for imaging the white
matter of the brain. DTI can generate multiple quantitative maps with markedly different qualities in terms of anatomical contrasts. The mean diffusivity (MD) contrast provides clear distinction between the tissue (generally within the range of 0.6 ∼ 0.9) and the CSF (approximately 3.0 ∼ 3.5), providing a strong constraint in defining the ventricles and the brain surface. The fractional anisotropy (FA) contrast provides sharp distinction between the gray (typically FA < 0.15 ∼ 0.25) and white matter structures (FA > 0.15 ∼ 0.25). The eigenvector (EV) can differentiate intra-white matter structures based on their characteristic orientations. It has been suggested that using a combination of multiple diffusion tensor measures (e.g. MD, FA, and EV) could maximize the specificity in characterizing microstructures in the human brain via DTI [124]. Previously, vector-to-vector or tensor-to-tensor registration algorithms have been introduced [34,125,126], which were further extended to multi-channel image registration, in which multiple contrast information, such as FA, diffusivity, and fiber orientation, was used simultaneously to drive the registration algorithm [36]. These ideas could improve the accuracy in registering a single atlas image and the subject image, but an incorporation of the multiple-contrast information into the multi-atlas likelihood fusion process has not been introduced so far.

The purpose of this chapter is to extend the single-contrast MALF algorithm proposed in Chapter 2 to multi-contrast MALF and apply it to segmenting DT images. The extension of MALF from a single contrast image, such as the T1-weighted images, to multi-contrast images (e.g. eigenvalues and eigenvectors of DTI) can be naturally
established by assuming conditional independence in computing $p(I^D|W, a)$, where $I^D$
denotes the observed vector-valued DT image, $W$ denotes a given segmentation label,
and $a$ the randomly selected atlas. We will proceed by introducing the framework for
incorporating the multi-contrast intensity information generated in DTI into MALF
and applying it to segmenting the human brain into a total of 159 structures from DT
images. In the T1 case, the distribution of the intensity in each structure is modeled
as a single Gaussian. In the DTI case, we use five intensity elements ([FA, MD, and
fiber angle (a unit vector)]) with the intensity distribution of each, in every single
structure, being modeled as a Gaussian Mixture Model (GMM), the parameters of
which are computed from the selected atlas via maximum-likelihood estimation.

The segmentation accuracy of the method is evaluated on 25 subjects with a
varying degree of pathology. In addition, we present the scan-rescan reproducibility
of the multi-contrast MALF in segmenting another dataset of 16 healthy subjects
which were scanned twice.

4.2 Multi-contrast MALF

In the context of diffusion tensor imaging, the algorithm will accommodate vector-
valued images. For each atlas $a$, $I^a = [I^a_{FA}, I^a_{MD}, I^a_x, I^a_y, I^a_z]$, where $I^a_{FA}$ denotes the
gray-scale FA image of the atlas subject, $I^a_{MD}$ denotes the gray-scale MD image of
the atlas subject, and $I^a_x, I^a_y, I^a_z$ denote the absolute values of the three elements of
the primary eigenvector. In this sense, the intensity of each atlas image, at every voxel of the image domain, is a 5-element vector $I^a(x) : x \in \Omega \rightarrow \mathbb{R}^5$, with $\Omega \subset \mathbb{R}^3$ being a finite grid where the images are defined. For segmenting the whole brain into a total of 159 anatomical structures, the label image $W^a$ of atlas $a$ is defined as a function from the image domain $\Omega$ to a subset of the non-negative integers $W^a(x) : x \in \Omega \rightarrow \{0, 1, 2, 3, \ldots, 159\}$, where $W^a(x) = 0$ for voxel $x$ belonging to the unlabeled background, and $W^a(x) = k$ for voxel $x$ labeled as the $k$-th structure, $k \in \{1, 2, 3, \ldots, 159\}$. Correspondingly, the to-be-segmented subject is defined as $(I^D, W)$, where $I^D = [I^D_{FA}, I^D_{MD}, I^D_x, I^D_y, I^D_z]$ and $W$ is the label image we aim to obtain.

Similar to the Algorithm 2 described in section 2.2.3 of Chapter 2, the multi-contrast MALF algorithm is summarized in Algorithm 3

**Algorithm 3 (Multi-contrast MALF via EM)**

We Initialize $W^{old}(x), \varphi^{old}_a(x), \forall x \in \Omega$ for each atlas $a$.

1. Compute the “atlas selector” function $\hat{p}(a, \varphi^{old}_a(x)|W^{old}(x), I^D(x))$ as in Eq. (2.14)

2. Obtain a new segmentation label via $W^{new} = \arg \max W Q(W; W^{old})$, where $Q(W; W^{old})$ is computed as in Eq. (2.15).

3. For each local ROI, recalculate the locally optimized diffeomorphisms of atlases onto segmentation labels via Eq. (2.16).

4. Update the parameter of interest $W^{old} \leftarrow W^{new}$ and the optimal diffeomorphism
\( \widehat{\varphi}_a^{\text{old}} \leftarrow \widehat{\varphi}_a^{\text{new}} \), go to Step 2.

The above iteration is continued until convergence of the parameter estimator or the number of performed iterations is bigger than a pre-specified threshold.

Compared with single-contrast MALF, there are two main differences in multi-contrast MALF:

- **Computation of** \( p(I^D(x)|W^{\text{old}}(x), a) \)

  In the single-contrast T1 case, the term \( p(I^D(x)|W^{\text{old}}(x), a) \) is computed as demonstrated in Eq. (2.22). In the multi-contrast DTI case, the scalar-valued \( I^D(x) \) is replaced with the vector-valued \( I^D(x) = [I^D_{\text{FA}}(x), I^D_{\text{MD}}(x), I^D_x(x), I^D_y(x), I^D_z(x)] \). Assuming conditional independence, we compute \( p(I^D(x)|W^{\text{old}}(x), a) \) as

  \[
  p(I^D(x)|W^{\text{old}}(x), a) = p(I^D_{\text{FA}}(x), I^D_{\text{MD}}(x), I^D_x(x), I^D_y(x), I^D_z(x)|W^{\text{old}}(x), a) \\
  = \prod_{m \in \{\text{FA, MD, x, y, z}\}} p(I^D_m(x)|W^{\text{old}}(x), a), \tag{4.1}
  \]

  with, for example, \( I^D_{\text{FA}}(x) : \Omega \to \mathbb{R} \) indicating the FA value of the test subject at the voxel \( x \). In calculating each factor of the product in Eq. (4.1) such as \( p(I^D_{\text{FA}}(x)|W^{\text{old}}(x), a) \), we model it as the probability density function of a mixture of Gaussians [127], the parameters of which are computed from the selected atlas \( a \). To be specific, for atlas \( a \), the distribution of the intensity
$I^a_{FA}(x)$ in each single label $k$ is estimated as a Gaussian Mixture Model (GMM),

$$p(I^a_{FA}(x)|W^a(x) = k) = \sum_{t=1}^{M} p(I^a_{FA}(x)|W^a(x) = k, t)\alpha^a_t,$$  \hspace{1cm} (4.2)

where $M$ denotes the total number of Gaussians in the mixture model,

$$p(I^a_{FA}(x)|W^a(x) = k, t)$$ represents the probability density function of a single Gaussian

$$p(I^a_{FA}(x)|W^a(x) = k, t) = \frac{1}{\sqrt{2\pi}\sigma^a_t} \exp\left\{ -\frac{(I^a_{FA}(x) - \mu^a_t)^2}{2(\sigma^a_t)^2} \right\},$$  \hspace{1cm} (4.3)

and $\sum_t \alpha^a_t = 1$ are the mixing coefficients for the different single Gaussians. To obtain the parameters of the GMM associated to a specific label $k$,

$$\theta^a_k = (\mu^a_t, \sigma^a_t, \alpha^a_t), \hspace{1cm} t = 1, 2, ..., M,$$  \hspace{1cm} (4.4)

we employ the EM algorithm to derive the maximum-likelihood estimators. The term $p(I^D_{FA}(x)|W^{old}(x), a)$ in Eq. (4.1) is computed according to

$$p(I^D_{FA}(x)|W^{old}(x) = k, a) = \sum_{t=1}^{M} \frac{1}{\sqrt{2\pi}\sigma^a_t} \exp\left\{ -\frac{(I^D_{FA}(x) - \mu^a_t)^2}{2(\sigma^a_t)^2} \right\} \alpha^a_t.$$  \hspace{1cm} (4.5)

The total numbers of Gaussians $M$ for the mixtures are pre-defined. We set $M = 2$ for structures with volume size smaller than 1000 $mm^3$ and $M = 4$ otherwise, in computing $p(I^D_i(z)|W^{old}(x) = k, a)$, for $i =$ FA, MD, x, y, z.

- **Initialization of the diffeomorphism $\hat{\varphi}^{old}_a$**

In the DTI case, the initial value of the optimal diffeomorphism associated to the atlas $a$, is obtained from a two-channel LDDMM-image mapping with one
channel being the FA images and the other the MD images. This two-channel LDDMM is a validated method for registering DT images [36]. Given a pair made up of atlas \( I \) and target \( J \), we compute a diffeomorphic deformation \( \varphi \) between the two vector-valued images \( I = [I_{FA}, I_{MD}] \) and \( J = [J_{FA}, J_{MD}] \) such that \( J = I \circ \varphi^{-1} \) or \([J_{FA}, J_{MD}] = [I_{FA} \circ \varphi^{-1}, I_{MD} \circ \varphi^{-1}]\). Similar to Eq. (2.19), the optimal diffeomorphic deformation in DTI is generated by integrating the vector field which is found to minimize the energy

\[
E = \int_0^1 \|v_t\|^2 dt + \frac{1}{\sigma_{FA}^2} \|I_{FA} \circ \varphi^{-1} - J_{FA}\|_{L^2}^2 + \frac{1}{\sigma_{MD}^2} \|I_{MD} \circ \varphi^{-1} - J_{MD}\|_{L^2}^2
\]

(4.6)

with \( v_t \) satisfying that \( \frac{\partial \varphi_t}{\partial t} = v_t (\varphi_t) \). The parameters \( \sigma_{FA} \) and \( \sigma_{MD} \) control the weighting of the two contrast-matching terms, relative to smoothness regularization term. We set \( \sigma_{FA} = \sigma_{MD} = 1 \).

## 4.3 Validation Analysis

### 4.3.1 Patient populations

All atlas subjects and the first testing dataset are obtained from the existing clinical database of pediatric brain MRI, and are older than 24 months of age. DT images from 16 subjects (Female = 7, Male = 9, age = 7.67 ± 4.12 years) are used to create the multiple atlases (Table 4.1). Among these 16 subjects, 10 subjects are
Table 4.1: Anatomical changes in the sixteen subjects used as the multiple DTI atlases.

<table>
<thead>
<tr>
<th>Atlas ID</th>
<th>Radiological findings</th>
<th>Radiological diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>White matter T2 hyperintensity involving the bilateral periventricular and deep white matter with restricted diffusion spots</td>
<td>Drug-induced leukoencephalopathy</td>
</tr>
<tr>
<td>2</td>
<td>T2-hyperintense lesions in periventricular and subcortical white matter</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>3</td>
<td>Multiple encephalomalacia/gliosis change related to sequelae from prior ischemic events</td>
<td>Moyamoya-disease</td>
</tr>
<tr>
<td>4</td>
<td>Diffuse CSF space dilatation</td>
<td>Associated finding with achondroplasia</td>
</tr>
<tr>
<td>5</td>
<td>Multiple T2-hyperintense lesions in white matter and gray matter</td>
<td>Neurofibromatosis type1</td>
</tr>
<tr>
<td>6</td>
<td>Mild ventricle dilatation with irregular shape and volume loss of periventricular white matter with posterior dominant</td>
<td>Periventricular leukomalacia</td>
</tr>
<tr>
<td>7 ~ 16</td>
<td>No abnormal finding</td>
<td>Diagnosed as normal</td>
</tr>
</tbody>
</table>

diagnosed as normal. In order to cover the wide range of anatomical phenotypes in the multiple atlases, the remaining 6 subjects included in the atlas set display various anatomical abnormalities, as shown in Table 4.1.

The accuracy of the segmentations obtained from multi-contrast MALF is tested based on another 25 DT images (Female = 10, Male = 15, age = 7.88 ± 4.80 years). As shown in Table 4.2, 10 subjects (Test #1 ~ #10) present a normal MR anatomy and the other 15 subjects present various anatomical abnormalities; 7 subjects (Test #11 ~ #17) are evaluated as mild to moderate anatomical change and 8 subjects
(Test #18 ∼ #25) are evaluated as severely abnormal based on the visual evaluation of a pediatric neuroradiologist.

For the scan-rescan reproducibility test, sixteen healthy volunteers with no history of neurological conditions (8 M/8 F, 22-61 years old, mean age: 31 years old) participated in this study. This is the same data used by [128], where details of the protocol can be found.

4.3.2 MRI scans

For the first dataset, MR imaging was performed using a 1.5T scanner (Avanto; Siemens, Erlagen, Germany). All patients underwent routine clinical multiplanar T1, T2, and FLAIR pulse sequences, including DTI that was obtained using a single-shot EPI with parallel acquisition. Diffusion weighting was performed along 21 independent axes with $b = 1000 \text{ s/mm}^2$, and repeated twice to enhance the signal-to-noise ratio (SNR) ($\text{TE} = 84 \text{ ms}, \text{TR} = 7700 \text{ ms}$). The DTI was scanned in the axial orientation with an imaging matrix of $96 \times 96 \text{ mm}$ (to $192 \times 192 \text{ mm}$ with zero-filled interpolation); FOV $240 \times 240 \text{ mm}$; and slice thickness $2.5 \text{ mm}$.

For the second dataset, subjects were scanned twice using a 3T MR scanner (Achieva, Philips Healthcare, Best, The Netherlands). The DTI dataset was acquired using a multi-slice, single-shot, echo-planar imaging (EPI), spin–echo sequence ($\text{TR}/\text{TE} = 6281/67\text{ms}, \text{SENSE factor} = 2.5$). Sixty-five transverse slices were acquired parallel to the line connecting the anterior commissure (AC) to the posterior
Table 4.2: Anatomical changes in the 25 subjects used for testing the multi-contrast MALF segmentation accuracy.

