Image Classification using Pair-wise Registration and Machine Learning with Applications to Neuroimaging

Xiaojing Long

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Committee Members:
Chris L. Wyatt
A. Lynn Abbott
William T. Baumann
A.A.(Louis) Beex
Ge Wang

December 3, 2010
Blacksburg, VA

Key words: Alzheimer's disease, mild cognitive impairment, MRI, disease diagnosis, pathology localization

Image Classification using Pair-wise Registration and Machine Learning with Applications to Neuroimaging

Xiaojing Long

(ABSTRACT)

Alzheimer's disease (AD) is the most frequent neurodegenerative dementia and a growing health problem. Early and accurate diagnosis and prediction of AD is crucial because treatment may be most efficacious if introduced as early as possible. Neuropsychological testing, which is clinically used, sometimes fails to recognize probable dementia, especially to recognize the disease at an early time point such as the mild cognitive impairment (MCI), which is the prodromal stage of AD.

Recently, there has been a realization that magnetic resonance imaging (MRI) may help diagnoses of AD and MCI. In this dissertation, we introduce an MRI-analysis based algorithm to help diagnose the illness before irreversible neuronal loss has set in, and to help detect brain changes between MCI patients who may convert and may not convert to AD. Given a set of brain MR images, the algorithm first calculates the distance between each pair of images via a registration process. Then images are projected from a high dimensional Euclidean space to a low dimensional Euclidean subspace based on the calculated distances, with a dimension reduction method. Finally classical supervised classification approaches are employed to assign images to appropriate groups in the low dimensional space. The classification accuracy rates we obtained in our experiments are higher than, or at least comparable to, those reported in recently published papers.

Moreover, this algorithm can be extended to explore the pathology distribution of AD. Exploring the distribution of AD pathology is of great importance to reveal AD related regional atrophy at specific stages of the disease and provide insight into longitudinal sequence of disease progression. Calculating distances between different brain structures produces different classification accuracy. Those structures yielding higher classification accuracy are considered as pathological regions. Our experimental results on pathology localization are also compared with the reproduced results using other existing popular algorithms; the observations are consistent.

Dedication

This work is dedicated to my parents, especially to my dear grandfather.

Acknowledgements

First and foremost, I would like to gratefully and sincerely thank my advisor Dr. Chris Wyatt for his guidance, understanding and patience. He encouraged me to question thoughts and express ideas. His patience and support helped me overcome many crisis situations in my research, writing and communication.

I also would like to thank all my committee members for reading my reports, commenting on my views and giving me useful suggestion to enrich my ideas. I am especially grateful to Dr. A. A. (Louis) Beex for the long discussions that helped me correct the grammar and notation in my writings.

I am also thankful to all my colleagues–Ginger (Xiaoxing) Li, Vidya Rajagopolan, Yuan Shen, Bonsung Koo, Micah Chambers, Shaza Zaghlool, Matt Carter and Aditya Mundle. Their support and care helped me overcome setbacks and stay focused on my graduate study. I am especially grateful for the friendship with Ginger, who helped me not only in my research but also in my life.

Most importantly, I can't finish my Ph.D study and my dissertation without the love, concern and support from my family. I would like to express my heart-felt gratitude to all of them, especially to my dear grandfather, who is in paradise still noticing me.

Contents

List of Figures

List of Tables

Abbreviations

AD = Alzheimer's Disease;
ADNI = Alzheimer's Disease Neuroimaging Initiative;
CDR = Clinical Dementia Rating;
cMCI = converting Mild Cognitive Impairment;
CSF = CerebroSpinal Fluid;
DBM = Deformation-Based Morphometry;
DOF = Degree Of Freedom;
FDR = False Discovery Rate;
FN = False Negative;
FP = False Positive;
FWE = Family-Wise Error;
GLM = General Linear Model;
GM = Gray Matter;
LDDMM = Large Diffeomorphic Deformation Metric Mapping;
LOOCV = Leave-One-Out Cross Validation;
mAD = mild Alzheimer's Disease;
MCI = Mild Cognitive Impairment;
MDS = MultiDimensional Scaling;
MMSE = Mini Mental State Exam;
MRI = Magnetic Resonance Imaging;
NC = Normal Control;
NPV = Negative Predictive Value;
OASIS = Open Access Series of Imaging Studies;
pMCI = potential Mild Cognitive Impairment;
PPV = Positive Predictive Value;
ROI = Region Of Interest;

sMCI = stable Mild Cognitive Impairment;
SVM = Support Vector Machine;
TBM = Tensor-Based Morphometry;
TN = True Negative;
TP = True Positive;
VBM = Voxel-Based Morphometry;
VENT = Ventricle;
WM = White Matter;

Chapter 1

Introduction

1.1 Motivation and significance

Alzheimer's disease (AD), the most common form of dementia, is known for placing a great burden on family members or caregivers and is also one of the most costly diseases to society.

Alois Alzheimer, a German neuropathologist, originally described the disease in 1906. In an autopsy of a 56-year-old woman with severe dementia, Alzheimer noted senile plaques outside neurons and neurofibrillary tangles within neurons [1]. One hundred years later, definitive diagnosis of Alzheimer's disease still can only be made with histopathological confirmation of the presence of plaques and neurofibrillary tangles as determined by biopsy or autopsy [2, 3, 4].

Neuropsychological testing is sometimes clinically used in the evaluation of Alzheimer's disease to learn the nature and level of a person's impairment [5]. The testing is often conducted by a neuropsychologist. Several tests are available that can narrow the range of possible diagnoses by producing test patterns that resemble Alzheimer's or other conditions. A neuropsychologist might use one test or a whole battery of them, depending on the individual. Some of the common tests include: (1) Mini-Mental State Exam (MMSE), a brief, structured test of mental status which assesses orientation, word recall, attention and calculation, language abilities, and visuospatial ability; (2) Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog), which is considered more thorough than the MMSE, focusing on attention, language, orientation, executive functioning, and memory skills; (3) NeuroPsychiatric Inventory (NPI), assessing several neuropsychiatric problems commonly seen

in Alzheimer's disease and other types of dementia, such as agitation, anxiety, apathy, delusions, hallucinations, eating difficulties, and sleep disturbances. Other tests involve blessed test, neurobehavioral cognitive status examination, clock drawing test and so on. Neuropsychological testing shouldn't be used independently to make a diagnosis, nor should any other singular aspect of the diagnostic process, such as the MMSE or a general physical. A diagnosis of Alzheimer's can only be made after a complete diagnostic workup rules out any other possible cause for the person's symptoms. However, a number of limitations of clinical criteria for diagnosing Alzheimer's disease are evident [7, 8, 9]. First, for those patients with severe depression, aphasia, and apraxia, clinical evaluation is difficult to make. Second, early memory loss also occurs in normal aging, which however has a slow progression. Longitudinal clinical testing is thus necessary to distinguish it from very early Alzheimer's disease which has a more rapid progression. Third, using the clinical criteria alone can not preclude the possibility of other degenerating dementias (e.g. primary progressive aphasia, posterior cortical atrophy, corticobasal degeneration, and frontotemporal dementia) which also show a similar decline in cognitive functions [6]. Moreover, the diagnosis based on neuropsychological scale requires rich clinical experience of neuropsychologists or physicians, and as a result it is subjective and less repeatable. For these reasons, increased emphasis has been placed on the role of neuroimaging techniques to establish the diagnosis of Alzheimer's disease.

With the development of therapeutic strategies for Alzheimer's disease, early diagnosis and monitoring of Alzheimer's disease have taken on increased importance. Magnetic resonance imaging (MRI) becomes a key diagnostic tool not only for accurate diagnosis of Alzheimer's disease but also for identification of preclinical stages of Alzheimer's disease that are not discernible using solely clinical testing [7]. The latter indication is targeted toward the identification of patients in whom the effects of neuronal loss seen on scans precede decline in cognitive performance scores and in whom early intervention may slow disease progression.

Mild cognitive impairment (MCI) is an operational diagnostic term developed to describe the preclinical stage of Alzheimer's disease. MCI is a condition in which a person has problems with memory, language, or another mental function severe enough to be noticeable to other people and to show up on tests, but not serious enough to interfere with daily life. Because the problems do not interfere with daily activities, the person does not meet criteria for being diagnosed with dementia. However, the rate at which MCI subjects convert to AD each year is ten times higher than the rate for normal subjects [24]. Thus MRI-analysis methods are desired to help detect subtle brain changes, or to help diagnose the illness before irreversible neuronal loss has set in [10].

Among all MRI-based diagnosis approaches, there has been recent interest in the application of machine learning techniques. As medical imaging technologies are used more and more, a large number of medical images are produced in hospitals and healthcare centers. Valuable medical knowledge bases are being created but the utilization is not being fully exploited from such data [11]. Specifically, image-supported medical information (diagnostic, prognostic, treatment, etc.) related to past cases can be very useful in the diagnosis and treatment of new cases. However, manually searching, retrieving, comparing and inferring over large-scale databases requires costly labor and it is time-consuming. More advanced techniques are consequently desired to complete these tasks. Machine learning methods have advantages that satisfy the practical requirements: (1) they can be fully automated; (2) they provide standard PC-based clinical decisions; (3) they are unaffected by individual neuroradiological expertise. In machine learning algorithms, registration is widely performed to measure the similarity between images, either the deformation field (which reflects displacement between images) or the residual (which reflects anatomic deviations after macroscopic differences in shape are discounted) is used. Such methods have increasingly been applied to a number of diagnostic problems in Alzheimer's disease and mild cognitive impairment (MCI) [12, 13, 14].

Besides early diagnosis and prediction, researchers are also interested in exploring pathology distribution in AD brain [15, 16, 18, 17, 19, 20, 21, 22, 23]. Spatially distribution maps of histopathological changes can reveal AD related regional atrophy at an early time point, or at a specific stage of disease [16]. For example, they can show profiles of brain changes that typically occur as subjects convert from normal cognition to MCI and from MCI to AD. Moreover, they provide insight into the longitudinal sequence of disease progression, or help detect subtle brain changes associated with prognosis or treatment.

1.2 Research goals

In this dissertation, we introduce a machine learning based method which uses embedding of registration distance. This method can not only be used to classify normal subjects, AD and MCI patients to appropriate groups, but also be used to detect regional deviations in the brains between groups.

For classification, the algorithm will be applied to

- discriminate brains between normal controls and AD patients;
- discriminate brains between normal controls and MCI subjects;

- discriminate brains between normal controls and potential MCI subjects (which were healthy at the time of scanning but later had converted to MCI);
- discriminate brains between stable MCI subjects (which were MCI at the time of scanning and had not converted to AD) and converting MCI subjects (which were MCI at the time of scanning but latter had converted to AD);
- discriminate brains between normal controls and AD patients when the MR images are obtained from different communities and scanned by different machines using different protocols.

Classification results will be compared with those from other existing methods.

For pathology localization, the algorithm will be applied to

- explore pathology distribution on MCI subjects compared with normal controls;
- explore pathology distribution on AD patients compared with MCI subjects;
- explore pathology distribution on potential MCI subjects (which were healthy at the time of scanning but latter had converted to MCI) compared with normal controls;
- explore pathology distribution on converting MCI subjects (which latter had progressed to AD) compared with stable MCI subjects (which were MCI at the time of scanning and had not converted to AD).

The results will be compared with the reproduced results using other existing methods.

1.3 Basic concepts

1.3.1 Machine learning

Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge. This capacity to learn from experience, analytical observation, and other means, results in a system that can improve its own speed or performance, i.e., its efficiency and/or effectiveness. Learning techniques and methods have been successfully applied to a variety of learning tasks in a broad range of areas, including text classification, image

segmentation/classification, gene discovery, financial forecasting, information filtering, design of adaptive web agents and others.

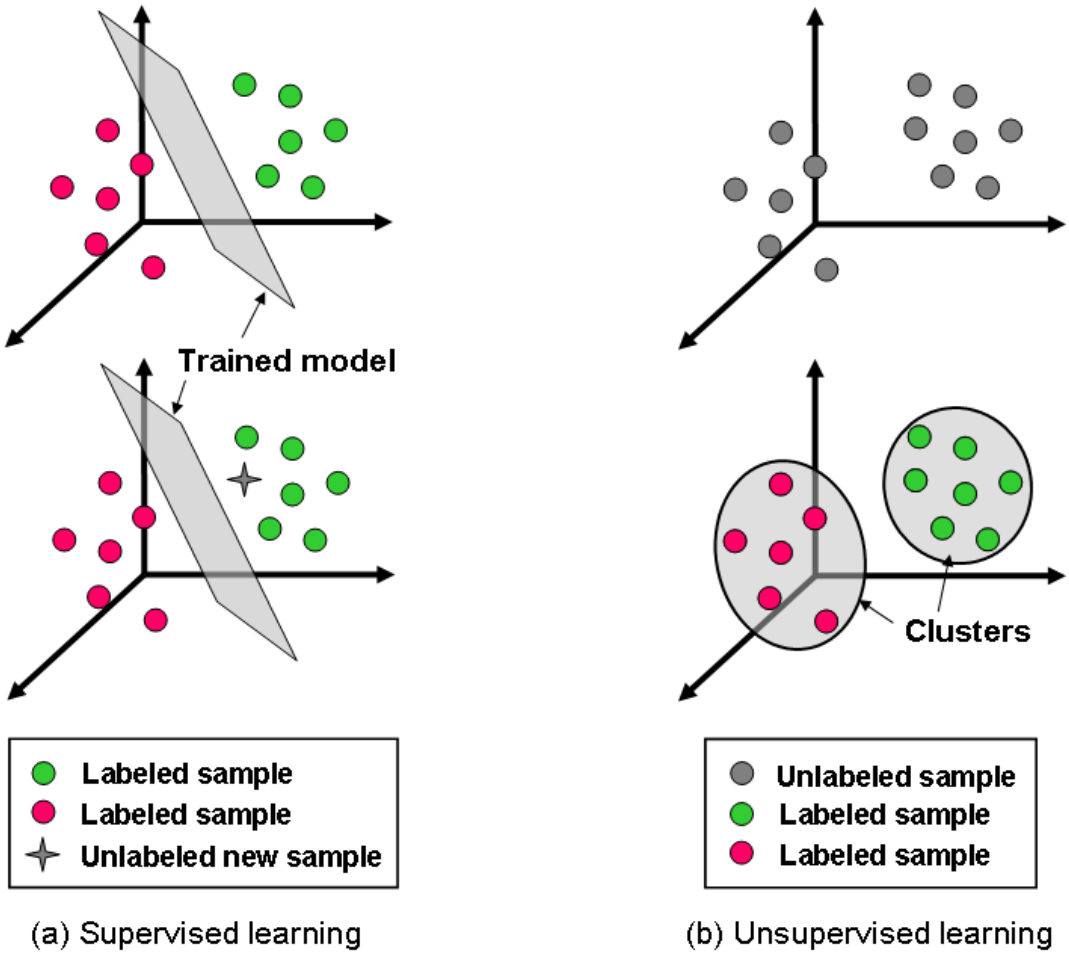


Figure 1.1: Supervised learning and unsupervised learning.

Machine learning algorithms are described as either "supervised" or "unsupervised". The distinction is drawn from how the learner classifies data (as shown in Fig.1.1). In supervised algorithms, the classes are predetermined. These classes can be conceived of as a finite set, previously assigned manually. In practice, a certain segment of data will be labeled with these classifications. The machine learner's task is to search for patterns and construct mathematical models. These models then are evaluated on the basis of their predictive capacity in relation to measures of variance in the data itself. Many of the well-known methods, such as decision tree learning, Bayesian learning, and support vector machines, are

examples of supervised learning techniques.

Unsupervised learners are not provided with predetermined classifications. In fact, the basic task of unsupervised learning is to develop classification labels automatically. Unsupervised algorithms seek out similarity between pieces of data in order to determine whether they can be characterized as forming a group. These groups are termed clusters, and there is a whole family of clustering machine learning techniques such as K-Means and meanshift.

1.3.2 Registration and diffeomorphism

Image registration is the process of aligning two or more images of the same scene taken at different times, from different viewpoints, on different subjects, and/or by different sensors. Medical image registration is a crucial step in many clinical events, not only within clinical diagnostic settings, but also in the area of planning, consummation, and evaluation of surgical and radiotherapeutical procedures.

The components of a image registration framework and their connections are shown in Fig.1.2. The inputs of a registration process are defined as the target (or fixed) image $I_0(p)$ and the source (or moving) image $I_1(p)$, where p denotes a position in N-dimensional space. Registration is treated as an optimization problem with the goal of finding the spatial mapping that will bring the source image into alignment with the target image. The transformation $T(p)$ represents the spatial mapping of points from the target image space to points in the source image space. The interpolator is used to evaluate source image intensities at non-rigid positions. The similarity measure $S(I_0, I_1 \circ T)$ provides a measure of how well the target image is matched by the transformed source image. This measure forms the quantitative criterion to be optimized by the optimizer over the search space defined by the parameters of the transform.

The transformation T always contains a global transformation and a local transformation:

$$T(x, y, z) = T_{global}(x, y, z) + T_{local}(x, y, z). \tag{1.1}$$

The global transformation describes the overall motion of images. A general class of global transformations is affine registration (as shown in Fig.1.3), which has 12 degrees of freedom, describing the rotation, translation, scaling, and shearing. In 3D, an affine transformation

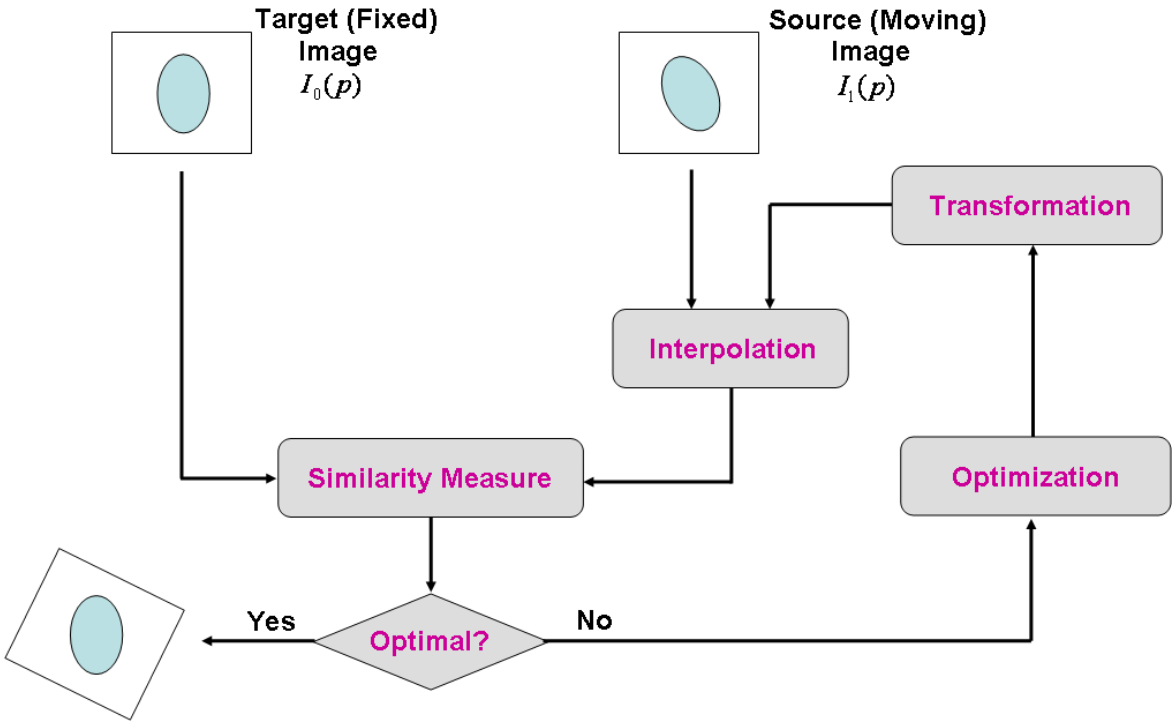


Figure 1.2: The basic components of the registration framework and their connections.

can be written as

$$T_{global}(x, y, z) = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} \theta_{14} \\ \theta_{24} \\ \theta_{34} \end{pmatrix}, \tag{1.2}$$

where the coefficients $\{\theta_{ij}\}$ parameterize the 12 degrees of freedom of the transformation.

The affine transformation captures only the global motion between images. An additional transformation is required which models the local deformation. The local transformation always has high degrees of freedom, therefore, it is difficult to describe the local deformation via parameterized transformation. In this work, a non-parametric registration will be introduced to find the optimal displacement field $T(p)$ (which is represented by $\phi(p)$ in the following chapters) between the target image $I_0(p)$ and the source image $I_1(p)$. The optimal

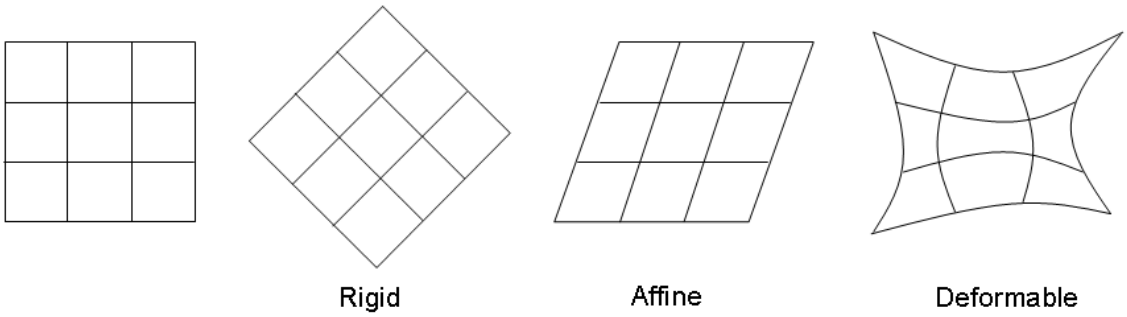


Figure 1.3: Types of transformations.

$T(p)$ is obtained by optimizing the following energy function:

$$E(T) = \frac{1}{\sigma_i^2} \|I_0 - I_1 \circ T\| + \frac{1}{\sigma_T^2} Reg(T), \tag{1.3}$$

where the first term measures the resemblance of the aligned images, the second term is a regularization estimating the likelihood of the transformation, and σ_i accounts for the noise on the image intensity, σ_T controls the amount of regularization. The operation "$\circ$" represents a composition, e.g. $I_1 \circ T$ is to warp I_1 by T.

In order to register the target image and the source image, we need to optimize Eq. 1.3 over a given space of spatial transformations. If this space is the entire space of displacement fields, which is called a *vector space*, then the transformation T is an element of the vector space and it can thus be simply added with other transformations belonging to the same space. If this space consists of diffeomorphisms which are smooth, continuous and bijective, then the transformation T is an element of the Lie group, and Euclidean statistics can not be applied directly to T. The distinction of the optimization spaces results in different types of registration, which will be introduced in Chapter 2.

1.4 Organization of this work

The remainder of this dissertation is organized as follows. In Chapter 2, we detail the classification procedure, including the registration algorithm, the calculation of distance as well as the embedding methods, and show our classification results and a comparison of results with other existing methods. In Chapter 3, we proceed to illustrate the utilization

of classification for pathology localization and introduce how recently popular approaches achieve this goal. Experiment results on clinical data are shown too. Conclusion and future work is provided in Chapter 4.

Chapter 2

Classification and prediction of mild cognitive impairment and Alzheimer's disease

2.1 Background

Since early diagnosis of AD or MCI is of increasing importance, many algorithms have been proposed, varying from conceptually simple measurement of volumes or mathematically complex description of shape differences in *a-priori* regions of interest (ROI) [27, 28], to voxel-wise modeling of tissue density changes on the whole brain region, e.g. voxel-wise morphometry [29, 30]. There has been recent interest in machine learning and computer-aided diagnostics in the field of medical imaging, where a machine learning algorithm is trained to produce a desired output from a set of input training data. In the diagnosis of AD, the input training data could be features obtained from voxel intensity, tissue density or shape descriptor. Machine learning diagnostics can be also divided into ROI-based and full-brain based methods. ROI-based algorithms always focus on the medial temporal lobe structures of the brain, including the hippocampus and entorhinal cortex. Trosset et al. [12] proposed a semisupervised learning method, which used multidimensional scaling based on LDDMM (Large Diffeomorphic Deformation Metric Mapping) distance between hippocampi for embedding, and used linear discriminate analysis for training. In the work of Gutman et al.[31] and Gerardin et al. [32], SVM is again used for classification of patients with AD or mild cognitive impairment (MCI), analyzing the hippocampal shape features. However

these ROI-based analyses typically do not make use of all the available information contained in the full brain, and require *a-priori* decisions concerning which structures to assess. In many other research [33, 18], atrophy has also been detected in the inferior and lateral temporal lobes, the cingulate gyri, and in the parietal and frontal lobes. Thus, analysis of hippocampus or medial temporal lobe may not be the best way to detect AD. Kloppel et al. [13] developed a supervised method using linear support vector machine (SVM) to group the gray matter segment of T1-weighted MR scans on a high dimensional space, treating voxels as coordinates and intensity value at each voxel as their location. This whole-brain technique shows high discriminative power that it can not only assign individual diagnoses but also differentiate brains obtained from different scanners.

In this paper, we extend advantages of the whole-brain technique and propose a machine learning algorithm which can be used for individual patient diagnosis, discrimination of brains from different scanners, as well as prediction of MCI. In our method, the efficient symmetric log-domain diffeomorphic demons registration algorithm is first employed to compute the pair-wise registration between whole-brain images, giving a diffeomorphism that deformably aligns each pair of brains. All the brain images are segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using the FreeSurfer [34] software. The whole-brain diffeomorphism or regional diffeomorphisms (which are the overlapped region between the whole-brain diffeomorphism and segmentations) are quantified as Riemannian distances. With the matrix constructed by distance between each pair of subjects, embedding algorithms, including the metric multi-dimensional scaling and the spectral method, project images onto a low-dimensional space. In the embedded space, we used labeled images to train a discriminant hyperplane separating underlying groups. Then we project unlabeled subjects onto the constructed embedded space using the extending embedding algorithm and assign them to appropriate groups based on the computed hyperplane. We conducted experiments on several datasets of MR scans, including (1) normal controls *vs* moderate-to-severe AD, (2) normal controls *vs* mild AD, (3) normal controls *vs* converting MCI (which were diagnosed as MCI at the time of scanning but latter had progressed to AD), (4) normal controls *vs* potential MCI (which were healthy at the time of scanning but had converted to MCI in 4 ± 1.41 years), and (5) stable MCI (which were diagnosed as MCI at the time of scanning and had not converted to AD in 2.81 ± 0.61 years) *vs* converting MCI (which were diagnosed as MCI at the time of scanning and had converted to AD in 1.88 ± 0.62 years). Discriminating normal controls and AD patients aims to test the diagnosis performance of the proposed algorithm, while separating normal controls and potential MCI subjects or separating stable MCI and converting MCI subjects aims to test

the prediction performance of the algorithm. In practice, the prediction performance is more important since treatment may be most efficacious if introduced as early as possible. The correct classification rate, sensitivity and specificity are calculated to evaluate the diagnosis and prediction performances with different distance metrics. The positive and negative predictive values are calculated to evaluate the reliability of the test.

2.2 Method

2.2.1 Registration and distance metric

2.2.1.1 Image registration

The framework and categories of image registration have been introduced in Chapter 1.3.2. Non-linear registration has been used in morphometric studies for characterizing and comparing anatomies or group analysis. Moreover, the output deformation fields of non-linear registration algorithms are often required to be invertible and symmetric with respect to the order of the inputs [35, 36]. Vercauteren et al. proposed the symmetric log-domain diffeomorphic demons algorithm [38] that guarantees the above properties.