<table>
<thead>
<tr>
<th>Test Subject ID</th>
<th>Radiological findings</th>
<th>Radiological diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 10</td>
<td>No abnormal finding</td>
<td>Diagnosed as normal</td>
</tr>
<tr>
<td>11</td>
<td>Mild deep white matter T2-hyperintense change and ventricle enlargement</td>
<td>Adrenoleukodystrophy</td>
</tr>
<tr>
<td>12</td>
<td>Right hemiatrophy, ventricle dilatation and mild T2-hyperintense change in deep gray matter</td>
<td>Chronic ischemic insult</td>
</tr>
<tr>
<td>13</td>
<td>Diffuse CSF space and ventricle dilatation</td>
<td>Associated finding with achondroplasia</td>
</tr>
<tr>
<td>14</td>
<td>Mild ventricle dilatation</td>
<td>Associated finding with achondroplasia</td>
</tr>
<tr>
<td>15</td>
<td>T2-hyperintense lesions in periventricular and subcortical white matter and mild ventricle enlargement</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>16</td>
<td>Porencephalic left ventricle dilatation and volume loss of left corticospinal tract</td>
<td>Prenatal hemorrhagic insult</td>
</tr>
<tr>
<td>17</td>
<td>Asymmetrical ventricle dilatation (right&gt;left)</td>
<td>Associated finding with achondroplasia</td>
</tr>
<tr>
<td>18</td>
<td>Ventricle enlargement with multiple T2-hyperintense lesions in white matter</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>19</td>
<td>Venticulomegaly associated periventricular volume loss (right&gt;left) and T2-hyperintense change</td>
<td>Perinatal hypoxic ischemic injury</td>
</tr>
<tr>
<td>20</td>
<td>Lateral ventricular enlargement with periventricular white matter volume loss and T2-hyperintense change</td>
<td>Periventricular leukomalacia</td>
</tr>
<tr>
<td>21</td>
<td>Right ventricle enlargement associated with periventricular white matter volume loss</td>
<td>Prenatal intraventricular hemorrhage</td>
</tr>
<tr>
<td>22</td>
<td>Diffuse parenchymal volume loss, CSF space dilatation and multiple ischemic lesions</td>
<td>Congenital metabolic disease</td>
</tr>
<tr>
<td>23</td>
<td>Venticulomegaly and thinning of corpus callosm</td>
<td>Ventriculomegaly</td>
</tr>
<tr>
<td>24</td>
<td>Left hemiatrophy</td>
<td>Sturge-Weber syndrome</td>
</tr>
<tr>
<td>25</td>
<td>Left parenchymal volume loss, gliosis and lateral ventricle enlargement</td>
<td>Perinatal stroke</td>
</tr>
</tbody>
</table>
commissure (PC) with no slice gap and 2.2 \( mm \) nominal isotropic resolution (FOV \( = 212 \times 212 \ mm \), data matrix \( = 96 \times 96 \ mm \), reconstructed to \( 256 \times 256 \ mm \)).

4.3.3 DTI processing

All DTI datasets are processed offline using DTIStudio software (H. Jiang and S. Mori, Johns Hopkins University, Kennedy Krieger Institute, lbam.med.jhmi.edu or www.MriStudio.org). The raw diffusion-weighted images are first co-registered to one of the b0 images using a 12-parameter affine transformation obtained from the Automated Image Registration (AIR), and then the six elements of the diffusion tensor, the fractional anisotropy (FA), and the mean diffusivity (MD) are calculated.

4.3.4 Manual delineation for accuracy measurement

Eighteen structures (sixteen white matter structures and two deep gray matter structures) are manually delineated on the pre-selected 2D slices of fourteen subjects from three groups - four from the normal group, five from the mildly abnormal group, and the other five from the severely abnormal group. To investigate the intra- and inter-rater variability of the manual delineation, two raters perform the manual delineations twice with more than 3-week intervals. To quantitatively evaluate the segmentation accuracy of our algorithm, we use four measurements:

1. The Dice overlap coefficients between the manually delineated 2D ROI and the
corresponding ROI in the automated segmentations.

2. The correlation between the size of the manually delineated ROI and that of the automated ROI.

3. The correlation between the mean FA value of the manual ROI and the mean FA of the automated ROI.

4. The correlation between the mean MD value of the manual ROI and the mean MD of the automated ROI.

To evaluate the improvement in the segmentation accuracy from the multi-contrast approach, we compare the segmentations from the 5-contrast multi-atlas approach (FA, MD, vector elements x, y, and z combined), with those obtained from the same multi-atlas but with only a single contrast – FA-only and MD-only, and a three-contrast approach – EV-only (x, y, z combined). These four methods vary from each other only in the computation of Eq. (4.1). To compare the four methods statistically, for each structure, we perform a one-way ANOVA to examine significant differences among the Dice results obtained from the four approaches. For statistical correlation analysis, we used William’s modification of Hotelling’s test.

For the scan-rescan reproducibility test, we investigate the volume difference between the automated segmentations of the same structure from the two scans of the same subject. The volume difference is computed as: 

\[ \text{VD} = \frac{|vol(ROI_1) - vol(ROI_2)|}{(vol(ROI_1) + vol(ROI_2))/2}, \]

where \( vol(ROI_1) \) denotes the volume of a specific ROI for the first scan and \( vol(ROI_2) \)
denotes the volume of the same ROI for the second scan of the same subject. In addition, we examine the difference between the mean FA value of the automated segmentation of each single structure for the first scan and that of the automated segmentation for the second scan, as well as the difference between the mean MD values.

4.4 Results

Figure 4.1 shows results of Dice measurements, reporting spatial agreement with manual delineation. The 5-contrast approach is compared with FA-only, MD-only, and EV-only approaches. Because of the unique contrast signature of each structure, the best single contrast to accurately define it varies. For example, to define the contracospinal tract (CST), the EV contrast provides the best accuracy, but it provides poor definition around the putamen, which is best defined by FA or MD. However, the 5-contrast approach performs well for all structures. According to the results from the one-way ANOVA, we find statistical differences among the 4 approaches in 11 structures, in which the 5-contrast approach consistently achieves one of the best results. These structures include: the caudate, the putamen, the cingulate gyrus, the middle cerebellar peduncle, and the corticospinal tract in both hemispheres. The absolute Dice level is 0.8 ~ 0.9. We note that some structures are difficult to define even manually with perfect reproducibility. A good example is the superior
longitudinal fasciculus (SLF), which has a vague structural boundary and the inter-rater variability is large (Dice = 0.8 ± 0.259). Because the manual definition is used as the gold standard, the spatial matching cannot be better than the inter-rater spatial matching (automated methods cannot be more accurate than manual delineation by definition).

Figure 4.1: The mean Dice overlaps and the standard deviations of the eighteen ROIs obtained from automated segmentations based on five contrasts (red), the single FA contrast (green), the single MD contrast (blue), the combined EV contrast (yellow), as well as the inter-rate (patterned). The mean values are calculated across fourteen different subjects. Star marks indicate significant difference among the four sets of Dice results by ANOVA ($p \ll 0.05$). Abbreviations are: GCC – genu of corpus callosum; BCC – body of corpus callosum; Caud – caudate; Put – putamen; ALIC – anterior limb of internal capsule; PLIC – posterior limb of internal capsule; CG – cingulate gyrus; MCP – middle cerebellar peduncle; SLF – superior longitudinal fasciculus; CST – corticospinal tract.

The correlation coefficients between the sizes of the manual and the automated segmentations obtained from the four approaches for all the eighteen ROIs are listed in Table 4.3. For some structures, the ROI sizes from all the four automated approaches
Table 4.3: Pearson’s correlations between the manual and four different automated ROI area measures, with bold typesetting indicating that the correlation between the automated and the manual measures is statistically significant compared to other methods ($p < 0.025$).

<table>
<thead>
<tr>
<th></th>
<th>GCC</th>
<th>Caud-r</th>
<th>Put-r</th>
<th>ALIC-r</th>
<th>PLIC-r</th>
<th>CG-r</th>
<th>MCP-r</th>
<th>SLF-r</th>
<th>CST-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-contrast</td>
<td>0.79541</td>
<td>0.96321</td>
<td>0.94928</td>
<td>0.98448</td>
<td>0.95195</td>
<td>0.83631</td>
<td>0.89982</td>
<td>0.89484</td>
<td>0.94142</td>
</tr>
<tr>
<td>MD-only</td>
<td>0.62269</td>
<td>0.91332</td>
<td>0.93717</td>
<td>0.98027</td>
<td>0.89410</td>
<td>0.09248</td>
<td>0.82841</td>
<td>0.59750</td>
<td>0.76529</td>
</tr>
<tr>
<td>FA-only</td>
<td>0.78993</td>
<td>0.55538</td>
<td>0.95956</td>
<td>0.97514</td>
<td>0.90322</td>
<td>0.21467</td>
<td>0.87219</td>
<td>0.69858</td>
<td>0.76673</td>
</tr>
<tr>
<td>EV-only</td>
<td>0.69702</td>
<td>0.49411</td>
<td>0.90326</td>
<td>0.96372</td>
<td>0.88237</td>
<td>0.87709</td>
<td>0.88976</td>
<td>0.94912</td>
<td>0.94356</td>
</tr>
</tbody>
</table>

are highly correlated with the ROI sizes of the manual delineations. However, for structures such as the caudate, the corticospinal tract (CST), and the cingulate gyrus (CG), the performance varies from approach to approach. In Figure 4.2 and Figure 4.3, we show examples of the correlation plot between the automated and manual approaches for gray matter (caudate) and white matter (CST) structures.

Figure 4.4 shows actual segmentation results of the CST in the brainstem of subjects with different degrees of abnormalities, which demonstrates how the integration of the 5-contrast information can accurately delineate the sizes. In this example, the fiber-orientation information in the EV contrast is necessary to accurately reflect the small CST sizes in Case #3. Namely, the CST has a characteristic Z-axis (blue) fiber orientation, which can uniquely differentiate the CST from the surrounding high-FA white matter structures. The integration of the EV information provides
Figure 4.2: A plot of the correlation between the automated and the manual measurements of the size of the caudate in both hemispheres in square millimeters. Results from the four automated segmentation methods are compared: 5-contrast (red), FA-only (green), MD-only (blue), and EV-only (yellow).
Figure 4.3: A correlation plot between the automated and manual measurements of the sizes of left and right corticospinal tracts (CST). Results from the four automated segmentation methods are compared: 5-contrast (red), FA-only (green), MD-only (blue), and EV-only (yellow).
strong constraints for the segmentation, particularly in defining the high-FA regions with a strong orientation alignment along the Z-axis. The FA-, MD-, and EV-only approaches could extract the CST accurately for Case #1 and #2, but grossly overestimate the CST size for Case #3.

Figure 4.4: Demonstration of the segmentation accuracy of the CST in three representative cases with different degrees of anatomical abnormalities. Results from four different approaches are compared. From left to right: the color-coded map without CST definition superimposed, the color-coded map with CST from the 5-contrast MALF superimposed, the color-coded map with CST from the FA-only single-contrast MALF superimposed, the color-coded map with CST from the MD-only single-contrast MALF superimposed, and the color-coded map with CST from the EV-only 3-contrast MALF superimposed.

Based on the comparison results shown in Table 4.3, significant improvement of the correlation between the sizes of the automated segmentations and that of the manual delineations is achieved by the 5-contrast approach compared with the other single contrast approaches. Likewise, we perform the manual-auto correlation analyses of the mean FA and MD values within each single ROI. As shown in Table 4.4 and
Table 4.4: Pearson’s correlations between the mean FA value within the manual ROI and the automated segmentations, with bold typesetting suggesting that the correlation between the automated measure and the manual one is statistically significant compared to the other methods ($p < 0.025$).

<table>
<thead>
<tr>
<th></th>
<th>GCC</th>
<th>Caud-r</th>
<th>Put-r</th>
<th>ALIC-r</th>
<th>PLIC-r</th>
<th>CG-r</th>
<th>MCP-r</th>
<th>SLF-r</th>
<th>CST-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-contrast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.8852</td>
<td>0.94208</td>
<td>0.98440</td>
<td>0.99331</td>
</tr>
<tr>
<td>MD-only</td>
<td>0.92838</td>
<td>0.86687</td>
<td>0.88287</td>
<td>0.98203</td>
<td>0.78654</td>
<td>0.78890</td>
<td>0.78991</td>
<td>0.83956</td>
<td>0.97452</td>
</tr>
<tr>
<td>FA-only</td>
<td>0.97415</td>
<td>0.91417</td>
<td>0.89183</td>
<td>0.99429</td>
<td>0.81501</td>
<td>0.90213</td>
<td>0.90952</td>
<td>0.98081</td>
<td>0.97258</td>
</tr>
<tr>
<td>EV-only</td>
<td>0.91020</td>
<td>0.81890</td>
<td>0.82255</td>
<td>0.98156</td>
<td>0.78727</td>
<td>0.90313</td>
<td>0.89046</td>
<td>0.98859</td>
<td>0.99241</td>
</tr>
</tbody>
</table>

Table 4.5, again, the 5-contrast approach is consistently superior to the other three approaches in terms of either FA or MD correlation.

Figure 4.5 demonstrates segmentation results of three patients with different degrees of abnormalities. A high level of segmentation accuracy is visually appreciable for the wide variety of anatomical states.

According to our test-retest experiments, the reproducibility results are 4.7%, 2.19%, and 2% for the volume, FA, and MD, respectively, averaged over all 193 structures. If we remove 26 small gray matter structures ($< 1000mm^3$), the reproducibility improves to 3.73%, 1.91%, and 1.79% respectively. These small gray matter structures have poor test-retest reproducibility because they lack clear contrasts in DTI and they are usually not the targets of DTI measurements.
Figure 4.5: Results of the whole brain segmentations into 159 structures in three representative cases with large anatomical variability. The segmentation results are superimposed on color (upper row) and MD (bottom row) images.
Table 4.5: Pearson’s correlations between the mean diffusivity (MD) value within the manual ROI and the automated segmentations with bold typesetting indicating that the correlation between the automated measure and the manual one is statistically significant compared to the other methods ($p < 0.025$).