To fully understand the symmetric log-domain diffeomorphic demons registration, we should first know the basic idea behind the demons algorithm. Non-linear registration aims at estimating a smooth, continuous mapping $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^N$ that best aligns two given images: the target image I_0 and the source image I_1. Define $Sim(I_0, I_1, \phi)$ be the similarity measure of the aligned images and $Reg(\phi)$ a regularization to estimate the likelihood of the transformation, the global energy functional is:

$$E(s) = \frac{1}{{\sigma_i}^2} Sim(I_0, I_1, \phi) + \frac{1}{{\sigma_T}^2} Reg(\phi), \tag{2.1}$$

where σ_i accounts for the noise on the image intensity and σ_T controls the amount of regularization.

In order to register the target and source images, non-parametric demons algorithms need to optimize Eq. (2.1) over a given space of spatial transformations. This space can be the entire space of the displacement fields, and it can also be the space of diffeomorphisms. The distinction of the optimization spaces results in three different types of algorithm in

Demons family. They are (1) additive demons; (2) diffeomorphic demons; (3) log-domain diffeomorphic demons.

a. Additive demons

Giving a target image I_0 and a source image I_1, the global energy functional of additive demons is:

$$E_{addi}(u) = \|I_0 - I_1 \circ (\phi + u)\|^2 + \frac{\sigma_i^2}{\sigma_x^2} \|u\|^2 , \tag{2.2}$$

where ϕ denotes the transformation modeling the spatial mapping of points from the target image space to the source image space, u denotes the update field, σ_i accounts for the noise on the image intensity and σ_x accounts for a spatial uncertainty on the correspondences.

The algorithm is initialized with $\phi = Id$, an identity transformation, and the optimization is performed over the entire space of the displacement field. At each iteration of the optimization procedure, the update field u is calculated and it is superimposed to ϕ. Since the spatial transformations form a vector space, transformations can thus be simply added. Classical descent methods, such as Newton's method, Powell's direction set method, downhill simplex method and so on, can be used for optimization based on the additive iterations of the form $\phi \leftarrow \phi + u$.

Algorithm (Additive Demons)

- *Choose a starting spatial transformation (a vector field) ϕ.*
- *Iterate until convergence*
  - *Compute the correspondence update field u using*

$$u(p) = -\frac{I_0(p) - I_1 \circ \phi(p)}{\|J^p\|^2 + \frac{\sigma_i^2(p)}{\sigma_x^2}} {J^p}^T ,$$

    where $J^p = -\nabla_p^T I_0$ or $J^p = -\nabla_p^T (I_1 \circ \phi)$ with Thirion's rule[1].
  - *If a fluid-like regularization is used, let $u \leftarrow K_{fluid} * u$.*
  - *Let $c \leftarrow \phi + u$.*
  - *If a diffusion-like regularization is used, let $\phi \leftarrow K_{diffusion} * c$ (else let $\phi \leftarrow c$).*

The convolution kernels K_{fluid} and $K_{diffusion}$ in the algorithm will typically be Gaussian with parameters σ_{fluid} and $\sigma_{diffusion}$, respectively.

b. Diffeomorphic demons

The additive demons algorithm is an optimization procedure on the entire space of displacement fields. Vercauteren et al. adapts the basic demons algorithm to make it diffeomorphic by optimizing Eq. (2.2) over a space of diffeomorphisms [37]. However, diffeomorphisms do not form a vector space but only a Lie group, where classical descent methods can not be used for optimization. In [37], Vercauteren et al. used a strategy known as geometric optimization [104], which guarantees convergence on Lie groups as the classical Newton methods do on vector spaces. To adapt the classical Newton-Raphson method, an intrinsic update step is used:

$$\phi \leftarrow \phi \circ exp\,(u). \tag{2.3}$$

The '$\circ$' in Eq. 2.3 also represents a composition operation. However, the composition between two deformation fields is different from the composition between an image and a deformation field. The latter is a simple warping procedure, while the former involves a space shift and an addition, e.g. computing $\phi_1 \circ \phi_2$ will first treat ϕ_1 as a fixed field and warp ϕ_2 to the space of ϕ_1 based on ϕ_1^{-1}; then add the warped ϕ_2 to ϕ_1 (as shown in Fig. A.2).

Let $\mathcal{G}$ be a Lie group for the composition '$\circ$' [37]. Any Lie group can be associated with a Lie algebra $\mathfrak{g}$. $\mathcal{G}$ and $\mathfrak{g}$ are related through the group exponential which is a diffeomorphism from a neighborhood of 0 in $\mathfrak{g}$ to a neighborhood of Id in $\mathcal{G}$. The exponential map can be used to get the Taylor expansion of a smooth function φ on $\mathcal{G}$:

$$\varphi(\phi \circ exp(u)) = \varphi(\phi) + J_\phi^\varphi u + O(\|u\|^2), \tag{2.4}$$

where $[J_\phi^\varphi]_i = \frac{\partial}{\partial u_i}\varphi(\phi \circ exp(u))|_{u=0}$, and $O(\|u\|^2)$ is neglected in calculation.

Then we use the following approximation:

$$I_0(p) - I_1 \circ \phi \circ exp\,(u)\,(p) \approx I_0(p) - I_1 \circ \phi(p) + J^p \cdot u(p),$$

[1]Even if a simple transformation space and a classical regularization term are used in Eq. 2.2, optimizing the energy functional needs the solution of an Euler-Lagrange equation which is computationally expensive. Thirion's demons algorithm uses an efficient two-step procedure at each optimization iteration: It first looks for an unconstrained update step with an optical flow computation; then it uses a simple Gaussian smoothing on the updated transformation.

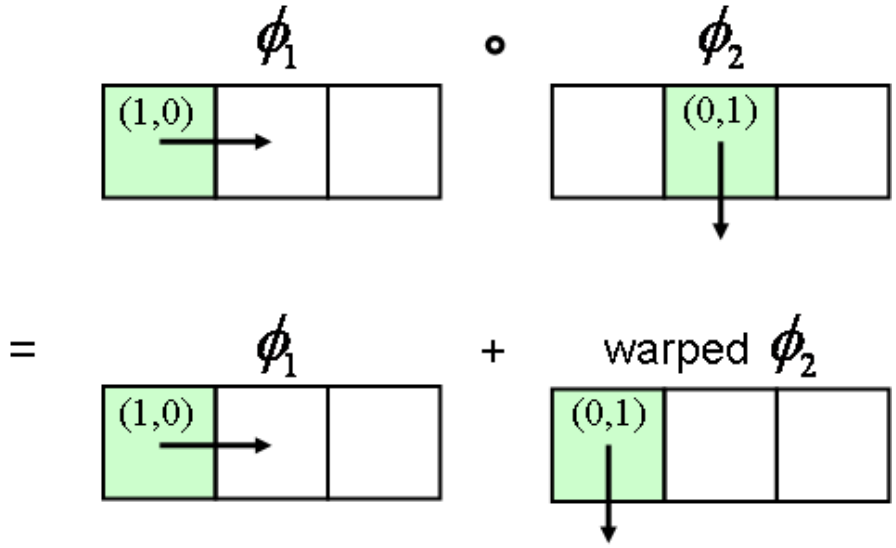


Figure 2.1: The composition of deformation fields.

where $J^p = -\nabla_p^T I_0$ or $J^p = -\nabla_p^T (I_1 \circ \phi)$.

Thus the energy functional for the diffeomorphic demons algorithm can be rewritten as:

$$\begin{aligned} E_{diffeo}(u) &= \|I_0 - I_1 \circ (\phi \circ exp(u))\|^2 + \frac{\sigma_i^2}{\sigma_x^2} \|u\|^2 \\ &\approx \|I_0 - I_1 \circ \phi + J^p \cdot u\|^2 + \frac{\sigma_i^2}{\sigma_x^2} \|u\|^2 \end{aligned} \tag{2.5}$$

The non-parametric diffeomorphic image registration algorithm is then obtained:

Algorithm (Diffeomorphic Demons)

- *Choose a starting spatial transformation (a vector field) ϕ.*
- *Iterate until convergence*
  - *Compute the correspondence update field u*

$$u(p) = -\frac{I_0(p) - I_1 \circ \phi(p)}{\|J^p\|^2 + \frac{\sigma_i^2(p)}{\sigma_x^2}} J^{p^T}.$$

– *If a fluid-like regularization is used, let* $u \leftarrow K_{fluid} * u$.

– *Let* $\phi \leftarrow \phi \circ exp(u)$.

– *If a diffusion-like regularization is used, let* $\phi \leftarrow K_{diffusion} * \phi$.

The convolution kernels K_{fluid} and $K_{diffusion}$ in the algorithm will typically be Gaussian with parameters σ_{fluid} and $\sigma_{diffusion}$, respectively.

c. Symmetric log-domain diffeomorphic demons

In diffeomorphic demons, the exponential is used only to encode the adjustment made at each iteration to the update field. In the log-domain diffeomorphic demons algorithm [38], the basic diffeomorphic demons $E_{diffeo}(I_0, I_1, \phi, u) = \|I_0 - I_1 \circ (\phi \circ exp(u))\| + \|u\|^2$ is extended to represent the complete spatial transformation in the log domain. The main idea of this algorithm is to represent the current transformation ϕ as an exponential of a smooth velocity field[2] v, i.e. $\phi = exp(v)$, and use the basic diffeomorphic demons to compute a field u for an update of the form $\phi \circ exp(u) = exp(v) \circ exp(u)$. Hence the updating rule is defined as

$$\phi = exp(v) \leftarrow exp(Z(v, u)) \approx exp(v) \circ exp(u) = \phi \circ exp(u), \tag{2.6}$$

where $Z(v, u)$ is a velocity field. Since the update field u is assumed to be small, the first order approximation of $Z(v, u)$ is computed by the Baker-Campbell-Hausdoff approximation:

$$Z(v, u) \approx v + u + \frac{1}{2}[v, u] + \frac{1}{12}[v, [v, u]], \tag{2.7}$$

where $[v, u]$ denotes a Lie bracket providing a velocity field defined at each voxel p by

$$[v, u](p) = [Jac(v)(p)]u(p) - [Jac(u)(p)]v(p). \tag{2.8}$$

In Eq. 2.8, $u(p)$ and $v(p)$ are 3×1 vectors in 3D image registration, $Jac(u)(p)$ and $Jac(v)(p)$ are 3×3 matrices:

[2]In this dissertation the velocity field is defined as a smooth stationary field in the log-domain. Since the exponential of velocity field is a diffeomorphism, the velocity field is invertible and can be applied directly with Euclidean statistics.

$$Jac(u)(x,y,z) = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(y-u_y)/\partial x & \partial(y-u_y)/\partial y & \partial(y-u_y)/\partial z \\ \partial(z-u_z)/\partial x & \partial(z-u_z)/\partial y & \partial(z-u_z)/\partial z \end{pmatrix}.$$

Thus, $[v,u](p)$ is still a 3×1 vector.

Log-domain diffeomorphic demons

The outputs of the log-domain diffeomorphic demons algorithm include not only the deformation field ϕ, but also the logarithm of the diffeomorphism, $v = log(\phi)$, which can be directly used in computational anatomic analysis [3]. The log-domain diffeomorphic demons algorithm is obtained as

Algorithm (Log-Domain Diffeomorphic Demons)

- *Choose a starting spatial transformation* ϕ
- *Iterate until convergence*
  - *Given the current transformation* $\phi = exp(v)$*, compute a correspondence update field* u *by minimizing* $E_{diffeo}(I_0, I_1, \phi, u)$ *with respect to* u*.*
  - *For a fluid-like regularization, let* $u \leftarrow K_{fluid} * u$*.*
  - *Let* $v \leftarrow Z(v,u) = v + u + \frac{1}{2}[v,u]$*.*
  - *For a diffusion-like regularization, let* $v \leftarrow K_{diffusion} * v$*.*

The convolution kernels K_{fluid} and $K_{diffusion}$ in the algorithm will typically be Gaussian with parameters σ_{fluid} and $\sigma_{diffusion}$, respectively.

Symmetric extension

The log-domain diffeomorphic demons registration has a symmetric (or inverse-consistent) extension [38]. The symmetric registration framework can be obtained by symmetrizing the global energy:

$$\phi_{opt} = \underset{\phi}{\operatorname{argmin}}(E(I_0, I_1, \phi) + E(I_1, I_0, \phi^{-1})). \tag{2.9}$$

[3] Computational Anatomy (CA) is emerging as a discipline focused on the quantitative analysis of variability of biological shape. Applications of CA in brain science have developed rapidly, with applications of large deformation brain mapping technologies that provide mechanisms for discovering neuropsychiatric disorders of many types.

Eq. 2.9 can be also formulated as a constrained optimization using two diffeomorphisms:

$$[\phi_{opt}, \phi_{opt}^{-1}] = \underset{[\phi,\psi]|\psi=\phi^{-1}}{\operatorname{argmin}} E(I_0, I_1, \phi) + E(I_1, I_0, \psi). \tag{2.10}$$

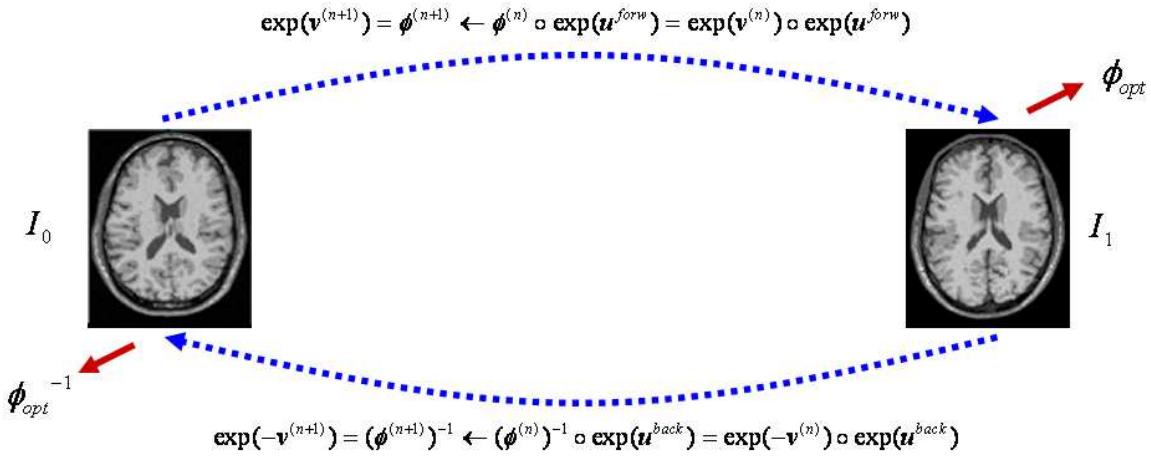


Figure 2.2: The symmetric extension of log-domain diffeomorphic demons algorithm.