<table>
<thead>
<tr>
<th></th>
<th>GCC</th>
<th>Caud-r</th>
<th>Put-r</th>
<th>ALIC-r</th>
<th>PLIC-r</th>
<th>CG-r</th>
<th>MCP-r</th>
<th>SLF-r</th>
<th>CST-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-contrast</td>
<td>0.97192</td>
<td>0.94152</td>
<td>0.99307</td>
<td>0.99115</td>
<td>0.98541</td>
<td>0.93410</td>
<td><strong>0.94336</strong></td>
<td>0.98455</td>
<td><strong>0.98557</strong></td>
</tr>
<tr>
<td>MD-only</td>
<td>0.94154</td>
<td><strong>0.88121</strong></td>
<td>0.99237</td>
<td>0.99452</td>
<td>0.97074</td>
<td>0.89304</td>
<td>0.81370</td>
<td>0.95021</td>
<td>0.89566</td>
</tr>
<tr>
<td>FA-only</td>
<td>0.95861</td>
<td>0.58329</td>
<td>0.99361</td>
<td>0.99191</td>
<td>0.97357</td>
<td>0.92795</td>
<td>0.90926</td>
<td>0.98013</td>
<td>0.76053</td>
</tr>
<tr>
<td>EV-only</td>
<td>0.88039</td>
<td>0.58453</td>
<td>0.98699</td>
<td>0.97750</td>
<td>0.97737</td>
<td>0.92538</td>
<td>0.84391</td>
<td>0.98684</td>
<td>0.91718</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BCC</th>
<th>Caud-I</th>
<th>Put-I</th>
<th>ALIC-I</th>
<th>PLIC-I</th>
<th>CG-I</th>
<th>MCP-I</th>
<th>SLF-I</th>
<th>CST-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-contrast</td>
<td>0.94524</td>
<td><strong>0.95291</strong></td>
<td>0.99699</td>
<td><strong>0.98404</strong></td>
<td>0.98605</td>
<td><strong>0.98599</strong></td>
<td>0.92627</td>
<td>0.95778</td>
<td><strong>0.97023</strong></td>
</tr>
<tr>
<td>MD-only</td>
<td>0.93038</td>
<td><strong>0.92279</strong></td>
<td>0.99173</td>
<td><strong>0.98246</strong></td>
<td>0.97558</td>
<td><strong>0.97598</strong></td>
<td>0.88370</td>
<td>0.95810</td>
<td>0.89994</td>
</tr>
<tr>
<td>FA-only</td>
<td>0.98557</td>
<td>0.81496</td>
<td>0.99083</td>
<td>0.10983</td>
<td>0.96575</td>
<td>0.93071</td>
<td>0.83919</td>
<td>0.96408</td>
<td>0.62843</td>
</tr>
<tr>
<td>EV-only</td>
<td>0.92346</td>
<td>0.87005</td>
<td>0.99003</td>
<td><strong>0.96607</strong></td>
<td>0.97322</td>
<td>0.73985</td>
<td>0.90548</td>
<td>0.98875</td>
<td>0.87714</td>
</tr>
</tbody>
</table>

4.5 Conclusion

In this chapter, we developed and tested an automated whole brain segmentation method based on the multi-contrast multi-atlas likelihood-fusion algorithm. We used mixtures of Gaussians to model the intensity distribution of individual contrasts within a single structure. If the intensity within the structure is homogeneous, the algorithm automatically assigns weight 1 to a single Gaussian. If there is high intra-structure variability, multiple Gaussians will be used to model the intensity, with each Gaussian given a small weight. In this way, it effectively reduces the contribution of this contrast information in computing the quantity $p(I_m(x)|W,a)$, $m \in \{FA, MD, vectorx, y, z\}$. By incorporating the consistent anatomical signatures into the segmenting criteria, we aim to achieve robust segmentations.
While the Dice measurements provide important information about the accuracy level of the automated segmentation, it is probable that the correlation results reported in Table 4.3 are more important for actual image-based studies. The anatomical delineations of brain structures depend on anatomical definitions. It is reasonable that there is a consistent difference in boundaries of defined structures between two different approaches. In this sense, low Dice values do not necessarily mean that the automated results are not useful. The high correlation between the manual and automated methods indicates that they have similar powers to differentiate different anatomical states, which is ultimately the goal of quantitative analyses. In this respect, the high correlation between the 5-contrast and manual approaches is an encouraging result.

The test-retest reproducibility shows less than 5% variability for the volume measurement of most of the white matter structures. Due to the effect of non-reproducible B0-susceptibility image distortion, we believe this is a reasonable level of reproducibility. The test-retest reproducibility measures of FA and MD indicate an even higher reproducibility (less than 3%).

To summarize, we have extended the MALF algorithm to the multi-contrast context. Validity of the multi-contrast MALF algorithm has been demonstrated in terms of both overlap accuracy when compared to manual delineations and reproducibility. This may suggest that the multi-contrast MALF algorithm is ready to be applied to clinical studies focusing on anatomical phenotypes of whole brain structures.
Chapter 5

Surface-based Morphometric Analysis

5.1 Introduction

Metric comparison based analyses have been dominant in the field of CA for the past two decades. It is a powerful tool for examining the anatomical abnormalities or functional variabilities in cross-sectional studies as well as longitudinal studies. In studying the structural abnormalities of a specific disease group, volumetric assessment, comparing the volumes of a specific brain structure in subjects from different groups, is the most popular and direct way. As shown in Chapter 3, statistical analyses based on the volumetric measurements are efficient from the standpoint of numerical computing and easily interpretable from that of biology, making it the most widely
used metric comparison method in medical image based clinical analyses.

One potential limitation of volumetric analysis is that a change in the volume size of a single structure does not provide detailed information about the locations of any regional differences between the disease group and a control group. In addition, an evaluation based on the volume size of a structure is incapable of identifying the variation in the structure that may be induced by the disease. Moreover, for a single structure, it is plausible that a part of it is undergoing atrophy while another part is expanding, which makes the overall volume size unaffected. In this case, simply evaluating whether there is a change in the structure volume size cannot indicate whether this structure is affected or not.

To identify the magnitude and the pattern of structural changes in a specific disease group, new tools that enable the detection of subtle changes in neuroanatomy have been sought. Brain warping techniques such as LDDMM have been reported to characterize region-specific variations in numerous neurodegenerative disease studies in terms of either volume or shape analysis.

In this chapter, we first describe a surface-based morphometric (SBM) analysis pipeline that was originally proposed in [45]. A simulation study is performed to validate the SBM analysis pipeline in detecting local shape differences between two groups. The SBM pipeline is then applied to characterizing the shape abnormalities of seven subcortical and ventricular structures (the lateral ventricular system, the memory related amygdala-hippocampal circuit, and the thalamic and basal ganglia
circuits, which receive projections from the amygdala and hippocampus) in subjects
with MCI or AD within the framework of LDDMM. The primary goal of this study
is to quantitatively assess whether the surface (a 2D manifold) contouring a single
structure, i.e. the shape, differs as a function of disease severity in prodromal and
mild AD, and whether it could provide accurate prognostic information in patients
with AD. Specifically, the shape differences among the three groups – HC, MCI, and
AD are investigated. The MCI subjects are stratified into three sub-groups according
to the longitudinal clinical information: 1) MCI patients who reverted to normal
status, 2) MCI patients who remained stable, 3) MCI patients who deteriorated to
AD. The shapes, at the baseline, of the aforementioned structures in these three MCI
sub-types are being studied here to determine whether differences exist that would
help in predicting cognitive decline.

5.2 Method

5.2.1 Template Surface Generation

The SBM analysis pipeline is built on the basis of a common anatomical coor-
dinate system to ensure statistical group comparisons at equivalent spatial locations.
Therefore, the first step is to create a common template surface for each structure
of interest. To reduce the difference between the template and the study population,
the template surface is usually generated from a subset of the population. For each
structure, every subject surface is first rigidly aligned (rotation and translation) to a common spatial position. The rigid registration algorithm computes an optimal transformation between the vertex sets of the two surfaces $S_0$ and $S_1$, by minimizing a score that combines registration and soft assignment, which is similar to the one considered in [129]. In detail, let $x_i$ ($i = 1, 2, ..., M$) denote the set of vertices on the template surface, and $y_j$ ($j = 1, 2, ..., N$) the set of vertices on the target surface, the cost function is:

$$E = \lambda(3 - \text{tr}(R)) + \sum_{i=1}^{M} \sum_{j=1}^{N} (w_{ij} + v_{ij})\|Rx_i + T - y_j\|^2 + t(w_{ij} \log w_{ij} + v_{ij} \log v_{ij}),$$

(5.1)

for some $\lambda > 0$, subject to constraints $\sum_{i=1}^{M} w_{ij} = 1$, $\sum_{j=1}^{N} v_{ij} = 1$, $w_{ij} \geq 0$, $v_{ij} \geq 0$. The term $(3 - \text{tr}(R))$ is the regularization term. The matching term and the soft assign term are given by $(w_{ij} + v_{ij})\|Rx_i + T - y_j\|^2$ and $t(w_{ij} \log w_{ij} + v_{ij} \log v_{ij})$ respectively.

$R$ is the rotation matrix and $T \in \mathbb{R}^3$.

After rigid registration, we computed an averaged template surface, using the algorithm described in [130]. Each observed subject surface is modeled as a random deformation of a hidden template plus additive Gaussian noise. Given this model, the template is estimated from the subject surfaces using an approximation of the EM algorithm, subject to some topology constraints. It is enacted by ensuring that the hidden template surface is a diffeomorphic deformation of a reference shape, called a hyper-template.
5.2.2 LDDMM-surface Registration

After obtaining the template surface for each structure, we performed LDDMM-surface mapping [35] to compute a diffeomorphic registration between the template surface and each target surface. In the LDDMM setting, the set of anatomical shapes is placed into a metric space. This is modeled by assuming that one shape can be generated from another via group actions of diffeomorphisms, i.e., that compared shapes are topologically equivalent, which is true for the subcortical structures and the ventricles which we focus on in this chapter. To compare shapes, we generate time-dependent diffeomorphisms by solving the ordinary differential equation

$$\dot{\phi}_t = v_t(\phi_t), \quad t \in [0, 1] \quad (5.3)$$

with

$$\phi_0 = id$$

where \(\dot{\phi}_t\) is the time derivative of \(\phi_t\).

Given a template surface \(S_{\text{temp}}\) and a target shape \(S_{\text{obs}}\) (the observed subject surface), the inexact matching registration algorithm minimizes the functional:

$$E(v) = \int_0^1 \|v_t\|^2 dt + D(\phi_1 \cdot S_{\text{temp}}, S_{\text{obs}}), \quad (5.2)$$

with

$$\dot{\phi}_t = v_t(\phi_t), \quad \phi_0 = id$$

where \(\phi_1 \cdot S_{\text{temp}}\) is the deformed template resulting from the action of the diffeomorphism at time \(t = 1\) on the template surface. The function \(D\) is a discrepancy measure between surfaces. After minimization, the integral term in the cost function can be interpreted as a squared geodesic distance in shape space between the template and
the deformed template. The norm $\|\cdot\|_V$ is a Hilbert norm, $V$ being a reproducing kernel Hilbert space of vector fields. To ensure that the solutions are diffeomorphisms, $V$ must be a space of smooth vector fields.

The solutions take a special form after discretization. Assume that the surfaces are triangulated, and let $x^j$ denote the vertices of the template surface $S_{temp}$. It has been proved that the solution of Eq. (5.2) and Eq. (5.3) must be of the form

$$v_t(x) = \sum_j k_V(x_t^j, x) \alpha_t^j,$$

(5.4)

where $k_V$ denotes the reproducing kernel of the space $V$ and $\alpha$ is the momentum vector $[31, 131]$. In practice, $k_V$ is selected to be a Gaussian kernel in the sense that $k_V(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma_V^2}\right)$. After substitution, Eq. (5.3) can be equivalently put in the form:

$$E(\alpha) = \int_0^1 \alpha_t^T K(x_t) \alpha_t dt + D(\phi_1 \cdot S_{temp}, S_{obs}),$$

(5.5)

with

$$\dot{\phi}_t = k_V \alpha_t (\phi_t), \quad \phi_0 = \text{id}.$$  

(5.6)

$K(x_t)$ is the matrix formed with $k_V(x_t^i, x_t^j)$. 

119
5.2.3 Vertex-based Statistical Analysis

From the LDDMM-surface mapping, we calculate a scalar field on the template surface according to

\[ J = \log(\det(D\phi_t)), \]  

(5.7)

where \( D\phi_t \) is the Jacobian matrix of \( \phi_t \). This scalar field measures the localized expansion or compression at each vertex of the subject relative to the template in the logarithmic scale: i.e. positive values correspond to surface expansion of the subject’s structure relative to the template at a particular location, while negative values denote surface compression of the subject’s structure relative to the template.

For each subject \( s \), we shall term this scalar field the “deformation marker” \( J_k(s) \) which is indexed at each vertex \( k \) of the template surface. This vertex-based analysis (which therefore restricts to the shape boundaries) arises naturally for studying shape changes since, in our case, the subcortical structures at 1 mm scale MRI appear constant in contrast within a single structure (therefore little information is available inside a single structure).

A binary group variable \( Y(s) \) is used to represent the diagnosis variable in our model. Let \( Y(s) = 1 \) if subject \( s \) belongs to the disease group in comparison and \( Y(s) = 0 \) if subject \( s \) belongs to the control group. The statistical model is given by (at each vertex \( k \)):

\[ J_k(s) = \beta_{k,0} + \beta_{k,1}Y(s) + \sum_{\text{cov}} \alpha_{\text{cov}}X_{\text{cov}}(s) + \varepsilon_k(s), \]  

(5.8)
where $X_{\text{cov}}(s)$ is the covariate information (confounding factors) included in the analysis, such as age, sex, and the estimated intracranial volume. We test for the null hypothesis that $\beta_{k,1} = 0$ separately for each vertex $k$. Statistics are therefore computed at each vertex of the triangulated template surface, and $p$-values are corrected for multiple comparisons. More precisely, for each $k$, we define the vertex-based statistic as

$$F_k = \frac{RSS_0(k)}{RSS(k)} - 1,$$

(5.9)

where $RSS_0$ is the residual sum of squares under the null hypothesis, and $RSS$ is the residual sum of squares under the general hypothesis. A joint statistic is defined as $F^* = \max_k F_k$, the maximum value of the statistics over all vertices. The statistical significance of group differences is measured using Fisher’s method of randomization. We utilized Monte Carlo simulations to generate 40,000 uniformly distributed random permutations of the group labels, which gives rise to a collection of the $F^*$ statistic coming from each permutation. The $p$-value for the significance of the group labels is then given by the fraction of the times that the values of $F^*$ from the permutations are larger than the value obtained from the true groups. The set of vertices on which the null hypothesis is not valid is estimated to be:

$$D = \{ k : F_k \geq q^* \},$$

(5.10)

where $q^*$ is the 95-th percentile bound of the collection of the $F^*$ statistic from the permutation tests [132] and $F_k$ is the observed statistic at vertex $k$ (with the true
labels). To quantify the group shape variation (compression or expansion), we define the degree of the group shape differences as the negative of the $\beta_{k,1}$ coefficients. Thus, negative values denote expansion in the latter group compared to the former group in comparison while positive values denote relative compression in the latter group.