By using a complete log-domain demons iteration starting from $\phi = exp(v)$ to optimize the first term $E(I_0, I_1, exp(\varsigma))$, one can get $\varsigma = K_{diffusion} * Z(v, K_{fluid} * u^{forw})$, where u^{forw} is the demons force from I_0 to I_1. Similarly, the second term $E(I_1, I_0, exp(-\tau))$ is optimized with $\tau = -K_{diffusion} * Z(v, K_{fluid} * u^{back})$, where u^{back} is the demons force from I_1 to I_0.

Since the log-domain is a vector space [4], an easy projection operator is designed that guarantees the symmetry of the results [38]:

$$v = \frac{1}{2} K_{diffusion} * [Z(v, K_{fluid} * u^{forw}) - Z(v, K_{fluid} * u^{back})]. \tag{2.11}$$

Algorithm (Symmetric Log-Domain Diffeomorphic Demons)

- *Choose a starting spatial transformation ϕ*
- *Iterate until convergence*

[4] Diffeomorphisms form a Lie group, thus the logarithms of diffeomorphisms form a Lie algebra which is a vector space.

Given the current transformation $\phi = exp(v)$,

- *compute the demons force* u^{forw} *to minimize* $E^{corr}_{diffeo}(I_0, I_1, exp(v), u^{forw})$.
- *compute the demons force* u^{back} *to minimize* $E^{corr}_{diffeo}(I_1, I_0, exp(-v), u^{back})$.
- *For a fluid-like regularization, let* $u^{forw} \leftarrow K_{fluid} * u^{forw}$, $u^{back} \leftarrow K_{fluid} * u^{back}$.
- *Let* $v \leftarrow \frac{1}{2}[Z(v, u^{forw}) - Z(-v, u^{back})]$.
- *For a diffusion-like regularization, let* $v \leftarrow K_{diffusion} * v$.

The convolution kernels K_{fluid} and $K_{diffusion}$ in the algorithm will typically be Gaussian with parameters σ_{fluid} and $\sigma_{diffusion}$, respectively.

d. Registration examples

Given two input images, a target image and a source image, we aligned the source image to the target space using the additive demons, diffeomorphic demons and log-domain diffeomorphic demons algorithm respectively. We selected two images that have topological changes for easier visual comparison over the registration results. The affine registration was applied to the inputs to remove global differences such as rotation, translation, scaling and shearing. We used the source code downloaded from Insight Journal (http://hdl.handle.net/10380/3060) to conduct the deformable registration. The source code is written in c++ using a set of Insight Toolkit (ITK, www.itk.org) reusable filters and classes. The parameters in this example were set as follows:

- -f —fixed-image <filename>
- -m —moving-image <filename>
- -o —output-image <filename>
- -O —output-field <filename>
- -i —number-of-iteration <45x65x85x105>
- -s —deformation-field-sigma-in-the-*diffusion*-like-kernel <2>
- -g —update-field-sigma-in-the-*fluid*-like-kernel <2>
- -l —max-step-length <2.0>
- -a —update-rule <*additive* or *diffeomorphic* or *log-domain diffeomorphic*>
- -t —gradient-type <0:symmetrized>
- -e —use-histogram-matching

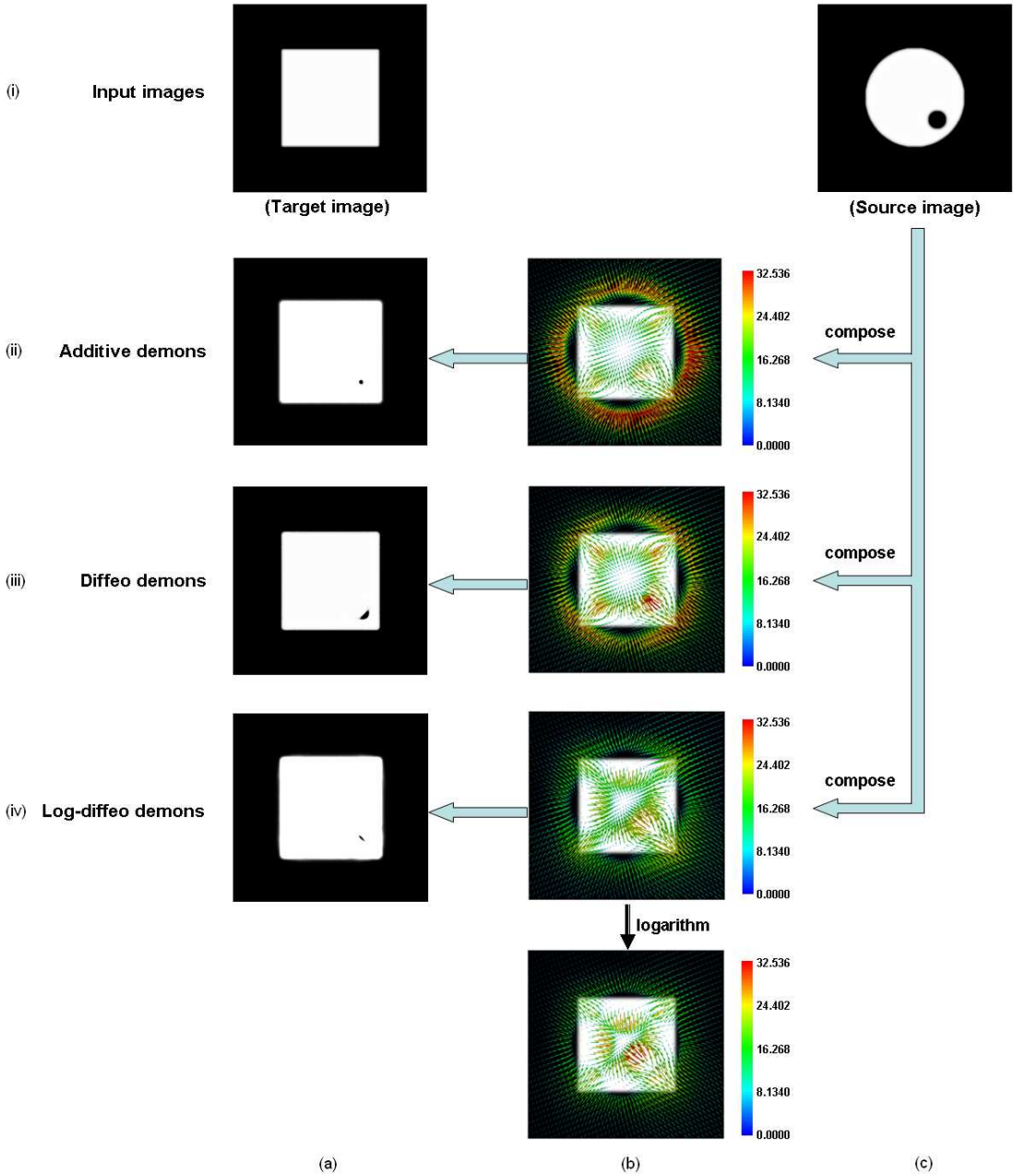


Figure 2.3: Registration results using the additive demons, diffeomorphic demons, log-domain diffeomorphic demons algorithm, respectively. Diffeo=diffeomorphic.

The aligned source images and corresponding deformation fields are shown in Fig. 2.3. Although the additive demons algorithm produced visually good registered images (for example, Fig. 2.3 (ii-a)) with sharp boundary and smaller residual, the deformation field (Fig. 2.3 (ii-

b)) is not as smooth as the field yielded by the log-domain diffeomorphic demons algorithm (Fig. 2.3 (iv-b)). Consequently, the field (ii-b) is not a diffeomorphism and thus could not be used directly in computational anatomic analysis, such as tensor-based morphometry.

2.2.1.2 Distance calculation

To compute the distance between images, a Riemannian distance [39] is defined. The Riemannian distance between two points is the length of the shortest geodesic curve between them. Given two elements a and b such that $a, b \in \mathcal{G}$, where $\mathcal{G}$ denotes a Lie group, the Riemannian distance between a and b is defined as

$$d(a,b) = \left\| log(a^{-1}b) \right\| . \tag{2.12}$$

Since diffeomorphisms are elements of a Lie group, Eq. 2.12 can be used to compute the Riemannian distance between diffeomorphisms. For each pair of images $\{I_j, I_k\}$, the symmetric log-domain diffeomorphic demons algorithm outputs a diffeomorphic deformation field $\phi_{jk} = \phi$ mapping from I_k to I_j, its logarithm $\boldsymbol{v}_{jk} = log(\phi_{jk})$ (that is, $\phi_{jk} = exp(\boldsymbol{v}_{jk})$), and an inverse mapping $\phi_{kj} = \phi^{-1} = exp(-\boldsymbol{v}_{jk})$ from I_j back to I_k (let $\boldsymbol{v}_{kj} = log(\phi_{kj})$, then $\boldsymbol{v}_{kj} = -\boldsymbol{v}_{jk}$). The following Eq.(2.13) is used to compute the Riemannian distance between I_j and I_k:

$$\begin{aligned}
dist(I_j, I_k) &= dist(I_k, I_j) \\
&= dist(Id, \phi_{jk}) = dist(Id, \phi_{kj}) \\
&= \left\| log(Id^{-1}\phi_{jk}) \right\| = \left\| log(Id^{-1}\phi_{kj}) \right\| \\
&= \| log(\phi_{jk}) \| = \| log(\phi_{kj}) \| \\
&= \| log(\phi) \| = \left\| log(\phi^{-1}) \right\| \\
&= \| \boldsymbol{v}_{jk} \| = \| \boldsymbol{v}_{kj} \| \\
&= \| \boldsymbol{v} \| = \| \text{-}\boldsymbol{v} \| .
\end{aligned} \tag{2.13}$$

In Eq.(2.13), ϕ can be either a diffeomorphism of the whole brain or a sub-region of a diffeomorphism overlapped with any segmented structure of the brain, for example, the gray matter (GM), white matter (WM), cortical lobes, hippocampus or other subcortical tissues. If the region of interest (whole brain or any brain structure) is denoted by Ω, the regional

distance between I_j and I_k is defined as:

$$\begin{aligned} REGdist(I_j, I_k) &= REGdist(I_k, I_j) \\ &= REGdist(Id, \phi_{jk}) = REGdist(Id, \phi_{kj}) \\ &= (\left\|log(Id^{-1}\phi_{jk})\right\|)_\Omega = (\left\|log(Id^{-1}\phi_{kj})\right\|)_\Omega \\ &= (\|log(\phi_{jk})\|)_\Omega = (\|log(\phi_{kj})\|)_\Omega \\ &= \sum_{p\in\Omega} \|v_{jk}(p)\| = \sum_{p\in\Omega} \|v_{kj}(p)\| = \sum_{p\in\Omega} \|-v_{jk}(p)\| \\ &= \sum_{p\in\Omega} \|v(p)\| = \sum_{p\in\Omega} \|-v(p)\|\,, \end{aligned} \tag{2.14}$$

where Id denotes an identity transformation, p denotes voxels belonging to the region of interest Ω.

2.2.2 Embedding methods

Unsupervised learning algorithms are a mainstay of machine learning. Examples include Laplacian Eigenmaps [40], Isomap [41], Principal Component Analysis (PCA) [42], Multi-Dimensional Scaling (MDS) [43]. Many such algorithms use an eigen-decomposition for obtaining a low-dimensional embedding of data lying on a non-linear manifold. All embedding algorithms can be cast in the same framework as follows [44]:

Algorithm (Framework of eigen-based learning algorithms)

Given a data set $D = \{x_1, \cdots, x_n\}$ with n points in $\mathbb{R}^d$.

1. *Construct an $n \times n$ matrix M with $M_{ij} = K(x_i, x_j)$ derived from the distance or affinity between each pair of points.*
2. *Optionally transform M, yielding a "normalized" matrix $\tilde{M}$.*
3. *Compute the m ($m << d$) largest positive eigenvalues λ and eigenvectors v of $\tilde{M}$.*
4. *The embedding of each sample x_i, $i = \{1, \ldots, n\}$, is the vector y_i with y_{ik} the i-th element of the k-th principal eigenvector v_k, $k = \{1, \ldots, m\}$, of $\tilde{M}$. Alternatively(for some algorithms), the embedding is e_i, with $e_{ik} = \sqrt{\lambda_k} y_{ik}$.* ■

In Sec. 2.2.1 we introduced how to compute the Riemannian distance between images. A distance matrix is constructed after the distance between each pair of training images is calculated. Embedding algorithms then project all the labeled images onto a low-dimensional space with this distance matrix and a discrimination hyperplane on the embedded space will be trained using the labeled images. To classify a new unlabeled image, an out-of-sample extension of embedding algorithms is used to project the new sample onto the constructed embedded space.

a. Spectral embedding

The spectral methods for clustering are based on major eigenvalues of an $n \times n$ matrix derived from the **affinity** between labeled subjects, where n denotes the number of labeled subjects in the dataset. The eigenvectors induce an embedding of the subjects onto a low-dimensional space. In spectral embedding, the affinity matrix is formed from the distance matrix using a Gaussian kernel with parameter σ. On the embedded space each training image is represented as a point, and nearby points correspond to images of high similarities. The advantage of this space is that similarity relationships are represented spatially, so that clusters can be more easily found.