5.3 Simulation Study

5.3.1 Mathematical Modeling

We first create two sets of artificial shapes that differ mainly in length, not in thickness. We start with a superquadric template $T$, which is parameterized by

$$|x|^p + \frac{|y|^p}{2} + |z|^p = 2^p,$$

(5.11)

where we select $p = 1.5$. To create surfaces for the group with relatively larger length, we use the following procedure:

1. Introduce a smooth random noise indexed at each vertex $k$ of the template $T$ as

$$\alpha(k) = \sum_{j=1}^{25} \left( \frac{x_i(j)}{1 + j} \right) e_j(k),$$

(5.12)

where $e_j$ denotes the $j$-th Laplace-Beltrami (LB) eigenvector on the template. The quantity $x_i(j)$ is modeled as: $x_i(j) \sim N(0, \sigma^2)$, with $\sigma = 0.05$. 

122
2. Define a vector field $v$ as

$$ v = K(\alpha N), \quad (5.13) $$

where $N$ denotes the normal vector at every vertex, and $K$ is a smoothing Gaussian kernel with a variance of 1 voxel.

3. Evolve the surface at all vertices according to

$$ \frac{dk}{dt} = v(k), \quad 0 < t < 1. \quad (5.14) $$

To create the surfaces with shorter length, we take exactly the same steps as above. The only difference lies in Step 2. To be specific, we compute $\alpha$ as

$$ \alpha(k) = \sum_j \left( \frac{x_i(j)}{1 + j} \right) e_j(k) + \mu \left( \max(|y(k)| - y_0, 0) \right), \quad (5.15) $$

where $y_0$ is a thresholding constant, the 95 percentile of all $y$, and $\mu$ is uniformly distributed over $[0, 0.1]$.

### 5.3.2 Experimental Results

The SBM analysis results of comparing the local surface areas between the two sets of artificial shapes are shown in Figure 5.1, from which we find that the shape variations occur mainly at the two tails of the template surface, the location where the change of length occurs. To be specific, compared with the set of shapes with larger length, we detect local compression in the two tails of the shapes with shorter
length. Also, according to the quantitative results, the farther a vertex is from the common center of the two sets of shapes, the stronger the shape difference is. This result matches what we were expecting and validates the effectiveness of the analysis pipeline.

By designing such a simulation study with the length being the single variable, we successfully showed that the SBM analysis pipeline is capable of detecting and quantifying the shape variations induced by a change of length but not thickness.

Figure 5.1: Shape differences between two sets of artificial shapes that vary mainly in length rather than thickness. Warm color suggests compression while cool color suggests expansion in the group with shorter length when compared with the other group. The scale value measures the ratio of the local volume of the group with larger length to that of the group with shorter length in logarithmic scale.
5.4 Clinical Study

5.4.1 Alzheimer’s Disease Neuroimaging Initiative

Data used in the preparation of this study were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to
participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

5.4.2 Participants

In this study, we include data from 210 HC subjects, 369 subjects with MCI, and 175 subjects with AD. Within the MCI group, 369 subjects are further divided into three groups: MCI-HC – those who reverted to normal cognitive status (13 subjects); MCI-MCI – those who remained stable (205 subjects); and MCI-AD – those who converted to AD (151 subjects), according to the clinical information of the population after a follow-up of one year. Since the MCI-HC group is very small, we exclude it from our analysis. We term the MCI-MCI group as MCI-stable. Group clinical and demographic data are presented in Table 5.1. Briefly, subjects are 55-92 years old, and are not depressed. The control subjects have Mini-Mental State Examination (MMSE) scores of 25-30 and a clinical dementia rating (CDR) of 0. The subjects with MCI have MMSE scores of 23-30, a CDR of 0.5, preserved ability to perform daily living activities, and absence of dementia. The subjects with AD have MMSE scores of 20-28 and a CDR of 0.5 or 1.0 and meet the criteria for probable AD.

The subject groups did not differ significantly in age ($F = 2.53$, $p = 0.081$). All
Table 5.1: Demographic information of the baseline dataset in the ADNI study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>HC Group (n = 210)</th>
<th>MCI Group (n = 369)</th>
<th>AD Group (n = 175)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject age (y)</td>
<td>76.25 ± 5.01</td>
<td>75.03 ± 7.32</td>
<td>75.28 ± 7.49</td>
</tr>
<tr>
<td>No. of male subjects</td>
<td>109</td>
<td>236</td>
<td>94</td>
</tr>
<tr>
<td>MMSE score</td>
<td>29.12 ± 1.02</td>
<td>27.01 ± 1.76</td>
<td>23.43 ± 2.01</td>
</tr>
<tr>
<td>CDR-SB score</td>
<td>0.03 ± 0.12</td>
<td>1.60 ± 0.89</td>
<td>4.23 ± 1.64</td>
</tr>
</tbody>
</table>

groups differed on MMSE and clinical dementia rating scale sum of boxes (CDR-SB) as expected based on diagnostic criteria (all $p < 0.001$).

5.4.3 Image protocol and volumetric segmentation

The volume segmentations of all the seven structures are created from raw DICOM MR scans downloaded from the public ADNI website. Locally, the raw MR data are automatically corrected for spatial distortion due to gradient nonlinearity [133] and B1 field inhomogeneity [104]. The two T1-weighted images from each subject are rigid-body aligned to each other and then averaged to improve signal-to-noise ratio and resampled to isotropic 1-mm voxels. Volumetric segmentations for the hippocampus, amygdala, caudate, putamen, globus pallidus, thalamus, and lateral ventricle, in both sides, are created using FreeSurfer [5]. Based on the transformation of the full brain mask into atlas space, total cranial vault value is estimated from the atlas scaling factor [134] to control individual differences in head size.
The quality of the automated volumetric segmentations has been reviewed. Failed subjects are excluded from the analysis. A qualitative review is performed, with blinding to the diagnostic status, by one of three technicians who have been trained and supervised by an expert neuroanatomist with more than 10 years of experience. The technicians had a minimum of 4 months of experience reviewing brain MR images prior to their involvement in this project. Images that suffer degradation due to motion artifacts, technical problems (change in scanner model or change in RF coil during the time-series), or significant clinical abnormalities (e.g., hemispheric infarction) are excluded. As a result, the number of scans is reduced by approximately 15%.

5.4.4 Surface generation

In preparation for SBM analysis, all volumetric segmentations of the seven structures in both sides are transformed into triangulated surfaces using a pipeline built on the LDDMM-image algorithm. A template set of the seven structures (left and right), the Computational Functional Anatomy (CFA) subcortical templates [135], were created from a separate set of 41 manually labeled volumes. In this CFA subcortical template set, each structure has its 3D binary volume representation as well as a smooth 2D surface contouring the volume. To be specific, the CFA subcortical template consists of 14 binary images $I_{\text{temp}} = \{I_{\text{temp,struct}_1}, I_{\text{temp,struct}_2}, \ldots, I_{\text{temp,struct}_{14}}\}$ and 14 surfaces bounding the corresponding structure images.
temp = \{S_{\text{temp}\text{-struct}_1}, S_{\text{temp}\text{-struct}_2}, \ldots, S_{\text{temp}\text{-struct}_{14}}\}. For each subject, the corresponding volume segmentation images of the 14 structures,

\(I_{\text{sub}_i} = \{I_{\text{sub}_i\text{-struct}_1}, I_{\text{sub}_i\text{-struct}_2}, \ldots, I_{\text{sub}_i\text{-struct}_{14}}\}\), are created by FreeSurfer. A 14-channel LDDMM-image mapping [36] is performed to obtain a diffeomorphic change between the template coordinate system and the subject coordinate system, with each individual channel being the volume image of a structure. To do this, we define a distance function, between the deformed template and the i-th subject, as:

\[
D \left( I_{\text{temp}} \circ \phi_i^{-1}, I_{\text{sub}_i} \right) = \sum_{j=1}^{14} \left\| I_{\text{temp}\text{-struct}_j} \circ \phi_i^{-1} - I_{\text{sub}_i\text{-struct}_j} \right\|^2_{L^2},
\]

where the optimizing deformation \(\phi_i\) is generated as the end point, \(\phi_i = \varphi_t^v\), of the flow of a smooth time-dependent vector field, \(v_t \in V, t \in [0, 1]\) with the ordinary differential equation, \(\frac{d\varphi_t^v}{dt} = v_t(\varphi_t^v), t \in [0, 1]\), for \(V\) a reproducing kernel Hilbert space with a smooth kernel and norm \(\| \cdot \|_V\). The optimal diffeomorphism solves the matching problem

\[
E(v) = \int_0^1 \| v_t \|^2_{V} dt + D \left( I_{\text{temp}} \circ \phi_i^{-1}, I_{\text{sub}_i} \right).
\]

The deformed template segmentations, corresponding to the subject, are given by

\[
\hat{I}_{\text{sub}_i} = \left\{ \hat{I}_{\text{sub}_i\text{-struct}_1}, \hat{I}_{\text{sub}_i\text{-struct}_2}, \ldots, \hat{I}_{\text{sub}_i\text{-struct}_{14}} \right\} = \left\{ I_{\text{temp}\text{-struct}_1} \circ \phi_i^{-1}, I_{\text{temp}\text{-struct}_2} \circ \phi_i^{-1}, \ldots, I_{\text{temp}\text{-struct}_{14}} \circ \phi_i^{-1} \right\}. \tag{5.17}
\]

These can be regarded as the “filtered” or “denoised” approximations of the subject’s structure segmentations. The surface representations of the subject’s structures are
then created by applying the deformation $\phi_i$ on the template surfaces:

$$S_{\text{sub}_i} = \{S_{\text{sub}_i, \text{struct}_1}, S_{\text{sub}_i, \text{struct}_2}, \ldots, S_{\text{sub}_i, \text{struct}_{14}} \}$$

$$= \{\phi_i \cdot S_{\text{temp, struct}_1}, \phi_i \cdot S_{\text{temp, struct}_2}, \ldots, \phi_i \cdot S_{\text{temp, struct}_{14}} \}. \quad (5.18)$$

The surfaces $S_{\text{sub}_i}, i = 1, 2, \ldots, 754$, are the ones our statistical analyses are based on in subsequent sections. LDDMM carries the smooth sub-manifold diffeomorphically, and thus is capable of maintaining the smooth boundary and the correct topology of the template surfaces in the target surfaces. We quantitatively compare the structure volumes after the de-noising procedure with the original FreeSurfer volumes in terms of kappa overlap [136] and volume difference. As shown in Figure 5.2, for each structure, an average kappa overlap above 0.85 is obtained. For each structure, the average volume of the segmentations from the de-noising procedure and that of the original FreeSurfer segmentations, as well as their differences, are listed in Table 5.2. For a majority of structures, the mean volume discrepancy is within 10%. The discrepancy mostly occurs where the FreeSurfer segmentations are not smooth, have topological errors, or thin structures that FreeSurfer is not able to identify in the MR image.

### 5.4.5 Results

**Comparisons of HC, MCI, and AD**

Results obtained from the vertex-based statistical analyses are summarized in
Table 5.2: Average volume measurements ($mm^3$) of the original segmentations from FreeSurfer and the ones “filtered” from the LDDMM-based pipeline as well as their mean differences. Abbreviations: lvent - left lateral ventricle, ltha - left thalamus, lcaud - left caudate, lputa - left putamen, lpall - left globus pallidus, lhipp - left hippocampus, lamyg - left amygdala, rvent - right lateral ventricle, rtha - right thalamus, rcaud - right caudate, rputa - right putamen, rpall - right globus pallidus, rhipp - right hippocampus, ramyg - right amygdala.

<table>
<thead>
<tr>
<th></th>
<th>Original Segmentation</th>
<th>Filtered Segmentation</th>
<th>Mean Volume Difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lvent</td>
<td>22132</td>
<td>21899</td>
<td>1.05%</td>
</tr>
<tr>
<td>lthal</td>
<td>6395</td>
<td>6255</td>
<td>2.18%</td>
</tr>
<tr>
<td>lcaud</td>
<td>3362</td>
<td>2981</td>
<td>11.34%</td>
</tr>
<tr>
<td>lputa</td>
<td>4648</td>
<td>4456</td>
<td>4.14%</td>
</tr>
<tr>
<td>lpall</td>
<td>1631</td>
<td>1517</td>
<td>6.98%</td>
</tr>
<tr>
<td>lhipp</td>
<td>3165</td>
<td>2971</td>
<td>6.12%</td>
</tr>
<tr>
<td>lamyg</td>
<td>1299</td>
<td>1212</td>
<td>6.72%</td>
</tr>
<tr>
<td>rvent</td>
<td>20360</td>
<td>20887</td>
<td>2.56%</td>
</tr>
<tr>
<td>rthal</td>
<td>6214</td>
<td>6033</td>
<td>2.92%</td>
</tr>
<tr>
<td>rcaud</td>
<td>3454</td>
<td>2964</td>
<td>14.21%</td>
</tr>
<tr>
<td>rputa</td>
<td>4608</td>
<td>4453</td>
<td>3.37%</td>
</tr>
<tr>
<td>rpall</td>
<td>1655</td>
<td>1513</td>
<td>8.61%</td>
</tr>
<tr>
<td>rhipp</td>
<td>3312</td>
<td>3093</td>
<td>6.60%</td>
</tr>
<tr>
<td>ramyg</td>
<td>1318</td>
<td>1228</td>
<td>6.88%</td>
</tr>
</tbody>
</table>
Figure 5.2: This figure shows the mean and the standard deviations of the Kappa overlaps between the segmentations from the denoising pipeline and the original FreeSurfer segmented volumes. Black and white bars respectively denote left and right structures. Key: vent – Ventricles, thal – Thalamus, caud – Caudate, puta – Putamen, pall – Pallidum, hipp – Hippocampus, amyg – Amygdala.

Figures 5.3 - 5.8, describing regionally-specific group surface differences between HC and AD, HC and MCI, as well as MCI and AD. The figures highlight vertices on the template surfaces at which significant compression or expansion is detected in the latter group at the significance level of $p = 0.05$. To be specific, Figure 5.3 - Figure 5.5 respectively show group shape differences detected in those structures near medial temporal regions – the amygdala and the hippocampus, as well as the lateral ventricle. Figure 5.6 shows the group differences in the basal ganglia regions as well as the thalamus. Figure 5.7 shows the relative shape variations of all the seven structures in the left hemisphere while Figure 5.8 shows shape differences in the right hemisphere.
Figure 5.3: Panels a)-b) respectively show the group shape differences between HC and AD, HC and MCI, as well as MCI and AD measured in the left and right amygdala. Both superior (left) and inferior (right) views are displayed for each group comparison. Negative color scale values indicate surface expansion in the latter group, and positive values indicate atrophy. The scale value quantifies the ratio of the local volume of the former group at a particular location to that of the latter group in logarithmic scale.
Figure 5.4: Panels a), b) demonstrate the group shape differences between HC and AD, HC and MCI, MCI and AD of the hippocampus in both hemispheres. Both superior (left) and inferior (right) views are displayed for each group comparison. Negative color scale values indicate surface expansion in the latter group, and positive values indicate atrophy. The scale value measures the ratio of the local volume of the former group to that of the latter group in logarithmic scale.
Figure 5.5: Group shape differences detected in the lateral ventricles in both hemispheres. Colors corresponding to negative values indicate local expansion of surfaces in the latter group whereas positive values imply atrophy, compared with the former group. The more negative the value, the more prominent the expansion. The scale value measures the ratio of the local volume of the former group to that of the latter group in logarithmic scale.
Comparisons of MCI-stable and MCI-AD

To compare the two subtypes of MCI at baseline: MCI that remained stable (MCI-stable) and MCI that deteriorated to AD (MCI-AD), we perform the same vertex-based statistical analysis for all the seven structures in both hemispheres. The vertices on the template surfaces revealing statistically significant group differences are shown in Figures 5.9 - 5.11. Comparing MCI-stable versus MCI-AD, we find group surface differences in three structures: the hippocampus, the amygdala, and the lateral ventricle. Little difference has been detected in other subcortical structures.
Figure 5.7: Group differences, between every two of the three groups, measured in the surfaces of all the seven structures in the left hemisphere. Warm color denotes regions where the corresponding structure has significant atrophy in the latter group when compared with the former group. Cool color suggests local expansion of the structure in the latter group when compared with the former group. Key: am – amygdala, hi – hippocampus, vent – lateral ventricle, thal – thalamus, puta – putamen, pal – globus pallidus, caud – caudate. The scale value measures the ratio of the local volume around a particular vertex of the former group to that of the latter group in logarithmic scale.
Figure 5.8: Group shape variations of all the seven structure in the right hemisphere. Colors corresponding to positive values indicate regions on the structure where there is significant atrophy in the latter group when compared with the former group. Colors corresponding to the negative values suggest local expansion on the structure in the latter group, compared with the former one. Key: am – amygdala, hi – hippocampus, vent – lateral ventricle, thal – thalamus, puta – putamen, pal – globus pallidus, caud – caudate. The scale value measures the ratio of the local volume of the former group to that of the latter group in the logarithmic scale.
Figure 5.9: Panels a)-b) display group differences, detected in the left and right amygdala, between the two subtypes: MCI-stable and MCI-AD. Positive values indicate atrophy in MCI-AD while negative indicates expansion as compared with MCI-stable. The scale value measures the ratio of the local volume around each vertex on the template surface of the MCI-stable group to that of the MCI-AD group in logarithmic scale.

Figure 5.10: Hippocampal shape differences between MCI-stable and MCI-AD. Warm color indicates regions on the hippocampus where significant atrophy has been detected in MCI-AD when compared with MCI-stable, whereas cool color indicates expansion. The scale value measures the ratio of the local volume of the hippocampus in MCI-stable to that in MCI-AD in the logarithmic scale.
Figure 5.11: Shape differences between MCI-stable and MCI-AD detected on lateral ventricles in both hemispheres. The color corresponding to value 0 suggests no regional shape variation. Colors corresponding to negative values suggest expanding regions in MCI-AD when compared with MCI-stable, while those corresponding to positive values indicate atrophy in MCI-AD. The scale value measures the ratio of the local volume of the lateral ventricle in MCI-stable to that in MCI-AD in logarithmic scale.

5.5 Conclusion

In this chapter, we presented a surface-based morphometric (SBM) analysis framework for detecting regional shape differences across various clinical groups. The SBM pipeline was first validated on a simulation study by constructing two sets of surfaces that vary in length but not thickness. The SBM pipeline was shown to be capable of detecting the elongated tails in the group with longer length. We then applied the SBM analysis pipeline to detect and quantify the subcortical and ventricular structural changes in mild cognitive impairment (MCI) and Alzheimer’s disease (AD) patients.

SBM analysis offers an alternative to volume-based analysis for detecting and quantifying abnormalities of deep subcortical and ventricular structures in dementia.
of the Alzheimer type. An analysis of the surfaces enclosing the brain structures allows for the characterization of shape abnormalities that are associated with MCI or AD, which do not necessarily involve changes in the overall size of the structures. In addition to measuring group differences in subjects with MCI or AD compared with HC ones, shape-based analysis may also help in identifying MCI individuals who are suffering from prodromal AD. In our study, we examined all the subcortical structures and the lateral ventricle. The shape abnormality of a single structure may be nonspecific to AD. Thus, a combination of neuroimaging measures from multiple structures may be more sensitive and specific to AD pathology, helping to identify abnormalities in MCI and AD and predict conversion from MCI to AD more accurately.

Our study revealed significant atrophy in the hippocampus and the amygdala and prominent expansion in the lateral ventricle, in both hemispheres, in MCI as well as AD groups, which suggested that histopathological changes occurred before being defined as AD clinically. In addition to structural abnormalities near the medial temporal area and the lateral ventricle, we also found mild regional atrophy of basal ganglia structures such as the left and right putamen and globus pallidus in MCI and AD groups. Both atrophy and expansion has been detected on some vertices of the surfaces of the thalamus at each side in MCI and AD, compared with HC. Mild expansion was also found in caudate nucleus in AD when compared with normal aging subjects. Since the number of vertices showing expansion on the caudate surface is very small, this may not influence the overall volume size, which makes it difficult
for this change to be measured according to the overall structure volume. Also, the expansion in either the caudate or the thalamus may be due to inaccurate automated volumetric segmentations or partial volume effects.

The purpose of comparing the baseline measures of the sub-groups of MCI is to help assess whether a person diagnosed with MCI has underlying AD pathophysiology based only on baseline information. In our datasets, among the subjects that had been diagnosed as MCI at baseline, 205 were diagnosed to have remained MCI (MCI-stable), while another 151 converted to AD (MCI-AD) over a fixed follow-up time of one year. Unlike the significant differences observed among HC, MCI, and AD groups in the basal ganglia and the thalamus, the MCI subgroups did not significantly differ in these structures. However, compared with MCI individuals that did not deteriorate to AD, atrophy in both hippocampus and amygdala and regional expansion of the lateral ventricle in both hemispheres was detected in MCI-AD.
Chapter 6

Statistical Shape Analysis of Longitudinal Neuroimage Data

6.1 Introduction

Longitudinal neuroimaging studies have become more and more widely used in the analysis of various neuropsychiatric diseases during the last decade [137, 138]. The advantages of longitudinal studies are two-fold. First, longitudinal analysis increases the statistical power of the study by reducing the confounding effect of inter-subject variability. In the second place, longitudinal studies unveil the temporal dynamics of different underlying biological processes.

There are typically two types of statistical analyses in longitudinal studies. The first one is the cross-sectional analysis, comparing different groups at one specific time
point. The second one focuses on the analysis of temporal dynamics, studying the
temporal evolution path of a single group of interest. In addition to these two types of
studies, we can also combine the cross-sectional analysis with the temporal evolution
analysis. Namely, we can compare different groups in terms of their temporal evolu-
tion dynamics. For example, based on longitudinal studies, it has been suggested that
the hippocampus atrophies over time in both normal aging and subjects with AD. We
can thus compare the two groups in terms of the atrophy rates of the hippocampus,
which is a cross-sectional comparison of the temporal dynamics of the hippocampus.

In Chapter 5, we studied the baseline shape abnormalities in seven subcortical
and ventricular structures in subjects with MCI and AD. This belongs to the realm
of the first type of analyses, cross-sectional analyses. In this chapter, we extend the
previously described surface based morphometric analysis pipeline to longitudinal
shape analysis, in which a single subject receives multiple scans over a period of
time. The longitudinal SBM analysis pipeline is again applied to the ADNI dataset,
examining the group differences of the temporal dynamics in terms of regional shape
change rates of the bilateral amygdala, hippocampus, and ventricle in MCI and AD
populations, compared to HC.
6.2 Method

In the longitudinal design, each subject has associated to it multiple scans (and thus multiple surfaces per structure) obtained at different time points. For convenience, for each subject, we denote the scan that was acquired the earliest as time 0, the next scan as time 1, and so on. The interval between a scan and its previous one is the same for different subjects. In this study, we included data from ADNI subjects with at least one follow-up scan, resulting in a total of 203 HC subjects, 343 MCI subjects, and 167 subjects with AD. Subjects were followed for up to 36 months, with a 6- or 12- month interval between every two sequential scans. The total number of available scans varies from subject to subject.

Figure 6.1 is a schematic of the longitudinal analysis pipeline for characterizing the differences in localized shape change rates, of each single structure, between different clinical groups. To measure the localized shape change rate at each vertex, for each subject, all follow-up surfaces are individually mapped to their baseline surfaces via an LDDMM-surface mapping. In this way, subject-specific deformation markers can be computed for each follow-up scan, measuring the regional surface changes along time. After that, for each subject, its baseline surface is mapped to the common coordinate system shared by all populations.
Figure 6.1: Demonstration of the longitudinal shape analysis pipeline. All statistical analyses are performed in a common registered coordinate system.

6.2.1 Longitudinal SBM Analysis

The morphometric shape feature in the longitudinal SBM analysis is again represented by the deformation marker $J$ that has been introduced in section 5.2.3. Let $J_k(s,j)$ be the deformation marker associated with subject $s$ at time point $j$. It is indexed at each vertex $k$ of the common template surface. To comparing two different groups, introduce a binary group variable $Y(s)$. The statistical model is given by

$$J_k(s,j) = \alpha_{k,0} + \alpha_{k,1} t(s,j) + (\beta_{k,0} + \beta_{k,1} t(s,j)) Y(s) + \sum_{\text{cov}} \alpha_{\text{cov}} X_{\text{cov}}(s) + \varepsilon_k(s,j), \quad (6.1)$$
where $X_{cov}(s)$ is the covariate information included in analysis. The quantity $t(s, j)$ denotes the age of subject $s$ at time point $j$. The noise structure $\varepsilon_k(s, j)$ is modeled using a mixed-effect model, the sum of two Gaussian processes

$$\varepsilon_k(s, j) = \eta_k(s) + \zeta_k(s, j), \quad (6.2)$$

with $\eta_k(s) \sim N(0, \rho \sigma_k^2)$ and $\zeta_k(s, j) \sim N(0, \sigma_k^2)$. The first one $\eta_k(s)$ is a random effect that models the intersubject variation while the second one $\zeta_k(s, j)$ models the intrasubject variation within a series. The parameter $\rho$ is assumed to be independent of vertex $k$ for the purpose of simplicity and computational efficiency. The parameters $(\alpha_{k,0}, \alpha_{k,1}, \beta_{k,0}, \beta_{k,1}, \alpha_{cov}, \sigma_k^2, \ k = 1, 2, \ldots)$ and $\rho$ are estimated via maximum-likelihood estimation.

Hypothesis testing is performed at each vertex $k$ with the null hypothesis being $\beta_{k,0} = \beta_{k,1} = 0$. The test statistic $F_k$ is defined at each vertex $k$ of the template surface, through

$$F_k = L_{H_1}(k) - L_{H_0}(k), \quad (6.3)$$

where $L_{H_0}(\cdot)$ is the log-likelihood under the null hypothesis, and $L_{H_1}(\cdot)$ is the log-likelihood under the general hypothesis.

**Statement 6.1** *The log-likelihood function of the parameters in the statistical model shown in Eq. (6.1) is computed as*

$$-2L(k) = cst + n_{subj}\log \sigma_k^2 + \sum_s \log(\rho n_s + 1), \quad (6.4)$$
where \( n_{subj} \) is the total number of all subjects in the study, \( n_s \) is the total number of scans for subject \( s \), and \( cst \) is a constant.