Let S_i be the i-th row sum of the **affinity matrix** M:

$$S_i = \sum_j M_{ij}. \tag{2.15}$$

The normalized Laplacian matrix is formed by

$$\tilde{M}_{ij} = \frac{M_{ij}}{\sqrt{S_i S_j}}. \tag{2.16}$$

For each training image x_i, the embedding coordinate on an $\mathbb{R}^m$ space is y_{ik}, $k = \{1, \ldots, m\}$.

Extending spectral embedding

Let $\tilde{K}(a, b)$ be a kernel function that gives rise to a symmetric matrix $\tilde{M}$ with entries $\tilde{M} = \tilde{K}(x_i, x_j)$ on a dataset $D = \{x_1, \ldots, x_n\}$:

$$\tilde{K}(a, b) = \frac{1}{n} \frac{K(a, b)}{\sqrt{E_x \left[K(a, x)\right] E_{x'} \left[K(b, x')\right]}}, \tag{2.17}$$

where $K(a, b)$ denotes the affinity between a and b, and the expectations are taken over the training data D.

Let (v_k, λ_k) be an $(eigenvector, eigenvalue)$ pair that solves $\tilde{M}v_k = \lambda_k v_k$ and $y_k(x)$ denote the embedding associated with a new point x. Then

$$y_k(x) = \frac{1}{\lambda_k} \sum_{i=1}^{n} v_{ki} \tilde{K}(x, x_i). \tag{2.18}$$

Readers can refer to the work of Bengio et al. [44, 48] for more details and a proof.

b. Metric multi-dimensional scaling

Metric Multi-Dimensional Scaling (MDS) transforms a **distance** matrix, which is computed on a set of training images, to coordinates for each subject such that the Euclidean distances derived from these coordinates approximate as well as possible the original distance [47]. The idea of metric MDS is to transform the distance matrix into a cross-product matrix and then to find its eigen-decomposition which gives a principal component analysis (PCA).

Let S_i be the i-th row sum of the **distance matrix** M, $S_i = \sum_j M_{ij}$. The cross-product matrix is obtained by using the "double-centering" formula [44]:

$$\tilde{M}_{ij} = -\frac{1}{2}(M_{ij} - \frac{1}{n}S_i - \frac{1}{n}S_j + \frac{1}{n^2}\sum_k S_k). \tag{2.19}$$

The embedding e_{ik} of sample x_i is $\sqrt{\lambda_k}v_{ki}$, $k = \{1, \ldots, m\}$.

Extending metric MDS

The kernel function $\tilde{K}$ yielding the symmetric matrix $\tilde{M}$ on the dataset D is defined as:

$$\tilde{K}(a, b) = -\frac{1}{2}(d^2(a, b) - E_x[d^2(x, b)] - E_{x'}[d^2(a, x')] + E_{x,x'}[d^2(x, x')]), \tag{2.20}$$

where $d(a, b)$ is the original distance between a and b, and the expectations are taken over the training data D.

Let (v_k, λ_k) be an $(eigenvector, eigenvalue)$ pair that solves $\tilde{M}v_k = \lambda_k v_k$ and e_k denote the

embedding associated with a new point x. Then

$$e_k(x) = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^{n} v_{ki} \tilde{K}(x, x_i). \tag{2.21}$$

See also the work of Bengio et al. [44, 48] for details and a proof.

c. Isomap

Isomap generalizes MDS to non-linear manifolds. They are based on replacing Euclidean distance by an approximation of the geodesic distance on the manifold. The geodesic distance is defined as

$$\tilde{D}(x_i, x_j) = \min_p \sum_r d(p_r, p_{r+1}), \tag{2.22}$$

where p is a sequence of points of length $l \geq 2$ with $p_1 = x_i$, $p_l = x_j$, $p_r \in D \forall r \in 2, \ldots, l-1$ and (p_r, p_{r+1}) are k-nearest-neighbor. The length l is a free parameter. Then the normalization matrix is calculated by

$$\tilde{M}_{ij} = -\frac{1}{2}(\tilde{D}^2(i,j) - \frac{1}{n} S_i - \frac{1}{n} S_j + \frac{1}{n^2} \sum_k S_k). \tag{2.23}$$

where S_i denotes the i-th row sum of the geodesic **distance matrix** $\tilde{D}^2$, $S_i = \sum_j \tilde{D}^2_{ij}$. The embedding is $e_{ik} = \sqrt{\lambda_k} v_{ik}$.

Extending Isomap

The kernel function $\tilde{K}$ is defined as:

$$\tilde{K}(a, b) = E_{x'}[\tilde{D}^2(x', a)] - \tilde{D}^2(a, b), \tag{2.24}$$

where $\tilde{D}(a, b)$ is the geodesic distance between a and b, and the expectations are taken over the training data D.

Let (v_k, λ_k) be an $(eigenvector, eigenvalue)$ pair that solves $\tilde{M} v_k = \lambda_k v_k$ and e_k denote the embedding associated with a new point x. Then

$$e_k(x) = \frac{1}{2\sqrt{\lambda_k}} \sum_{i=1}^{n} v_{ki} \tilde{K}(x_i, x). \tag{2.25}$$

d. Local linear embedding (LLE)

The LLE algorithm looks for an embedding that preserves the local geometry in the neighborhood of each data point. First a sparse matrix of local predictive weights W_{ij} is computed, such that $\sum_j W_{ij} = 1$, $W_{ij} = 0$ if x_j is not a k-nearest-neighbor of x_i and $\sum_j (W_{ij}x_j - x_i)^2$ is minimized. Then the normalized matrix is calculated as

$$\tilde{M} = I - (I - W')(I - W). \tag{2.26}$$

The embedding is given by $y_{ik} = v_{ik}$.

Extending LLE

To obtain a kernel function we must first associate a function $w(a, b)$ to the matrix W of regression weights such that we obtain $W_{ij} = w(x_i, x_j)$. A local $k \times k$ Gram matrix around an arbitrary point x is defined:

$$C(x)_{ij} = (x - x_{n(i)})(x - x_{n(j)})' 1_{x_{n(i)} \in \mathcal{N}(x)} 1_{x_{n(j)} \in \mathcal{N}(x)}, \tag{2.27}$$

where $\mathcal{N}(x)$ is the subset of k elements from D that are the k nearest neighbors of x and $n(i)$ is the index of the i-th such neighbor of x. Then $w(a, b)$ is defined as follows:

$$w(a, b) = 1_{b = x_{n(j)} \in \mathcal{N}(a)} \frac{\sum_q C^{-1}(x)_{jq}}{\sum_{pq} C^{-1}(x)_{pq}}. \tag{2.28}$$

The kernel function is then defined as

$$\tilde{K}(a, b) = w(a, b) + w(b, a) - \sum_i w(x_i, a) w(x_i, b). \tag{2.29}$$

Then

$$e_k(x) = \sum_{i=1}^{n} v_{ki} \tilde{K}(x, x_i). \tag{2.30}$$

In this work, we use two of the existing embedding algorithms: the spectral embedding [45, 46] and the metric multi-dimensional scaling [47]. The idea of metric MDS is to minimize the error between the real locations of data and the locations estimated. The practical advantage of metric MDS is that it provides experimenter with the clearest and most traceable relations between the original data and the test solution by maintaining a simple linear re-

lation between them. The spectral embedding is however a non-linear dimension reduction method. Relationships within sets of non-linear data are sometimes mis-rendered into a low dimensional space by linear methods. Nonlinear methods, in contrast, attempt to model important aspects of the underlying data structure, often requiring parameters fitting to the data type of interest. In our experiments, we performed a linear embedding and a non-linear embedding respectively, and compare the classification results.

e. An example on comparison between spectral embedding and metric MDS

A simple example on embedding 10 US cities based on airline distances is used to compare the spectral embedding and metric MDS.

The distances are listed in the Table 2.1 and they are used as dissimilarities between cities.

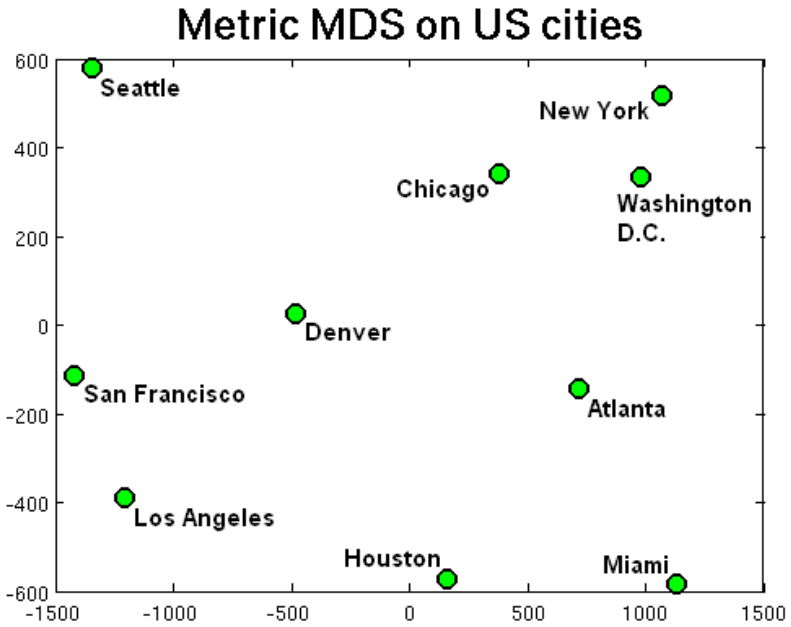
Table 2.1: Airline distances between 10 US cities

		1 Atlanta	2 Chicago	3 Denver	4 Houston	5 LA	6 Miami	7 NY	8 SF	9 Seattle	10 DC
1	Atlanta	0	587	1212	701	1936	604	748	2139	2182	543
2	Chicago	587	0	920	940	1745	1188	713	1858	1737	597
3	Denver	1212	920	0	879	831	1726	1631	949	1021	1494
4	Houston	701	940	879	0	1374	968	1420	1645	1891	1220
5	LA	1936	1745	831	1374	0	2339	2451	347	959	2300
6	Miami	604	1188	1726	968	2339	0	1092	2594	2734	923
7	NY	748	713	1631	1420	2451	1092	0	2571	2408	205
8	SF	2139	1858	949	1645	347	2594	2571	0	678	2442
9	Seattle	2182	1737	1021	1891	959	2734	2408	678	0	2329
10	DC	543	597	1494	1220	2300	923	205	2442	2329	0

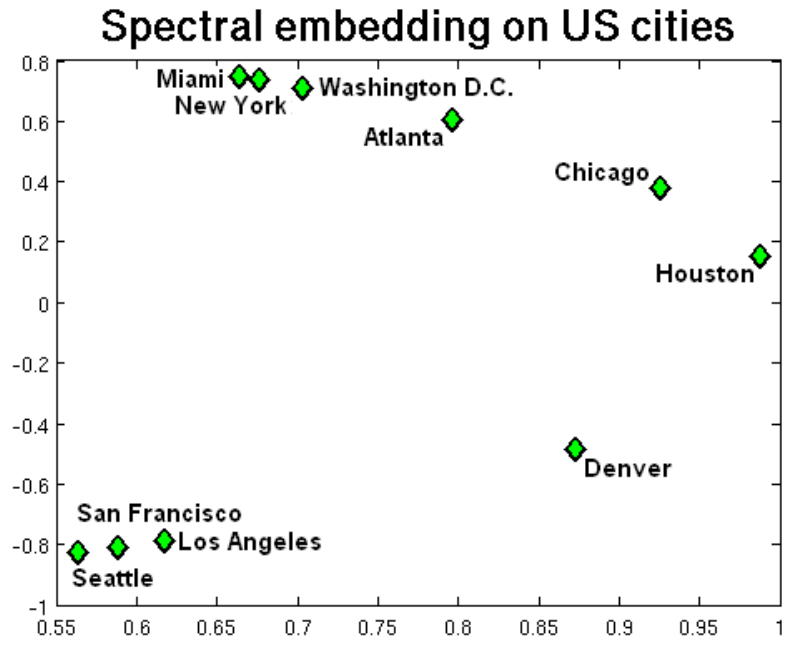
The metric MDS method directly uses the distance matrix and produces the embedding as shown in Figure 2.4 (a). In spectral embedding, the distance matrix is first converted to an affinity matrix using the Gaussian kernel:

$$a_{ij} = exp(-d_{ij}/\sigma^2), \tag{2.31}$$

where d_{ij} is the distance between the i-th and j-th cities, $i, j = \{1, \ldots, 10\}$, and σ is a free parameter. In this example, σ is set as 5000. The embedding using the spectral method is shown in Figure 2.4 (b).



(a)



(b)

Figure 2.4: Comparison on metric MDS and spectral embedding.

Comparing the embeddings of the spectral and the MDS, we found the different ideas between these two methods. The metric MDS aims to project the samples as close as possible to their real positions thus the distance between samples on the embedded space is as close as possible to the real distance. However, the spectral embedding strengthens the similarity within the underlying cluster and the dissimilarity between clusters through the free parameter σ in the Gaussian kernel. In this example, the grouping of Seattle, San Francisco and Los Angeles is more compact in the spectral embedding since these cities are all considered to be the west coastal cities in the algorithm and similarly for those east coastal cities. The three cities away from coastal regions, Denver, Houston and Chicago, are dispersed on the embedded space.

In our application, images will be treated like the cities used in this example. Pair-wise distances between images will be calculated as in Table 2.1. Images will then be projected onto a low-dimensional Euclidean space and be classified.