**Proof.** The statistical model in Eq. (6.1) can be written as

\[
J_k(s, j) = \sum_a \beta_{ka} x_{ak}(s, j) + \varepsilon_k(s, j). \tag{6.5}
\]

Define the following notations:

\[
x_k(s) = \begin{bmatrix}
x_{1k}(s, 1) & x_{2k}(s, 1) & \cdots & x_{pk}(s, 1) \\
x_{1k}(s, 2) & x_{2k}(s, 2) & \cdots & x_{pk}(s, 2) \\
\vdots & \vdots & \ddots & \vdots \\
x_{1k}(s, n_s) & x_{2k}(s, n_s) & \cdots & x_{pk}(s, n_s)
\end{bmatrix}_{n_s \times p}
\]

\[
J_k(s) = \begin{bmatrix}
J_k(s, 1) & J_k(s, 2) & \cdots & J_k(s, n_s)
\end{bmatrix}_{n_s \times 1}
\]

\[
\beta_k = \begin{bmatrix}
\beta_{k1} & \beta_{k2} & & \beta_{kp}
\end{bmatrix}_{1 \times p}
\]

\[
\varepsilon_k(s) = \begin{bmatrix}
\varepsilon_k(s, 1) & \varepsilon_k(s, 2) & \cdots & \varepsilon_k(s, n_s)
\end{bmatrix}_{n_s \times 1}
\]

From Eq. (6.5), we thus have

\[
J_k(s) = x_k(s) \beta_k^T + \varepsilon_k(s). \tag{6.6}
\]

In our statistical model, the noise structure is modeled as a multivariate normal distribution \( N(\mu, \Sigma) \), the probability density function of which is defined as

\[
f_X(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right). \tag{6.7}
\]
In our case, $\mu = 0$, $\forall k,s$. From Eq. (6.7), we calculate the log-likelihood $L_k(s)$ as

$$L_k(s) = \frac{-1}{2} (J_k(s) - X_k(s)\beta_k^T)(\Sigma_k(s))^{-1} (J_k(s) - X_k(s)\beta_k^T)$$

$$- \frac{1}{2} \log |\Sigma_k(s)| - \frac{n_s}{2} \log 2\pi. \quad (6.8)$$

Therefore,

$$-2L = \sum_k \sum_s (J_k(s) - X_k(s)\beta_k^T)^T(\Sigma_k(s))^{-1} (J_k(s) - X_k(s)\beta_k^T)$$

$$+ \sum_k \sum_s \log |\Sigma_k(s)| + \sum_k \sum_s n_s \log 2\pi. \quad (6.9)$$

Now, we describe the calculation of $\Sigma_k(s)$.

$$\Sigma_k(s) = \begin{bmatrix}
E(\varepsilon_k(s,1)^2) & E(\varepsilon_k(s,1)\varepsilon_k(s,2)) & \cdots & E(\varepsilon_k(s,1)\varepsilon_k(s,n_s)) \\
E(\varepsilon_k(s,2)\varepsilon_k(s,1)) & E(\varepsilon_k(s,2)^2) & \cdots & E(\varepsilon_k(s,2)\varepsilon_k(s,n_s)) \\
\vdots & \vdots & \ddots & \vdots \\
E(\varepsilon_k(s,n_s)\varepsilon_k(s,1)) & E(\varepsilon_k(s,n_s)\varepsilon_k(s,2)) & \cdots & E(\varepsilon_k(s,n_s)^2)
\end{bmatrix}$$

$$= \sigma_k^2 (I + \rho 11^T), \quad (6.10)$$

where $E(\cdot)$ denotes the expectation function, $I$ denotes the identity matrix of size $n_s \times n_s$, and $1 = \begin{bmatrix} 1 & 1 & \ldots & 1 \end{bmatrix}_{n_s \times 1}$.

Assume that

$$(I + \rho 11^T)^{-1} = I + c11^T, \quad (6.11)$$

for some constant $c$, then

$$I + \rho 11^T + c11^T + \rho cn_s 11^T = I \Rightarrow c = -\frac{\rho}{1 + \rho n_s}. \quad (6.12)$$
Therefore, we have

\[
(\Sigma_k(s))^{-1} = \frac{1}{\sigma_k^2} \left( I - \frac{\rho}{1 + \rho n_s} \mathbf{1} \mathbf{1}^T \right).
\]  

(6.13)

Calculating the eigenvalues of \(\Sigma_k(s)\), we get

\[
|\Sigma_k(s)| = (\rho n_s + 1)\sigma_k^2,
\]

(6.14)

and thus

\[
\sum_k \sum_s \log |\Sigma_k(s)| = n_{subj} \left( \sum_k \log \sigma_k^2 \right) + \sum_k \sum_s \log(\rho n_s + 1),
\]

(6.15)

given that \(\sum_s 1 = n_{subj}\).

Therefore

\[
-2L = \sum_k \sum_s (J_k(s) - X_k(s)\beta_k^T)^T \frac{1}{\sigma_k^2} \left( I - \frac{\rho}{1 + \rho n_s} \mathbf{1} \mathbf{1}^T \right) (J_k(s) - X_k(s)\beta_k^T)
\]

\[+ n_{subj} \left( \sum_k \log \sigma_k^2 \right) + \sum_k \sum_s \log(\rho n_s + 1) + \text{cst}.
\]

(6.16)

To get the maximum-likelihood estimators (MLEs) for the parameters, we first fix \(\rho\) and find the \(\hat{\beta}_k\) that minimizes the first term on the right hand side of Eq. (6.16).

The MLEs for \(\beta_k\) and \(\sigma_k^2\) are

\[
\hat{\beta}_k = \left[ \sum_s \left( X_k(s)^T (I - \frac{\rho \mathbf{1} \mathbf{1}^T}{1 + \rho n_s} X_k(s)) \right) \right]^{-1} \left[ \sum_s \left( J_k(s)^T (I - \frac{\rho}{1 + \rho n_s} \mathbf{1} \mathbf{1}^T) X_k(s) \right) \right],
\]

(6.17)

and

\[
\hat{\sigma}_k^2 = \frac{1}{n_{subj}} \sum_s \left( J_k(s) - X_k(s)\hat{\beta}_k^T \right)^T \left( I - \frac{\rho}{1 + \rho n_s} \mathbf{1} \mathbf{1}^T \right) \left( J_k(s) - X_k(s)\hat{\beta}_k^T \right).
\]

(6.18)
After that, we fix $\hat{\beta}_k$ and $\hat{\sigma}^2_k$. Differentiating the right hand side of Eq. (6.16) with respect to $\rho$ to get the MLE for $\rho$. Substituting the parameters with their MLEs, we have
\[
-2\hat{L} = \sum_k n_{subj} + n_{subj} \left( \sum_k \log \sigma^2_k \right) + \sum_k \sum_s \log(\rho n_s + 1) + \text{cst.} \tag{6.19}
\]
Therefore, the conclusion in the statement is correct (see Eq(6.4)).

QED.

### 6.2.2 Statistical Model in Examining Shape Change Rates

To examine the group differences in the temporal dynamics of regional shape, we measure the locally computed shape change rate. Driven by the model proposed in Eq. (6.1), we use the following statistical model in comparing the regional shape change rates between two groups

\[
\exp \{ \Delta J_k(s, j) \} - 1 = \alpha_{k,1} (\Delta t(s, j)) + \beta_{k,1} (\Delta t(s, j)) Y(s) + \varepsilon_k(s, j), \tag{6.20}
\]
for $j = 2, 3, ..., n_s$. In Eq. (6.20), $\Delta J_k(s, j) = J_k(s, j) - J_k(s, 1)$, $\Delta t(s, j) = t(s, j) - t(s, 1)$.

Statistical significance is again quantified based on Fisher’s method of randomization via Monte Carlo simulations, as we described in section 5.2.3. The familywise
error rate (FWER) is controlled at a level of 0.05 using the same procedure described in section 5.2.3. The group difference in terms of regional shape change rates is quantified as $-\beta_{k,1}$.

### 6.3 Results

#### 6.3.1 Three-structure Analysis

Significant group differences between every two of the three groups (HC, MCI, and AD), in terms of the annualized regional shape change rates of the amygdala, the hippocampus, and the ventricle, are demonstrated in Figure 6.2. This figure highlights vertices on the template surfaces at which a significantly smaller or larger shape change rate was detected in the latter group relative to the former at the significance level of $p = 0.05$ after multiple comparison correction. For the amygdala and the hippocampus, positive values suggest larger atrophy rates while negative values suggest smaller atrophy rates in the latter group when compared to the former group in comparison. For the ventricle, a negative value represents a greater expansion rate and a positive value represents a smaller expansion rate in the latter group relative to the former group.

According to our analysis, the atrophy of the amygdala and the hippocampus, as well as the ventricular expansion, progresses at different local rates and with different spatial patterns depending on the subject’s grouping. As shown in Figure
Figure 6.2: Illustrations of the regional shape change rate differences, located on the three structures (amygdala, hippocampus, and ventricle), in both hemispheres, between every two of the three groups (HC, MCI, and AD). For the amygdala and the hippocampus, colors with positive values represent larger atrophy rates whereas colors with negative values represent smaller atrophy rates in the latter group relative to the former. For the ventricle, colors with negative values indicate larger expansion rates while those with positive values indicate smaller expansion rates in the latter group when compared with the former group.
6.2, compared with HC, both MCI and AD subjects have higher, but non-uniform, shape change rates in certain regions of the three structure surfaces. At any specific time point, comparing the rate of shape change generally orders the populations as AD ≥ MCI ≥ HC with the localized rate difference varying from vertex to vertex on the template surface.

### 6.3.2 Subregion Analysis

Figure 6.3, Figure 6.4 and Figure 6.5 give maps of the pairwise group differences in localized atrophy rates, on the whole structure surface and the restriction to each of its four subregions, for the left hippocampus, right hippocampus and left amygdala respectively. The atrophy rate difference maps restricted to each subregion of the left and right hippocampus, which respectively comprise Figure 6.3 and Figure 6.4, indicate that the largest hippocampal atrophy rate difference, in each group comparison, occurs at CA1, followed by the subiculum and CA2, whereas the smallest occurs at the compartment containing CA3 and the dentate gyrus. In each subregion, the observed spatial pattern of atrophy rate difference was similar for each group comparison. In the left amygdala subregion analysis, similar spatial shape atrophy patterns were observed for all group comparisons of each compartment, as illustrated in Figure 6.5. Vertices showing the largest atrophy rate differences were found to belong to the basolateral compartment in the HC versus AD comparison. This subregion analysis revealed that, in each group compar-
ison for the left amygdala, the atrophy rate differences are in the following order
basolateral $>$ lateral nucleus $\geq$ basomedial $>$ centromedial.

Figure 6.3: Spatial maps of the atrophy rate differences of the left hippocampus, between every group pair (out of HC, MCI, and AD), for the entire surface (the first column from left) and the restrictions to the CA1 zone (the second column from left), the CA2 zone (the middle column), the CA3/dentate gyrus zone (the fourth column from left), and the subiculum zone (the fifth column from left). Colors with positive values indicate greater atrophy in the latter group relative to the former group.

6.4 Conclusion

We have extended the surface based morphometric analysis pipeline to accommodate for longitudinal analysis. Two types of analyses have been proposed: first, cross-sectional analysis of local shape volume/area differences between two groups at a specific time point; second, cross-sectional analysis of localized shape change rates
Figure 6.4: Spatial maps of the atrophy rate differences of the right hippocampus, between every group pair (out of HC, MCI, and AD), for the entire surface (the first column from left) and the restrictions to the CA1 zone (the second column from left), the CA2 zone (the middle column), the CA3/dentate gyrus zone (the fourth column from left), and the subiculum zone (the fifth column from left). Colors with positive values indicate greater atrophy in the latter group relative to the former group.
between two groups over time. The group differences in the first type are measured by $-(\beta_{k,0} + \beta_{k,1} \times \text{age})$, where age denotes the average age of the population at that specific time point, and $\beta_{k,0}, \beta_{k,1}$ are computed from Eq. (6.1). The group differences, in terms of the regional shape change rates (atrophy rates or expansion rates around each vertex of the template surface), are quantified by $-\beta_{k,1}$ from Eq. (6.20).

The longitudinal SBM statistical model was applied to the ADNI dataset to evaluate the group differences, in terms of localized shape change rates, in the memory-related amygdala-hippocampus circuitry as well as the surrounding ventricles. The proposed statistical model is capable of detecting and quantifying the abnormal shape change rates in dementia groups with MCI or AD. Facilitated by the high-field 7 Tesla
subregion segmentations of the hippocampus in both hemispheres and the left amygdala, we evaluated the atrophy processes restricted to each subregion compartment. This allowed for a detailed characterization of the local temporal progression dynamics of those three structures and the local comparison of those dynamics in the three populations. The proposed method not only helps to identify how the disease progresses over time but also reveals the spatial spreading pattern within each structure. The results presented in this study, obtained from a large sample of subjects, strongly suggest that this regional shape change rate method could be utilized as a diagnostic and prognostic tool in future clinical trials in Alzheimer dementia.
Chapter 7

Disease Detection and Prediction using Shape Deformation Patterns

7.1 Introduction

After detecting and quantifying the shape abnormalities of certain brain structures in populations with a specific disorder, the next natural step is to design algorithms that can automatically detect and predict the disease status based on the shape abnormality features. The procedure is comprised of three steps. The first step is to identify features or patterns that are capable of discriminating the disease group from the healthy control group. In the realm of CA, these features are usually related with various anatomical manifolds such as the cortical thickness, the volumetric measurement of certain structures, and the shape patterns of certain regions, whereas in our
case we use the shape deformation patterns of subcortical and ventricular structures. The shape deformation patterns are of high dimension (the features are indexed at each vertex on a template surface for each structure) while the number of participants is relatively small. Considering the complexity of the brain structure and the high dimensionality of the feature data, the second step is to perform dimension reduction. After that, the last step is to perform automated disease classification via supervised or unsupervised machine learning.

In previous chapters (Chapter 5 and Chapter 6), we identified and measured the regional shape abnormalities in diseased populations via an examination of the shape deformation patterns. In the framework of LDDMM, given a fixed template, the anatomical variability in the targets is encoded by the geodesics from the template to each target. The fundamental conservation of momentum property of these geodesics [80] allows for representing the entire flow of a geodesic by the initial momentum configuration. Since the geodesic flow at any point is completely determined by the momentum at the origin, this means that, once a template is fixed, the space of initial momenta becomes an appropriate linear vector space [51] for studying shape. Anatomical differences among different target groups can, therefore, be studied by applying linear statistical analysis such as principal component analysis (PCA) to the initial momentum vectors. PCA followed by linear discriminant analysis (LDA) on the initial momentum may be able to provide a shape-associated biomarker to discriminate between different clinical groups.
In this chapter, we propose a classification procedure based on the shape deformation patterns created in LDDMM-surface mapping of the fourteen subcortical structures – the left and right amygdala, hippocampus, caudate, putamen, globus pallidus, thalamus, and ventricle. The pipeline is then applied to the discrimination of subjects with AD from the healthy controls using a leave-one-out (LOO) cross-validation procedure. In addition, we apply those shape deformation patterns to predict the conversion from MCI to AD.

7.2 Discrimination between HC and AD via Shape Deformation Patterns

7.2.1 Principal Component Analysis

It has been proved that the optimal momentum, \( \alpha_t \), the solution of Eq. (5.5) satisfies a conservation property: the initial momentum \( \alpha_0 \) encodes the geodesic connecting the template surface to the target surface via

\[
\alpha_t^i = (D\phi_t(x^i))^{-T} \alpha_0^i,
\]

(7.1)

where \((\cdot)^{-T}\) denotes matrix inverse and then transpose, \(i\) denotes the \(i\)-th vertex on the template surface, and \(\phi_t\) is the diffeomorphism associated to \(v_t\) in Eq. (5.3). The deformation from the LDDMM-surface mappings is completely encoded by the initial
momentum \( \alpha_0 \) in the template surface coordinates. Therefore, we select the initial momentum vectors to be our features. These vectors form a very high dimensional space because, for each subject, \( \alpha_0 \) is of dimension \( N \times 3 \), where \( N \) denotes the number of vertices on the template surface.