2.2.3 Algorithm summation

Given a set of images $I_1, I_2, \ldots, I_N$, pair-wise registration is performed using the symmetric log-domain diffeomorphic demons algorithm with default parameters [38]:

- -f —fixed-image <filename>
- -m —moving-image <filename>
- -o —output-image <filename>
- -O —output-field <filename>
- -i —number-of-iteration <15x10x5>
- -s —deformation-field-sigma-in-the-*diffusion*-like-kernel <1.5>
- -g —update-field-sigma-in-the-*fluid*-like-kernel <0>
- -l —max-step-length <2.0>
- -a —update-rule <*log-domain diffeomorphic*>
- -t —gradient-type <0:symmetrized>
- -e —use-histogram-matching

Each diffeomorphism output from registration is used to compute the regional distances. The distance matrix is constructed after pair-wise distance is calculated, with zero diagonal

entries. The embedding methods are then applied to project all the images onto an $\mathbb{R}^m$ space, where leave-one-out cross validation is employed to evaluate the classification performance. In this work we compared the results with m selected from 2 to 15. We found almost no differences when m is larger than 3 and a slightly worse result when $m = 2$. Thus, we chose $m = 3$ in this dissertation for a lower computational cost and an easier visualization of results.

The entire algorithm is summarized as follows.

Algorithm (Classification Algorithm)

Given a set of images $\{I_1, I_2, \ldots, I_N\}$,

- *Compute the pair-wise registration using the symmetric log-domain diffeomorphic demons algorithm;*
- *Compute the pair-wise distance and construct the distance matrix;*
- *All images are projected onto an* $\mathbb{R}^m$ *subspace using the spectral embedding method and the metric MDS embedding method with manually selected* $m = 3$*;*
- *The leave-one-out cross validation is performed in the constructed subspace. The classification accuracy along with the sensitivity, specificity, positive/negative predictive values are calculated (the evaluation method will be introduced in the following section).*

2.3 Experiments and results

2.3.1 Data

Data used in this paper are selected from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI) and the Open Access Series of Imaging Studies (OASIS) database [49].

"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 OASIS is a project aimed at making MRI data sets of the brain freely available to the scientific community. By complying and freely distributing MRI data sets, we hope to facilitate future discoveries in basic and clinical neuroscience. OASIS is made available by the Washington University Alzheimer's Disease Research Center, Dr. Randy Buckner at the Howard Hughes Medical Institute (HHMI) at Harvard University, the Neuroinformatics Research Group (NRG) at Washington University School of Medicine, and the Biomedical Informatics Research Network (BIRN)."

2.3.2 MRI acquisition and experiment datasets

The MR scans are T1-weighted MR images. For the ADNI database, all the volumes were acquired on a 1.5-T scanner. MRI acquisition had been done according to the ADNI acquisition protocol [105]. The project combines data from 55 participating sites and includes MRI data from 3 different vendors (GE Healthcare, Philips Medical Systems, or Siemens Medical Solutions, http://www.loni.ucla.edu/ADNI/Data/ADNI_Data.shtml). The detailed protocol for each scanner is publicly available (http://www.loni.ucla.edu/ADNI/Research/Cores/). A sophisticated system for quality control including phantom scanning is in place [106, 107]. The target voxel size is approximately 1 mm^3, with a maximum of 1.5 mm in any one direction.

For the OASIS database, scans were obtained from a single acquisition site [49]. Multiple (three or four) high-resolution structural T1-weighted magnetizationprepared rapid gradient echo (MP-RAGE) images were acquired on a 1.5-T Vision scanner (Siemens, Erlangen, Germany) in a single imaging session. Image parameters are: TR= 9.7 msec., TE= 4.0 msec., Flip angle= 10, TI= 20 msec., TD= 200 msec., 128 sagittal 1.25 mm slices without gaps and pixels resolution of 256×256 ($1 \times 1mm$).

Before used in our experiments, the images in the ADNI database were geometry distortion corrected, intensity nonuniformity corrected, and histogram peak sharpened. We additionally processed the images with an affine registration to a common space. The images in the

OASIS were provided skull-stripped, bias corrected, and affinely registered to the Talairach atlas space.

We constructed five experiment datasets from the ADNI and OASIS database:

The first dataset consists of 77 T1-weighted MR scans from the ADNI, with 33 AD patients (aged 58.75∼88.84) and 44 age-matched normal controls (aged 59.98∼89.67). This dataset is referred to as "ADNI (NC *vs* AD)" in the following.

The second dataset is another subset of the ADNI database, with 39 brains from MCI subjects (aged 56.25∼87.81) which later had progressed to AD and 44 age-matched normal controls (aged 59.98∼89.67) which are the same set of subjects used in the first dataset. This dataset is referred to as "ADNI (NC *vs* cMCI[5])" in the following.

The third dataset consists of 75 MR scans from the OASIS database, with 35 brains from patients with very mild to mild dementia (aged 69∼96) and 40 age-matched normal controls (aged 67∼91). This dataset is referred to as "OASIS (NC *vs* mAD[6])" in the following.

The fourth dataset consists of 17 MR scans from the OASIS database, with 10 normal controls (aged 61∼80) and 7 brains from potential diseased subjects (aged 69∼87) which were healthy at the time of scanning but had developed to MCI after 4±1.41 years. This dataset is referred to as "OASIS (NC *vs* pMCI[7])" in the following.

The fifth dataset consists of 38 MR scans from the ADNI database, with 19 patients with MCI (aged 57.92∼87.81) who had not converted to AD within 2.81±0.61 years, and 19 patients with MCI (aged 56.25∼87.24) who had converted to AD in 1.88±0.62 years. This dataset is referred to as "ADNI (sMCI[8] *vs* cMCI)" in the following.

The diseased subjects were clinically diagnosed using the Clinical Dementia Rating (CDR) scale, which combines the information from mental examination, patient history and interviews with family members or caregivers [50]. This score is used to characterize and to track a patient's level of impairment: 0 = Normal, 0.5 = Very Mild Dementia, 1 = Mild Dementia, 2 = Moderate Dementia, 3 = Severe Dementia.

Age information and diagnosis characteristics on the selected subjects are listed in Table 2.2.

[5]cMCI = converting MCI
[6]mAD = mild AD
[7]pMCI = potential MCI
[8]sMCI = stable MCI

Table 2.2: Age and diagnosis characteristics of the five datasets

Datasets		# of subjects	Age (range)	MMSE (range)	CDR scale at scanning	Latter CDR scale
Group 1 **ADNI (NC *vs* AD)**	control	44	76.06 (59.98∼89.67)	28.07 (24∼30)	0	0
	AD	33	77.66 (58.75∼88.84)	15.64 (2∼26)	1∼3	1∼3
Group 2 **ADNI (NC *vs* cMCI)**	control	44	76.06 (59.98∼89.67)	28.07 (24∼30)	0	0
	cMCI	39	75.48 (56.25∼87.81)	20.97 (5∼30)	0.5	1∼2
Group 3 **OASIS (NC *vs* mAD)**	control	40	77.28 (67∼91)	28.98 (26∼30)	0	0
	mAD	35	80.37 (69∼96)	22.63 (15∼29)	0.5∼1	0.5∼1
Group 4 **OASIS (NC *vs* pMCI)**	control	10	73.60 (61∼80)	29.30 (28∼30)	0	0
	pMCI	7	78.43 (69∼87)	29.28 (27∼30)	0	0.5
Group 5 **ADNI (sMCI *vs* cMCI)**	sMCI	19	75.16 (57.92∼87.81)	26.95 (24∼29)	0.5	0.5
	cMCI	19	77.36 (56.25∼87.24)	26.05 (24∼29)	0.5	1∼2

NC=Normal Control; AD=Alzheimer's Disease; cMCI=converting Mild Cognitive Impairment (MCI subjects which later progressed to AD); mAD=mild Alzheimer's Disease; pMCI=potential Mild Cognitive Impairment (normal subjects which later progressed to MCI); sMCI=stable Mild Cognitive Impairment (MCI patients which had not converted to AD); MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

2.3.3 Experiment design and parameter setting

The segmentations of gray matter, white matter, and CSF were obtained using the FreeSurfer [34] software. Pair-wise registration was done using the symmetric log-domain diffeomorphic demons algorithm with default parameters (the details were presented in Section 2.2.3), providing deformation fields and logarithms of deformation fields (using "log fields" for shorthand in the following). Riemannian distance between images was measured from the log field using the equation (2.14) described in Section 2.2.1. In this dissertation, the dimension of the embedded space was selected as 3. Thus all images were projected onto an $\mathbb{R}^3$ Euclidean space, using the spectral embedding and the metric MDS respectively. A discriminant hyperplane was then trained on the embedded space using the labeled subjects. We implemented five different learning algorithms: linear discriminant, Mahalanobis distance, support vector machine (SVM) with a linear kernel, SVM with a non-linear kernel and k nearest neighbor (knn) classification. The unlabeled subject was projected onto the constructed space using the out-of-sample embedding. The leave-one-out cross validation was adopted to evaluate the performance of each embedding algorithm and classification with different regional distances.

Six experiments were conducted for this work. In the first five experiments, the classification was performed on the five datasets respectively. We used three different distance metrics: the gray matter (GM) distance, the white matter (WM) distance and the ventricle (VENT) distance. The GM distance was calculated from the log field of gray matter, which was the overlapped region between the whole-brain diffeomorphism and the gray matter segmentation. Similarly, the WM/VENT distance was calculated from the log field of white matter/ventricles, which was the overlapped region between the whole-brain diffeomorphism and the white matter/ventricle segmentation. In addition, the distances on hippocampus and the entire temporal lobe were calculated for comparison.

In the sixth experiment, the generalized classification power of the proposed algorithm was tested. We first treated the 75 images from Group 3 (OASIS NC *vs* mAD) as the training data and the 77 images from Group 1 (ADNI NC *vs* AD) as the test data. Then we reversed the training dataset and test dataset. The spectral embedding method and the linear discriminant analysis were used for classification in this experiment.

2.3.4 Evaluation

From **Experiment 1 to 5**, the leave-one-out cross validation (LOOCV) was used to estimate the performance of classification. The LOOCV involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data. The LOOCV is a standard method that is used in the other papers and so is the best cross-validation for comparison.

In our LOOCV, each time the data set is divided into 2 subsets, one with $n-1$ subjects as the training group and the other with 1 subject as the test group. For the training group, pair-wise registration and pair-wise distance are computed over the group. Embedding coordinates of each training subject are calculated using the main embedding algorithm to train a classifer. For the test group, the registration and distance between the test subject and all training subjects are computed. The embedding coordinates of the test subject are calculated using out-of-sample embedding and the test subject is classified to a group based on the trained classifier. The procedure was repeated until each input subject was used as the test sample once. The proportion of correct classification is calculated.

In **Experiment 6**, data are divided into two groups according to the database they belong

to (OASIS or ADNI). Each group is composed of both normal control brains and AD brains. One group is used for training and the other is used for validation.

		Ground Truth		
		Positive	Negative	
Experiment Observation	Positive	True Positive (TP)	False Positive (FP)	→ *Positive predictive value*
	Negative	False Negative (FN)	True Negative (TN)	→ *Negative predictive value*
		↓ *Sensitivity*	↓ *Specificity*	

Figure 2.5: Evaluation terms and their relationships.

For each cross validation method, we computed the true positives (TP) - defined as the number of diseased subjects correctly assigned to the appropriate class, the true negatives (TN) - defined as the number of normal subjects correctly assigned to the appropriate class, the false positives (FP) - defined as the number of normal subjects misclassified by the classifier, and the false negatives (FN) - defined as the number of diseased subjects misclassified by the classifier. We then computed the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy as:

$$Sensitivity = \frac{TP}{TP+FN}; \tag{2.32}$$

$$Specificity = \frac{TN}{TN+FP}; \tag{2.33}$$

$$Positive\ predictive\ value = \frac{TP}{TP+FP}; \tag{2.34}$$

$$Negative\ predictive\ value = \frac{TN}{TN+FN}; \tag{2.35}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}. \tag{2.36}$$

Sensitivity, specificity, PPV, and NPV, collectively known as "test characteristics", are important ways to express the usefulness of diagnostic tests. Sensitivity measures the proportion of actual patients which are correctly identified as such. Specificity measures the proportion of normal subjects which are correctly identified. PPV is the proportion of patients with positive test results who are correctly diagnosed. NPV is the proportion of normal subjects with negative test results who are correctly diagnosed. The following example provides an easy way to clarify the concepts of sensitivity, specificity, PPV and NPV. Assuming

that we have 4 patients, for the first 2 we know only their disease status; for the last 2 patients we know only their test results [122].

We know the patient's disease status:

- Sensitivity: "I know my patient has the disease. What is the chance that the test will show that my patient has it?"
- Specificity: "I know my patient doesn't have the disease. What is the chance that the test will show that my patient doesn't have it?"

We have just gotten a test result and do not know the patient's disease status:

- PPV: "I just got a positive test result back on my patient. What is the chance that my patient actually has the disease?"
- NPV: "I just got a negative test result back on my patient. What is the chance that my patient actually doesn't have the disease?"

Sensitivity and specificity are fixed for a particular type of test. However, for a particular type of test PPV and NPV depend upon the prevalence of a disease in a population [122]. For example, though current screening tests for HIV have high sensitivity and specificity, the low prevalence of HIV in the general population cannot justify universal screening since the majority of positive tests would be falsely positive (i.e., low PPV).

2.3.5 Results

Experiment 1: Classification on the ADNI (NC *vs* AD) dataset

The spectral and metric MDS embedding methods along with 5 classifiers construct different combinations for classification. The results are shown in Table 2.3.

The best classification rates for each regional distance and each embedding method were taken out for comparison (Fig.2.6). Among the five regional distances used in this experiment, the **temporal-lobe** distance was observed giving the highest accuracy of 98.70%. The **GM** distance, the **WM** distance and the **hippocampal** distance using both embedding methods also showed good discrimination performance, yielding correct rates around 95%. The worst classification came from the **ventricle** distance using the MDS embedding method, with a correct rate of 88.31%.

Table 2.3: Classification results on the ADNI NC *vs* AD dataset

COMBINATIONS		SPECTRAL EMBEDDING					MDS EMBEDDING				
		Accuracy	SEN*	SPE	PPV	NPV	Accuracy	SEN*	SPE	PPV	NPV
GM	Linear	**96.10%**	96.97%	95.45%	94.12%	97.67%	**96.10%**	96.97%	95.45%	94.12%	97.67%
	Mahalanobis	90.91%	81.82%	97.33%	96.43%	87.76%	93.51%	90.91%	95.45%	93.75%	93.33%
	SVM linear	90.91%	84.85%	95.45%	93.33%	89.36%	93.51%	90.91%	95.45%	93.75%	93.33%
	SVM non-linear	96.10%	96.97%	95.45%	94.12%	97.67%	94.81%	93.94%	95.45%	93.94%	95.45%
	Knn	94.81%	93.94%	95.45%	93.94%	95.45%	94.81%	96.97%	93.18%	91.43%	97.62%
WM	Linear	**96.10%**	96.97%	95.45%	94.12%	97.67%	96.10%	96.97%	95.45%	94.12%	96.97%
	Mahalanobis	93.51%	90.91%	95.45%	93.75%	93.33%	**97.40%**	100%	95.45%	94.29%	100%
	SVM linear	93.51%	90.91%	95.45%	93.75%	93.33%	96.10%	96.97%	95.45%	94.12%	97.67%
	SVM non-linear	90.91%	78.79%	100%	100%	86.27%	96.10%	96.97%	95.45%	94.12%	97.67%
	Knn	96.10%	96.97%	95.45%	94.12%	97.67%	97.40%	100%	95.45%	94.29%	100%
VENT	Linear	**94.81%**	90.91%	97.33%	96.77%	93.48%	**88.31%**	90.91%	86.36%	83.33%	92.68%
	Mahalanobis	81.82%	60.61%	97.33%	95.24%	76.79%	87.01%	75.76%	95.45%	92.59%	84.00%
	SVM linear	89.61%	78.79%	97.33%	90.91%	93.18%	84.42%	84.85%	84.09%	93.94%	95.45%
	SVM non-linear	92.21%	84.85%	97.33%	96.55%	89.58%	85.71%	87.88%	84.09%	80.56%	90.24%
	Knn	90.91%	81.82%	97.33%	96.43%	87.76%	88.31%	90.91%	86.36%	83.33%	92.68%
HIPPO	Linear	93.51%	93.94%	93.18%	91.18%	95.35%	**94.81%**	93.94%	95.45%	93.94%	95.45%
	Mahalanobis	92.21%	84.85%	97.33%	96.55%	89.58%	94.81%	93.94%	95.45%	93.94%	95.45%
	SVM linear	90.91%	90.91%	93.18%	90.91%	93.18%	93.51%	93.94%	95.45%	93.94%	95.45%
	SVM non-linear	90.91%	87.88%	93.18%	90.62%	91.11%	93.51%	90.91%	95.45%	93.75%	93.33%
	Knn	**96.10%**	96.97%	95.45%	94.12%	97.67%	94.81%	93.94%	95.45%	93.94%	95.45%
TEMPO	Linear	**98.70%**	100%	97.33%	97.06%	100%	97.40%	96.97%	97.33%	96.97%	97.73%
	Mahalanobis	93.51%	87.88%	97.33%	96.67%	91.49%	96.10%	93.94%	97.33%	96.88%	95.66%
	SVM linear	93.51%	87.88%	97.33%	96.67%	91.49%	96.10%	93.94%	97.33%	96.88%	95.56%
	SVM non-linear	94.81%	93.94%	95.45%	93.94%	95.45%	**98.70%**	100%	97.33%	97.06%	100%
	Knn	98.70%	100%	97.33%	97.06%	100%	97.40%	100%	95.45%	94.29%	100%

∗ Considering a correctly identified AD case as a true positive; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value; GM=Gray Matter; WM=White Matter; VENT=Ventricle; HIPPO=Hippocampus; TEMPO=Temporal.

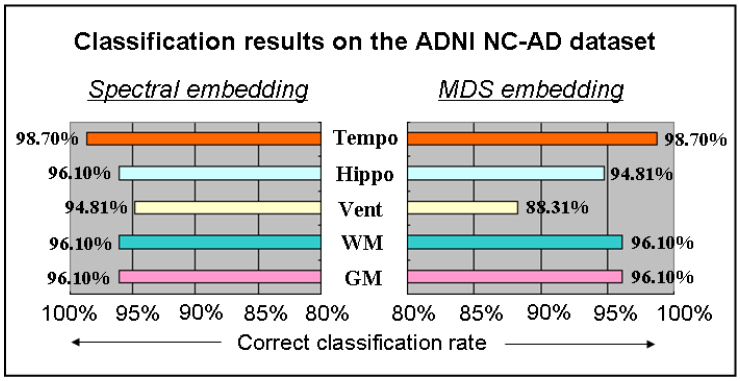


Figure 2.6: Classification results on the ADNI NC *vs* AD dataset. Vent=Ventricle; Hippo= Hippocampus; Tempo=Temporal lobe.

Experiment 2: Classification on the ADNI (NC *vs* cMCI) dataset

The combinations of the spectral and metric MDS embedding methods along with 5 classifiers were used for classification. The results are shown in Table 2.4.

The best classification rates for each regional distance and each embedding method again were taken out for comparison (Fig.2.7). The **ventricle** distance in this experiment was

Table 2.4: Classification results on the **ADNI NC *vs* cMCI** dataset

COMBINATIONS		SPECTRAL EMBEDDING					MDS EMBEDDING				
		Accuracy	SEN*	SPE	PPV	NPV	Accuracy	SEN*	SPE	PPV	NPV
GM	Linear	**92.77%**	94.87%	90.91%	90.24%	95.24%	**91.57%**	94.87%	88.64%	88.10%	95.12%
	Mahalanobis	90.36%	92.31%	88.64%	87.80%	92.86%	90.36%	89.74%	90.91%	89.74%	90.91%
	SVM linear	90.36%	92.31%	88.64%	87.80%	92.86%	91.57%	94.87%	88.64%	88.10%	95.12%
	SVM non-linear	92.77%	94.87%	90.91%	90.24%	95.24%	89.16%	94.87%	84.09%	84.09%	94.87%
	Knn	91.57%	94.87%	88.64%	88.10%	95.12%	91.57%	97.44%	86.36%	86.36%	97.44%
WM	Linear	**92.77%**	97.44%	88.64%	88.37%	97.50%	84.54%	84.62%	84.09%	82.50%	86.05%
	Mahalanobis	87.95%	84.62%	90.91%	89.19%	86.96%	**93.98%**	92.31%	95.45%	94.74%	93.33%
	SVM linear	91.57%	94.87%	88.64%	88.10%	95.12%	85.54%	84.62%	86.36%	84.62%	86.36%
	SVM non-linear	90.36%	87.18%	93.18%	91.89%	89.13%	91.57%	89.74%	93.18%	92.11%	91.11%
	Knn	92.77%	94.87%	90.91%	90.24%	95.24%	89.16%	87.18%	90.91%	89.47%	88.89%
VENT	Linear	**96.39%**	97.44%	95.45%	95.00%	97.67%	87.95%	79.49%	77.27%	93.94%	84.00%
	Mahalanobis	95.18%	94.87%	95.45%	94.87%	95.45%	95.18%	94.87%	95.45%	94.87%	95.45%
	SVM linear	96.39%	97.44%	95.45%	95.00%	97.67%	86.75%	76.92%	95.45%	93.75%	91.11%
	SVM non-linear	96.39%	97.44%	95.45%	95.00%	97.67%	91.57%	89.74%	93.18%	93.75%	91.11%
	Knn	96.39%	97.44%	95.45%	95.00%	97.67%	**96.39%**	94.87%	97.73%	97.37%	95.56%
HIPPO	Linear	**78.31%**	79.49%	77.27%	75.61%	80.95%	84.34%	84.62%	84.09%	82.50%	86.05%
	Mahalanobis	75.90%	76.92%	75.00%	73.17%	78.57%	84.24%	87.18%	81.82%	80.95%	86.67%
	SVM linear	78.31%	79.49%	77.27%	75.61%	80.49%	**86.75%**	84.62%	88.64%	86.84%	86.05%
	SVM non-linear	77.11%	79.49%	75.00%	73.81%	80.49%	84.34%	84.62%	84.09%	82.50%	86.05%
	Knn	78.31%	84.62%	72.73%	73.33%	84.21%	85.54%	89.74%	81.82%	81.40%	90.00%
TEMPO	Linear	91.57%	97.44%	86.36%	86.36%	97.44%	**87.95%**	89.74%	86.36%	85.37%	90.48%
	Mahalanobis	**92.77%**	97.44%	88.64%	88.37%	97.50%	87.95%	87.18%	88.64%	87.18%	88.64%
	SVM linear	92.77%	97.44%	88.64%	88.37%	97.44%	87.95%	89.74%	86.36%	85.37%	88.10%
	SVM non-linear	91.57%	97.44%	86.36%	86.36%	97.44%	85.54%	87.18%	84.09%	82.93%	88.10%
	Knn	91.57%	97.44%	86.36%	86.36%	97.44%	87.95%	89.74%	86.36%	85.37%	90.48%

∗ Considering a correctly identified converting MCI case as a true positive; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value; GM=Gray Matter; WM=White Matter; VENT=Ventricle; HIPPO=Hippocampus; TEMPO=Temporal.

observed giving the highest accuracy of 96.39% in both spectral embedding and MDS embedding. The **GM** distance, **WM** distance and **temporal-lobe** distance were slightly worse than the ventricle distance, yielding rates around 91%. While the spectral embedding method worked relatively better with the **GM** distance and **temporal-lobe** distance, the MDS embedding worked better with the **WM** distance. The worst classification was from the **hippocampal** distance, which gave an accuracy around 80%.

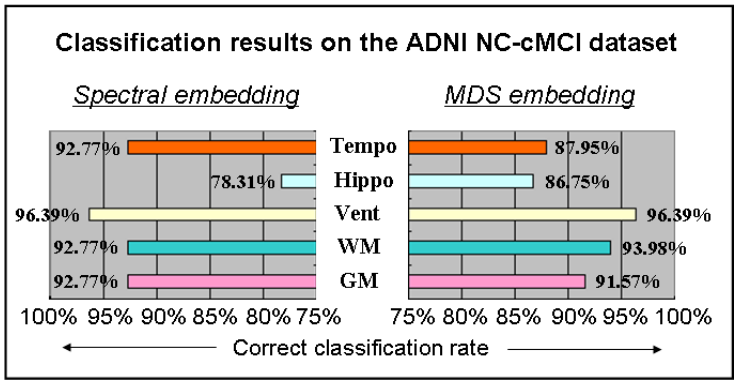


Figure 2.7: Classification results on the ADNI NC *vs* cMCI dataset. Vent=Ventricle; Hippo=Hippocampus; Tempo=Temporal lobe.

Experiment 3: Classification on the OASIS (NC *vs* mAD) dataset

The same experiments were performed on the OASIS datasets. The classification results are shown in Table 2.5, with different combinations of embedding methods and learning algorithms.

Table 2.5: Classification results on the **OASIS NC *vs* mAD** dataset

COMBINATIONS		SPECTRAL EMBEDDING			MDS EMBEDDING						
		Accuracy	SEN*	SPE	PPV	NPV	Accuracy	SEN*	SPE	PPV	NPV
GM	Linear	**97.33%**	100%	95.00%	94.59%	100%	**93.33%**	97.14%	90.00%	89.47%	97.30%
	Mahalanobis	82.67%	100%	67.50%	72.92%	100%	92.00%	91.43%	92.50%	91.43%	92.50%
	SVM linear	93.33%	85.71%	100%	100%	88.89%	93.33%	91.43%	95.00%	94.12%	92.68%
	SVM non-linear	92.00%	88.57%	95.00%	93.94%	90.48%	89.33%	91.43%	87.50%	86.49%	92.11%
	Knn	96.00%	94.29%	97.50%	97.06%	95.12%	93.33%	94.29%	92.50%	91.67%	94.87%
WM	Linear	**96.00%**	94.29%	97.50%	97.06%	95.12%	**96.00%**	94.29%	97.50%	97.06%	95.12%
	Mahalanobis	88.00%	85.71%	90.00%	88.24%	87.80%	96.00%	94.29%	97.50%	97.06%	95.12%
	SVM linear	96.00%	94.29%	97.50%	97.06%	95.12%	96.00%	94.29%	97.50%	97.06%	95.12%
	SVM non-linear	82.67%	68.57%	95.00%	92.31%	77.55%	94.67%	94.29%	95.00%	94.29%	95.00%
	Knn	94.67%	94.29%	95.00%	94.29%	95.00%	94.67%	94.29%	95.00%	94.29%	95.00%
VENT	Linear	92.00%	88.57%	95.00%	94.29%	90.48%	**89.33%**	97.14%	82.50%	82.93%	97.06%
	Mahalanobis	93.33%	88.57%	97.50%	96.88%	90.70%	84.00%	74.29%	92.50%	89.66%	80.43%
	SVM linear	85.33%	71.43%	97.50%	96.15%	79.59%	81.33%	85.71%	85.00%	81.82%	80.95%
	SVM non-linear	92.00%	90.00%	94.29%	93.94%	90.48%	82.67%	77.50%	88.57%	86.67%	80.00%
	Knn	**93.33%**	91.43%	95.00%	94.12%	92.68%	89.33%	91.43%	87.50%	86.49%	92.11%
HIPPO	Linear	**90.67%**	88.57%	92.50%	91.18%	90.24%	82.67%	82.86%	82.50%	80.56%	84.62%
	Mahalanobis	84.00%	74.29%	92.50%	89.66%	80.43%	**84.00%**	80.00%	87.50%	84.85%	83.33%
	SVM linear	86.67%	80.00%	92.50%	90.32%	84.09%	81.33%	80.00%	82.50%	80.00%	82.50%
	SVM non-linear	85.33%	85.71%	92.50%	90.00%	82.22%	81.33%	82.86%	80.00%	78.38%	84.21%
	Knn	85.33%	82.86%	87.50%	85.29%	85.37%	84.00%	82.86%	85.00%	82.86%	85.00%
TEMPO	Linear	**93.33%**	85.71%	100%	100%	88.89%	**90.67%**	94.29%	87.50%	86.84%	94.59%
	Mahalanobis	62.67%	100%	30%	55.56%	100%	89.33%	91.43%	87.50%	86.49%	92.11%
	SVM linear	89.33%	94.29%	85.00%	84.62%	94.44%	89.33%	91.43%	87.50%	86.49%	92.11%
	SVM non-linear	92.00%	91.43%	92.50%	91.43%	92.50%	90.67%	94.29%	87.50%	86.84%	94.59%
	Knn	89.33%	85.71%	100%	100%	83.33%	88.00%	88.57%	87.50%	86.11%	89.74%

∗ Considering a correctly identified AD case as a true positive; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value; GM=Gray Matter; WM=White Matter; VENT=Ventricle; HIPPO=Hippocampus; TEMPO=Temporal.