Let \( N_{HC} \) be the number of HC subjects and \( N_{AD} \) be the number of AD subjects. For each structure, arranging the initial momentum vectors \( \alpha_0 \) for all subjects, both HC and AD, into a common matrix yields

\[
M_0 = [\alpha_{0,1}, \alpha_{0,2}, \ldots, \alpha_{0,N_{HC}+N_{AD}}]^T \quad (\in \mathbb{R}^{(N_{HC}+N_{AD}) \times 3N}).
\]

We then perform principal component analysis (PCA) on \( M_0 \) to construct an orthonormal basis. It is worth noting that in our context, the inner product used in the PCA is derived from the Riemannian metric that leads to the geodesic equation in LDDMM. Details of PCA on the initial momentum space can be found in [51]. The feature space constructed via the initial momentum vectors is then linearly projected to the orthogonal directions that carry the greatest shape variance.

To test whether there exist group differences in our initial momentum features, we perform a nonparametric statistical test on the first \( M \) (\( M < N_{HC} + N_{AD} \)) coefficients in the principal component basis (PCs) that account for 95% of the total variance. Let \( \tilde{Z}^{HC} \) and \( \tilde{Z}^{AD} \) be the sample means of the first \( M \) principal component values for the two groups, and \( \tilde{\Sigma} \) be the pooled sample covariance. To test the null hypothesis \( H_0 : \tilde{Z}^{HC} = \tilde{Z}^{AD} \), we compute the Hotelling’s statistic as:

\[
T^2 = \frac{N_{HC} \times N_{AD}}{N_{HC} + N_{AD}} (\tilde{Z}^{HC} - \tilde{Z}^{AD})^T \tilde{\Sigma}^{-1} (\tilde{Z}^{HC} - \tilde{Z}^{AD}).
\]  

(7.2)
Similar to the vertex-based analysis described in section 5.2.3 and 6.2.2, we use Monte Carlo simulations to generate random permutations to correct the p-values. We also perform the same statistical tests on the structure volumetric measurements (replacing the principal component values with the structure volume measurements) for comparison purposes.

In our PCA, the value, $M$, is 21 for left amygdala, 24 for right amygdala, 48 for left hippocampus, 49 for right hippocampus, 44 for left ventricle, 40 for right ventricle, 45 for left caudate, 45 for right caudate, 57 for left putamen, 59 for right putamen, 61 for left thalamus, 60 for right thalamus, 30 for left pallidum, and 29 for right pallidum. The value $M$ is related to the variability of the structure in the general population. The fact that this value varies from structure to structure may indicate that some structures are more stable than others, in the sense that the variance of those structures with small $M$ is concentrated in a small number of components.

Results obtained from the PCA-based analyses, on the ADNI dataset, are in strong agreement with those found in the vertex-based analyses shown in Figure 5.7 and Figure 5.8. In Figure 7.1, we plot the empirical distributions of the amygdala and the hippocampus, in both hemispheres, from randomized Hotelling's $T^2$ tests with 40,000 group permutations, between the two groups. The p-values are calculated from the random permutation tests. All three comparisons show group differences in the amygdala and the hippocampus, both left and right, as revealed by the p-values. The upper bounds of the confidence intervals for all the p-values, obtained from the
PCs and the structure volumes, are listed in Table 7.1.

Figure 7.1: PCA of the initial momentum matrices of the bilateral amygdala and hippocampus. Each sub-figure shows permutation test results for a structure’s group comparison based on the first $M$ PCs that account for 95% of the total variability. Shown are: 1) value (solid blue line) of each group comparison (total of three comparisons); 2) $p = 0.0001$ (red dotted line), $p = 0.001$ (blue dotted line), and $p = 0.05$ (black dotted line) for reference; 3) p-value derived from the 40000 permutation tests (solid green line).

7.2.2 Linear Discriminant Analysis

Via non-parametric statistical tests on each set of PCs, we find that group differences between AD and HC exist in a majority of structures. It is natural to try to discriminate between the two groups based on those PC features showing group differences, and we choose linear discriminant analysis (LDA) to do so. In this framework, for the two-class problem, the discriminating direction is the projection of the differences between the two class means onto the common covariance, yielding
Table 7.1: A comparison of the upper bounds of the confidence intervals for p-values obtained from the volume analysis and the PCA based shape analysis for each structure. Abbreviations: lvent - left lateral ventricle, ltha - left thalamus, lcaud - left caudate, lputa - left putamen, lpall - left globus pallidus, lhipp - left hippocampus, lamyg - left amygdala, rvent - right lateral ventricle, rtha - right thalamus, rcaud - right caudate, rputa - right putamen, rpall - right globus pallidus, rhipp - right hippocampus, ramyg - right amygdala.

<table>
<thead>
<tr>
<th></th>
<th>lamyg</th>
<th>lhipp</th>
<th>lvent</th>
<th>lcaud</th>
<th>lputa</th>
<th>lthal</th>
<th>lpall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>volume analysis</strong></td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 0.46</td>
<td>p &lt; 0.02</td>
<td>p &lt; 0.07</td>
<td>p &lt; 0.08</td>
</tr>
<tr>
<td><strong>shape analysis</strong></td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
</tr>
<tr>
<td></td>
<td>ramyg</td>
<td>rhipp</td>
<td>rvent</td>
<td>rcaud</td>
<td>rputa</td>
<td>rthal</td>
<td>rpall</td>
</tr>
<tr>
<td><strong>volume analysis</strong></td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 0.89</td>
<td>p &lt; 0.04</td>
<td>p &lt; 0.05</td>
<td>p &lt; 0.45</td>
</tr>
<tr>
<td><strong>shape analysis</strong></td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
<td>p &lt; 1E-4</td>
</tr>
</tbody>
</table>

\[ K^{-1}(\mu^{HC} - \mu^{AD}), \] where \( K \) is the common covariance of the two populations, and \( \mu^{HC}, \mu^{AD} \) are the two class means.

In terms of discriminant analysis, we have several different goals: 1) evaluate the performance of the LDA classifiers, trained on the PCs of each single structure, to determine which structure has the most discriminating shape information; 2) determine whether combining the shape information of multiple structures will strengthen the classification effect; 3) determine whether increasing the number of structures we use would improve the accuracy; 4) determine the best LDA classifier we can build based on the shape information of the seven structures; 5) estimate the true classification rate we would obtain from our classification procedure.

To reduce the dimension of the feature space, we select only the PCs that show significant group differences between HC and AD. For each PC, we perform the same
non-parametric statistical test as described in the PCA analysis and select those PCs
with a p-value obtained in the permutation test of less than 0.05.

To fulfill our goals 1) – 4), we test all the possible LDA classifiers we could build
from the PCs of the seven structures (group the PCs of the left and right for the
same structure together). Considering each possible combination, we build different
classifiers and compare their classification performance with each other based on leave-
one-out cross validation. The procedure is demonstrated in Figure 7.2, and consists
of three steps. The first step is to create all the possible feature spaces by combining
the different sets of PCs. Since we have 7 sets of PCs, the result is a total of 127
feature spaces. The next step is to test the classification rate based on the feature
information in each combination. In this step, we adopt leave-one-out as the cross
validation procedure: leave one subject out and train an LDA classifier based on the
feature vectors of the remaining subjects, then use this LDA to classify the subject
excluded at the beginning. In the end, we select the LDA classifier with the highest
average correct classification rate in the previous step as our optimal LDA classifier.
To test whether shape information is more discriminating than volume information,
we perform the same procedure using only the volume information of the structures
and compare.

To estimate the true classification rate we would obtain from our classification
procedure, we use leave-one-out as our cross-validation strategy. The cross-validation
process is summarized in Figure 7.3. We exclude the initial momentum of one subject
Figure 7.2: This figure summarizes the procedure of selecting the optimal LDA classifier based on the PCs from all the fourteen structures, comprised of three steps. Step 1 is to create all possible combinations of PCs, resulting in a total of 127. Step 2 is to test the mean classification rate for each set of PCs, based on leave-one-out LDA. Step 3 is to select the LDA classifier with the highest mean correct classification rate in Step 2. Key: hi - Hippocampus, am - Amygdala, vl - Lateral Ventricle, th - Thalamus, pu - Putamen, pa - Pallidum, ca - Caudate.
at the very beginning, and then build a PCA basis, formed from the initial momentums of all the other remaining subjects, for every structure. Then, we take two steps to reduce the dimensionality of each feature matrix: 1) we select the first \( N \) components that account for a 95\% of the total variance; 2) among the \( N \) feature vectors we keep in 1), we sift for the ones with significant group difference, for which we use Student’s t-test. After that, we select the optimal LDA classifier via the procedure described in Figure 7.2. The optimal LDA classifier is then used to classify the subject removed at the very beginning. This procedure is done for each subject. To determine which structure, among all the seven, exhibits the most discriminating shape information, we test the classification rate using each single structure. For example, in examining the classification rate of the LDA classifier built from hippocampal shape information, we do not select the optimal LDA classifier for each subject. Instead, we use the LDA classifier built from the shape features of the hippocampus to classify the subject excluded at the very beginning and then take the average.

In testing the true classification rate as described in Figure 7.3, there will be different PCA bases for different left-out subjects since we do leave-one-out on PCA. Among all the 385 (210 HC subjects and 175 AD subjects) PCA bases that have been computed in that way, the numbers of selected PCs are the same for the left amygdala (21 PCs), right amygdala (23 PCs), left hippocampus (46 PCs), right hippocampus (47 PCs), left thalamus (58 PCs), right thalamus (57 PCs), left putamen (55 PCs), right putamen (56 PCs), left pallidum (29 PCs), right pallidum (28 PCs), and left
Figure 7.3: This figure summarizes the leave-one-out cross-validation procedure of testing the true classification rate that we would be able to yield using our procedure. For the left ventricle, 368 out of 385 PCA bases have 39 vectors (the mean value being 39.0442), while for the right ventricle, 384 out of 385 have 39 vectors (the mean value being 39.0078). Finally, for the right caudate, 303 out of 385 runs select 43 PCs, with mean value equal to 42.7870.

In terms of discriminating between the two groups, HC and AD, we find that the shape PC information associated with each individual structure is uniformly significantly more discriminating than the volume information of that structure – i.e. using shape PCs yields better classification accuracy than volume for every single structure. Generally, based on the shape information, the classification errors are reduced by more than 10% for each single structure. The two sets of classification results are listed in Table 7.2. In addition, Table 7.2 demonstrates that, among all the seven...
Table 7.2: A comparison of the specificity and the sensitivity obtained from the LDA classifiers built respectively from the volume information and the shape information for each structure.

<table>
<thead>
<tr>
<th></th>
<th>Specificity</th>
<th></th>
<th>Sensitivity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>volume info</td>
<td>shape info</td>
<td>volume info</td>
<td>shape info</td>
</tr>
<tr>
<td>Amyg</td>
<td>75%</td>
<td>80%</td>
<td>72%</td>
<td>78%</td>
</tr>
<tr>
<td>Hipp</td>
<td>76%</td>
<td>86%</td>
<td>74%</td>
<td>81%</td>
</tr>
<tr>
<td>Vent</td>
<td>73%</td>
<td>83%</td>
<td>53%</td>
<td>79%</td>
</tr>
<tr>
<td>Caud</td>
<td>47%</td>
<td>68%</td>
<td>58%</td>
<td>68%</td>
</tr>
<tr>
<td>Puta</td>
<td>53%</td>
<td>75%</td>
<td>58%</td>
<td>78%</td>
</tr>
<tr>
<td>Thal</td>
<td>49%</td>
<td>76%</td>
<td>58%</td>
<td>75%</td>
</tr>
<tr>
<td>Pall</td>
<td>51%</td>
<td>68%</td>
<td>54%</td>
<td>68%</td>
</tr>
</tbody>
</table>

structures, the hippocampus exhibits the highest discriminating ability.

Comparing the mean classification rates of all the LDA classifiers obtained from different combinations of sets of PCs, we find that, based on the initial momentum of all the subjects, the optimal LDA classifier comes from a combination of three structures – the hippocampus, the amygdala, and the ventricle. This is in agreement with our vertex-based analysis findings, as well as those in the non-parametric statistical analysis, where we again found that the strongest shape differences occurred at the hippocampus, the amygdala, and the lateral ventricle. In estimating the true classification rate that could be achieved using our classification pipeline, the optimal LDA classifiers may be different across iterations, since the process of selecting the optimal LDA classifier is embedded in the leave-one-out process. For example, when subject 1 is left out, the optimal LDA classifier comes from a combination of the hippocampus
and the amygdala while for subject 2, the optimal one comes from a combination of the hippocampus, the amygdala, and the ventricle. We therefore calculate the percentage of each structure being included in the optimal classifier. We find that among all optimal LDA classifiers, the hippocampus has been selected 88% of the time, the amygdala 83% of the time, the ventricle 71% of the time, the thalamus 45% of the time, the caudate 36% of the time, the putamen 37% of the time, and the pallidum 26% of the time. This, to some degree, confirms our conclusion that if we were to design a single classifier based on the information from all the subjects, the combination of the hippocampus, the amygdala, and the ventricle would likely yield an optimal LDA classifier. According to the leave-one-out cross-validation procedure, as described in Figure 7.2, the correct classification rates are: 88% for the HC group, 86% for the AD group, and 86% for the two groups together. In comparison, the correct classification rates using volume information, via the same procedure, are: 76% for the HC group, 75% for the AD group, and 75% for the two groups combined.

7.3 Predicting MCI-to-AD Conversion

7.3.1 Background

The cause and the mechanism of progression for most AD cases are still unknown. AD worsens during its progression until death and currently there are no treatments that can cure or reverse this progression. Definite diagnosis of AD can only be made
based on histopathologic evidence obtained from a biopsy or autopsy. MCI is a syndrome that is regarded as a risk state for dementia and is associated with an increased risk of progression to probable AD. More than half of the individuals with MCI deteriorate to dementia within 5 years at a rate of about 10% to 15% per year. Considerable heterogeneity exists among MCI patients: some convert to AD with varying progression rates, and others remain stable for a long period of time or even revert to normal cognitive status. The former is called MCI converters (MCI-C) and the latter is called MCI non-converters (MCI-NC).

The ability to identify an MCI patient’s risk of developing AD is important for clinical decision making and timing therapy. Structural neuroimaging measures have been shown to be sensitive to the degeneration that occurs in MCI and AD, which may provide robust biomarkers for predicting the conversion form MCI to AD. Methods of detecting MCI that represents prodromal AD would aid clinical practice by allowing attention to be focused on those with the highest risk of conversion. However, differentiating the MCI-C from the MCI-NC is very challenging [54], especially when using only the baseline information. This difficulty is due to the lag between brain atrophy and cognitive decline. During the last decade, there have been many methods developed, using structural imaging, to discriminate between MCI-C and MCI-NC [53,56,59].

In this experiment, the training process is performed on the baseline HC and AD subjects while the testing process is carried out on a subset of the baseline MCI
subjects. Such a design comes from the observation that the differences, in terms of the shape deformation patterns, between MCI-C and MCI-NC are similar to those detected between AD and HC, which has been shown in section 5.4.5. It has also been reported that other patterns of change, within the brain, of MCI-C are similar to those of AD while those of MCI-NC are similar to HC [139].

In this section, we first describe the procedure for generating the HC-AD classifier used in the prediction of MCI conversion, which is built from the shape deformation features of a subset of the fourteen structures from 210 HC subjects and 175 AD subjects, all of which come from the ADNI baseline dataset. We then evaluate the predictor on a total of 222 MCI subjects, including 135 MCI-C subjects and 87 MCI-NC subjects, the conversion of which is determined by a follow-up of 36 months.

### 7.3.2 Participant

In this experiment, we include data from 210 HC subjects, 222 subjects with MCI, and 175 subjects with AD. Within the MCI group, 135 subjects converted to AD (MCI-C) within a follow-up of 36 months. The conversion time within the MCI-C group is heterogeneous since an MCI patient may convert at any time over the course of 6 months to 3 years. In the case of this study, 17 subjects converted in 6 months, 40 in 12 months, 27 in 18 months, 34 in 24 months, and the other 17 in 36 months. For the group MCI-NC, we only include those MCI subjects that have been followed for at least 3 years, yielding a total of 87 subjects in MCI-NC.
7.3.3 Generation of Shape Deformation Features

Assume that we have a total of $N_{MCI}$ MCI subjects. For each structure, let $M_0^* \in \mathbb{R}^{N_{MCI} \times 3N}$ be the matrix comprising of the initial momentum vectors for all the MCI subjects. In the previous section, section 7.2, on discriminating HC and AD, we performed PCA on the matrix $M_0$, comprising of the initial momentum vectors for the HC and AD subjects, to get the first $M$ PCs that account for 95% of the total variance. Given these PCs, $M_0$ and $M_0^*$ are then linearly projected to the orthogonal directions that carry the greatest shape variance (the first $M$ principal components) to get $\tilde{M}_0 \in \mathbb{R}^{(N_{HC}+N_{AD}) \times M}$ and $\tilde{M}_0^* \in \mathbb{R}^{N_{MCI} \times M}$ respectively. We then keep only the PCs in $\tilde{M}_0 \in \mathbb{R}^{(N_{HC}+N_{AD}) \times M}$ that show significant group difference between HC and AD. For each PC, we perform a Student’s t-test between the PC coefficients for HC and those for AD, and select the PCs with a p-value less than 0.05. We also keep the corresponding PC coefficients in $\tilde{M}_0^* \in \mathbb{R}^{N_{MCI} \times M}$. Finally, we end up with $\hat{M}_0 \in \mathbb{R}^{(N_{HC}+N_{AD}) \times M_1}$, $M_1 \leq M$ as our training features from HC and AD and $\hat{M}_0^* \in \mathbb{R}^{N_{MCI} \times M_1}$ as our testing features from MCI.

7.3.4 Selecting the Optimal Predictor

After performing PCA on the initial momentum space for each structure, we get fourteen $\hat{M}_0 \in \mathbb{R}^{(N_{HC}+N_{AD}) \times M_1}$, $M_1 \leq M$ as the training features and correspondingly fourteen $\hat{M}_0^* \in \mathbb{R}^{N_{MCI} \times M_1}$ as the testing features. As we have shown in 7.2.2,
it is plausible that the combination of features from a subset of structures gives the best classification results. To find this optimal combination, we test all the possible classifiers we could build from the PCs of the fourteen structures. Again we use LDA to construct our predictors. Considering each possible combination, we build 16383 different classifiers and compare their classification performance with each other. The procedure to find the optimal LDA MCI-to-AD predictor is demonstrated in Figure 7.4.

![Flowchart](image)

Figure 7.4: Flowchart demonstrating the procedure of selecting the optimal linear discriminant analysis (LDA) classifier in classifying MCI-C from MCI-NC.

To compute the classification rate in terms of classifying MCI-C versus MCI-NC, we adopt LOO again as the cross-validation procedure. To prevent over-fitting, one MCI subject is excluded before we select the optimal LDA classifier. To be specific, we exclude one MCI subject at the beginning and then find the optimal LDA classifier based on the feature information from the other $N_{MCI} - 1$ MCI subjects as described
in the previous section. Then we use the optimal LDA classifier to classify the MCI subject excluded at the beginning. The LOO procedure is demonstrated in Figure 7.5. It is important to notice that we obtain a unique optimal classifier for each LOO test. Since the test subject has been removed from both the initial PCA and the process of selecting the optimal classifier, we avoid any bias or overestimation in predicting the conversion from MCI to AD.

Figure 7.5: Flowchart demonstrating the procedure of performing leave-one-out (LOO) cross-validation without being biased by the subject removed at the beginning. This process is repeated for each MCI subject.

7.3.5 MCI-to-AD Prediction Results

According to our LOO experiment, we obtain the same optimal LDA classifier after excluding each MCI subject, which suggests robustness of our approach. The optimal LDA classifier comes from a combination of PCs from six structures – the left hippocampus, left lateral ventricle, right thalamus, right caudate, left and right putamen, the PCs of which are obtained from the HC and AD initial momentum...
space. To be specific, 9 PCs of left hippocampus, 16 PCs of left lateral ventricle, 16 PCs of right thalamus, 7 PCs of right caudate, 9 PCs of left putamen, and 15 PCs of right putamen have been selected as the final shape feature used in the LDA classification.

Using the optimal predictor built from the baseline shape deformation information of HC and AD subjects, we are capable of achieving a total classification accuracy of 77.03%, a sensitivity of 78.52% (classifying MCI-C) and a specificity of 74.71% (classifying MCI-NC). The area under the receiving operating characteristic curve is found to be 76.51%.

Figure 7.6 shows the performance of the proposed LDA predictor in predicting MCI patients with different conversion times. To be specific, using the baseline shape deformation information alone, we are capable of correctly predicting 82.35% (14/17) for subjects converting in 6 months, 80% (32/40) for subjects converting in 12 months, 77.78% (21/27) for subjects converting in 18 months, 78.13% (25/32) for subjects converting in 24 months, and 73.68% (14/19) for subject converting in 36 months. Generally, the closer the MCI subject is to conversion, the more accurately the classifiers can predict that conversion.
7.4 Conclusion

We developed and validated a novel structural imaging based biomarker, for the discrimination between HC and AD, and also the prediction of conversion from MCI to AD, using shape deformation patterns from LDDMM. The patterns were extracted from a subset of subcortical structures and the lateral ventricles, based on a training dataset of 210 HC and 175 AD subjects. The pipeline was first validated using a leave-one-out cross-validation procedure to differentiate the AD subjects from the HC ones. In addition, the procedure was validated on a total of 135 progressive MCI subjects and 87 MCI subjects who remained stable after a follow-up of 3 years. The prediction accuracies on each subgroup of MCI-C, which were grouped according to the conversion time, were found to be related to the conversion time.
Appendix

**Statement** The iterations $W^{\text{old}} \leftarrow W^{\text{new}}$ updated via

$$
W^{\text{new}} = \arg\max_W E_{p(A, \varphi|W^{\text{old}}, I^D)} \left\{ \log p(W, I^D, A, \varphi)|I^D, W^{\text{old}} \right\}
$$

\[ = \arg\max_W \int \log p(W, I^D, A, \varphi) dp(A, \varphi|W^{\text{old}}, I^D) \]

is monotonically increasing in the log-likelihood as $\log p(I^D, W^{\text{new}}) \geq \log p(I^D, W^{\text{old}})$.

**Proof.** According to Bayes’ rule, we have

$$
p(I^D, W) = \frac{p(A, \varphi, I^D, W)}{p(A, \varphi|I^D, W)},
$$

and thus

$$
\log p(I^D, W) = \log p(A, \varphi, I^D, W) - \log p(A, \varphi|I^D, W).
$$

Introduce a distribution $q(A, \varphi|I^D)$, defined over the latent variable $A$ and $\varphi$, such that $\sum_A \int q(A, \varphi|I^D) d\varphi = 1$. Inserting $q(A, \varphi|I^D)$ into the above equation, we have

$$
\log p(I^D, W) = \sum_A \int q(A, \varphi|I^D) \log \left\{ \frac{p(A, \varphi, I^D, W)}{q(A, \varphi|I^D)} \right\} d\varphi
$$

$$
- \sum_A \int q(A, \varphi|I^D) \log \left\{ \frac{p(A, \varphi|I^D, W)}{q(A, \varphi|I^D)} \right\} d\varphi.
$$

For simplicity, let

$$
\log p(I^D, W) = L(q, W) + KL(q||p),
$$

where $L(q, W) = \sum_A \int q(A, \varphi|I^D) \log \left\{ \frac{p(A, \varphi, I^D, W)}{q(A, \varphi|I^D)} \right\} d\varphi$ and

$$
KL(q||p) = - \sum_A \int q(A, \varphi|I^D) \log \left\{ \frac{p(A, \varphi|I^D, W)}{q(A, \varphi|I^D)} \right\} d\varphi.
$$

The term $KL(q||p)$ is known.
as the Kullback-Leibler divergence and, according to Gibbs’ inequality, we have $KL(q||p) \geq 0$. Suppose that the current value of the parameter is $W^{old}$, viewing the EM algorithm as two alternating maximization steps:

**E-step:** Maximize $L(q, W^{old})$ with respect to $q(A, \varphi|I^{D})$ while holding $W^{old}$ fixed;

**M-step:** Fix $q(A, \varphi|I^{D})$ and maximize $L(q, W)$ with respect to $W$;

We now show that the maximizer $W$ obtained in the above two steps satisfy that

$$\log p(I^{D}, W^{new}) \geq \log p(I^{D}, W^{old}).$$

In the E-step, given that the value of $\log p(I^{D}, W^{old})$ does not depend on $q(A, \varphi|I^{D})$ and that $L(q, W^{old}) = \log p(I^{D}, W^{old}) - KL(q||p)$, the largest value of $L(q, W^{old})$ is thus obtained when $KL(q||p) = 0$. Therefore, in the E-step,

$$q(A, \varphi|I^{D}) = p(A, \varphi|I^{D}, W^{old})$$

and

$$\log p(I^{D}, W^{old}) = L(q, W^{old}).$$

Then, we have

$$\log p(I^{D}, W^{new}) - \log p(I^{D}, W^{old}) = L(q, W^{new}) - L(q, W^{old}) + KL(q||p).$$

Gibb’s inequality tells us that $KL(q||p) \geq 0$, so we can conclude that

$$\log p(I^{D}, W^{new}) - \log p(I^{D}, W^{old}) \geq L(q, W^{new}) - L(q, W^{old}).$$

In words, choosing $W^{new}$ to improve $L(q, W^{new})$ beyond $L(q, W^{old})$ will improve $\log p(I^{D}, W^{new})$ beyond $\log p(I^{D}, W^{old})$ at least as much.
Now, we proceed by showing that a maximization of $L(q, W)$ with respect to $W$ in the M-step is equivalent to maximizing $E_{p(A, \varphi|W^{old}, I^D)} \{ \log p(W, I^D, A, \varphi)|I^D, W^{old} \}$.

We have

\[
L(q, W) = \sum_A \int q(A, \varphi|I^D) \log \left\{ \frac{p(A, \varphi, I^D, W)}{q(A, \varphi|I^D)} \right\} d\varphi
\]

\[
= \sum_A \int p(A, \varphi|I^D, W^{old}) \log \left\{ \frac{p(A, \varphi, I^D, W)}{p(A, \varphi|I^D, W^{old})} \right\} d\varphi
\]

\[
= \sum_A \int p(A, \varphi|I^D, W^{old}) \log p(A, \varphi, I^D, W) d\varphi
\]

\[
- \sum_A \int p(A, \varphi|I^D, W^{old}) \log p(A, \varphi|I^D, W^{old}) d\varphi.
\]

Therefore

\[
W^{new} = \arg \max_W L(q, W)
\]

\[
= \arg \max_W \sum_A \int p(A, \varphi|I^D, W^{old}) \log p(A, \varphi, I^D, W) d\varphi
\]

\[
= \arg \max_W \int \log p(A, \varphi, I^D, W) dp(A, \varphi|I^D, W^{old})
\]

\[
= E_{p(A, \varphi|W^{old}, I^D)} \{ \log p(W, I^D, A, \varphi)|I^D, W^{old} \}.
\]

QED.
Bibliography


[16] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert, “Multi-


[22] C. Davatzikos, “Spatial transformation and registration of brain images using


“Regionally selective atrophy of subcortical structures in prodromal HD as revealed by statistical shape analysis,” *Human Brain Mapping*, pp. n/a–n/a, 2012.


[51] M. Vaillant, M. Miller, L. Younes, and A. Trouve, “Statistics on diffeomor-


integrated perspective from hippocampal shape, cortical thickness, and integrity of white matter bundles,” *NeuroImage*, vol. 52, no. 4, pp. 1181 – 1189, 2010.


[93] A. Asman and B. Landman, “Formulating spatially varying performance in the


[100] D. W. Shattuck, S. R. Sandor-Leahy, K. A. Schaper, D. A. Rottenberg, and


[127] C. E. Priebe, M. I. Miller, and J. T. Ratnanather, “Segmenting magnetic reso-


Xiaoying Tang

**Education**

**The Johns Hopkins University**, Baltimore, MD, USA

- Ph.D., Electrical and Computer Engineering, *Expected:* May 2014
- M.S., Electrical and Computer Engineering, May 2011

**Huazhong University of Science and Technology**, Wuhan, Hubei, China

- B.E., Control System Engineering, July 2009
- B.S., Foreign Language Education, July 2009

**Publications**