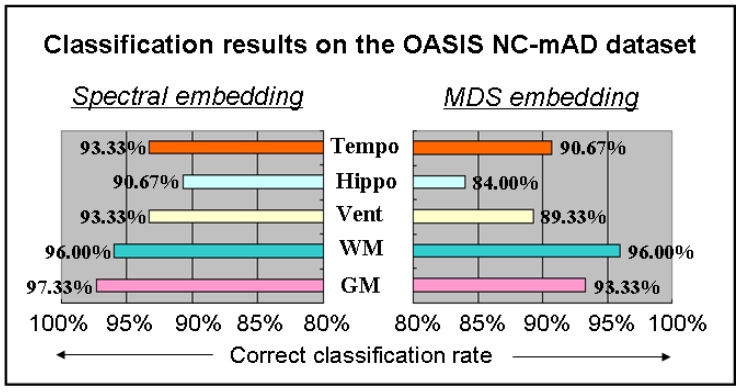


Figure 2.8: Classification results on the OASIS NC *vs* mAD dataset. Vent=Ventricle; Hippo=Hippocampus; Tempo=Temporal lobe.

The best classification rates for each regional distance and each embedding method were compared in Fig. 2.8. The spectral embedding method again performed better than the

MDS embedding on all regional distances. On this dataset the **GM** distance showed the greatest power for discrimination, with an accuracy of 97.33%. The **WM** distance gave a correct rate of 96% second only to the GM distance. The correct classified rates of **ventricle** distance, **hippocampal** distance and **temporal-lobe** distance were all higher than 90% in the spectral embedding.

Experiment 4: Classification on the OASIS (NC *vs* pMCI) dataset

The classification results, with different combinations of embedding methods and learning algorithms, are shown in Table 2.6.

Table 2.6: Classification results on the **OASIS NC *vs* pMCI** dataset

COMBINATIONS		SPECTRAL EMBEDDING					MDS EMBEDDING				
		Accuracy	SEN*	SPE	PPV	NPV	Accuracy	SEN*	SPE	PPV	NPV
GM	Linear	**94.12%**	100%	90.00%	87.50%	100%	70.59%	57.14%	80.00%	66.67%	72.73%
	Mahalanobis	64.71%	42.86%	80.00%	60.00%	66.67%	70.59%	57.14%	80.00%	66.67%	72.73%
	SVM linear	94.12%	100%	90.00%	87.50%	100%	**82.35%**	57.14%	100%	100%	76.92%
	SVM non-linear	76.47%	42.86%	100%	100%	71.43%	82.35%	57.14%	100%	100%	76.92%
	Knn	94.12%	100%	90.00%	87.50%	100%	82.35%	71.43%	90.00%	83.33%	81.82%
WM	Linear	**76.47%**	100%	60.00%	63.64%	100%	76.47%	71.43%	80.00%	71.43%	80.00%
	Mahalanobis	58.82%	28.57%	80.00%	50.00%	61.54%	70.59%	28.57%	100%	100%	66.67%
	SVM linear	76.47%	100%	60.00%	63.64%	100%	64.71%	28.57%	90.00%	66.67%	64.29%
	SVM non-linear	64.71%	57.14%	70.00%	57.14%	70.00%	**82.35%**	85.71%	80.00%	75.00%	88.89%
	Knn	76.47%	100%	60.00%	63.64%	100%	76.47%	85.71%	70.00%	66.67%	87.50%
VENT	Linear	**94.12%**	85.71%	100%	100%	90.91%	**82.35%**	71.43%	90.00%	83.33%	81.82%
	Mahalanobis	58.82%	0%	100%	0%	58.82%	70.59%	42.86%	90.00%	75.00%	69.23%
	SVM linear	76.47%	42.86%	100%	100%	71.43%	70.59%	42.86%	90.00%	75.00%	69.23%
	SVM non-linear	82.35%	57.14%	100%	100%	76.92%%	70.59%	42.86%	90.00%	75.00%	69.23%
	Knn	70.59%	85.71%	60.00%	60.00%	85.71%	70.59%	57.14%	80.00%	66.67%	72.73%
HIPPO	Linear	**76.47%**	57.14%	90.00%	80.00%	75.00%	**82.35%**	57.14%	100%	100%	76.92%
	Mahalanobis	58.82%	14.29%	90.00%	50.00%	60.00%	82.35%	57.14%	100%	100%	76.92%
	SVM linear	76.47%	57.14%	90.00%	80.00%	75.00%	76.47%	42.86%	100%	100%	71.43%
	SVM non-linear	76.47%	57.14%	90.00%	80.00%	75.00%	70.59%	28.57%	100%	100%	66.67%
	Knn	52.94%	71.43%	40.00%	45.45%	66.67%	41.18%	14.29%	60.00%	20.00%	50.00%
TEMPO	Linear	**70.59%**	57.14%	80.00%	66.67%	72.73%	**76.47%**	71.43%	80.00%	71.43%	80.00%
	Mahalanobis	58.82%	42.86%	70.00%	50.00%	63.64%	70.59%	42.86%	90.00%	75.00%	69.23%
	SVM linear	58.82%	28.57%	80.00%	50.00%	61.54%	64.71%	28.57%	90.00%	66.67%	64.29%
	SVM non-linear	64.71%	57.14%	70.00%	57.14%	70.00%	70.59%	85.71%	60.00%	60.00%	85.71%
	Knn	58.82%	42.86%	70.00%	50.00%	63.64%	76.47%	71.43%	80.00%	71.43%	80.00%

∗ Considering a correctly identified potential MCI case as a true positive; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value; GM=Gray Matter; WM=White Matter; VENT=Ventricle; HIPPO=Hippocampus; TEMPO=Temporal.

The best classification rates for each regional distance and each embedding method were compared in Fig. 2.9. Compared with Experiment 1 to 3, the classification accuracy was generally lower in this experiment. In the MDS embedding, the classification results using different regional distances showed the same accuracy of 82.35% except for the temporal-lobe distance with a lower rate of 76.47%. However, in the spectral embedding the **GM** distance and the **ventricle** distance dominated the others, with correct classified rates of 94.12% respectively; other correct rates were all lower than 77%. The results indicate that whereas no large deformation occurs or is detected within the hippocampus region or the

entire temporal lobe before healthy subjects progress to MCI, large differences are identified on the whole gray matter region and ventricle region.

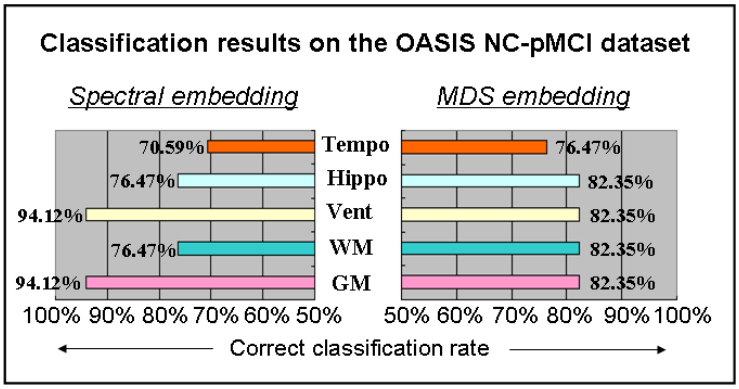


Figure 2.9: Classification results on the OASIS NC *vs* pMCI dataset. Vent=Ventricle; Hippo=Hippocampus; Tempo=Temporal lobe.

Experiment 5: Classification on the ADNI (sMCI *vs* cMCI) dataset

The embedding methods and learning algorithms were combined again to classify stable MCI and converting MCI subjects in this experiment. The results are shown in Table 2.7.

Table 2.7: Classification results on the **ADNI sMCI *vs* cMCI** dataset

COMBINATIONS		SPECTRAL EMBEDDING					MDS EMBEDDING				
		Accuracy	SEN*	SPE	PPV	NPV	Accuracy	SEN*	SPE	PPV	NPV
GM	Linear	**76.32%**	68.42%	84.21%	81.25%	72.73%	**73.68%**	68.42%	78.95%	76.47%	71.43%
	Mahalanobis	57.89%	73.68%	42.11%	56.00%	61.54%	60.53%	68.42%	52.63%	59.09%	62.50%
	SVM linear	68.42%	68.42%	68.42%	68.42%	68.42%	71.05%	68.42%	73.68%	72.22%	70.00%
	SVM non-linear	42.11%	84.21%	0%	45.71%	0%	71.05%	42.11%	100%	100%	63.33%
	Knn	68.42%	63.16%	73.68%	70.59%	66.67%	71.05%	57.89%	84.21%	78.57%	66.67%
WM	Linear	**92.11%**	89.47%	94.74%	94.44%	90.00%	**78.95%**	89.47%	68.42%	73.91%	86.67%
	Mahalanobis	73.68%	94.74%	52.63%	66.67%	90.91%	78.95%	89.47%	68.42%	73.91%	84.62%
	SVM linear	84.21%	94.74%	73.68%	78.26%	93.33%	78.95%	89.47%	68.42%	73.91%	86.67%
	SVM non-linear	89.47%	89.47%	89.47%	89.47%	89.47%	68.42%	68.42%	68.42%	68.42%	68.42%
	Knn	89.47%	89.47%	89.47%	89.47%	89.47%	71.05%	78.95%	63.16%	68.18%	75.00%
VENT	Linear	89.47%	94.74%	84.21%	85.71%	94.12%	81.58%	89.47%	73.68%	77.27%	87.50%
	Mahalanobis	86.84%	78.95%	94.74%	93.75%	81.82%	60.53%	31.58%	89.47%	75.00%	56.67%
	SVM linear	**92.11%**	89.47%	94.74%	94.44%	90.00%	78.95%	84.21%	73.68%	76.19%	82.35%
	SVM non-linear	89.47%	94.74%	84.21%	85.71%	94.12%	78.95%	84.21%	73.68%	76.19%	82.35%
	Knn	86.84%	94.74%	78.95%	81.82%	93.75%	**86.84%**	94.74%	78.95%	81.82%	93.75%
HIPPO	Linear	**68.42%**	68.42%	68.42%	68.42%	68.42%	**76.32%**	89.47%	63.16%	70.83%	85.71%
	Mahalanobis	68.42%	63.16%	73.68%	70.59%	66.67%	71.05%	78.95%	63.16%	68.18%	75.00%
	SVM linear	68.42%	68.42%	68.42%	68.42%	68.42%	76.32%	89.47%	63.16%	70.83%	85.71%
	SVM non-linear	60.53%	68.42%	52.63%	59.09%	62.50%	68.42%	63.16%	73.68%	70.59%	66.67%
	Knn	57.89%	78.95%	36.84%	55.56%	63.64%	73.68%	68.42%	78.95%	76.47%	71.43%
TEMPO	Linear	76.32%	68.42%	84.21%	81.25%	72.73%	**71.05%**	68.42%	73.68%	72.22%	70.00%
	Mahalanobis	71.05%	63.16%	78.95%	75.00%	68.18%	63.16%	73.68%	52.63%	64.00%	66.67%
	SVM linear	71.05%	57.89%	84.21%	78.57%	66.67%	68.42%	78.95%	57.89%	65.22%	73.33%
	SVM non-linear	**78.95%**	73.68%	84.21%	82.35%	76.19%	68.42%	78.95%	57.89%	65.22%	73.33%
	Knn	68.42%	63.16%	73.68%	70.59%	66.67%	71.05%	73.68%	68.42%	70.00%	72.22%

∗ Considering a correctly identified converting MCI case as a true positive; SEN=sensitivity; SPE=specificity; PPV=positive predictive value; NPV=negative predictive value; GM=Gray Matter; WM=White Matter; VENT=Ventricle; HIPPO=Hippocampus; TEMPO=Temporal.

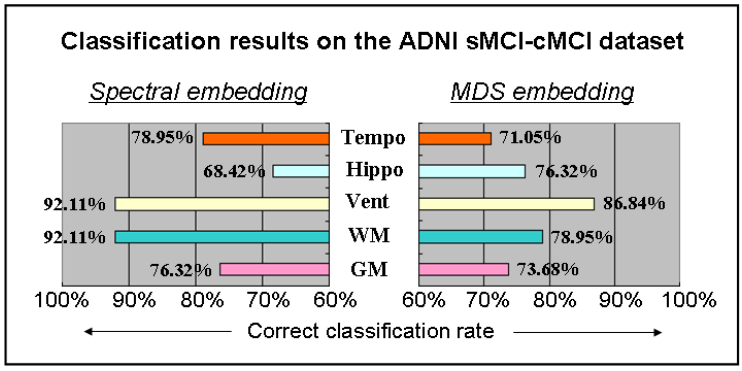


Figure 2.10: Classification results on the ADNI sMCI *vs* cMCI dataset. Vent=Ventricle; Hippo=Hippocampus; Tempo=Temporal lobe.

The best classification rates for each regional distance and each embedding method were again compared (as shown in Fig. 2.10). In this experiment, the classification accuracy was also generally lower than those between normal subjects and AD patients. The accuracy using **ventricle** distance with both spectral embedding and MDS embedding were the highest among the five, with rates of 92.11% and 86.84% respectively. The **WM** distance also yielded relatively good accuracy of 92.11% in the spectral embedding. The **GM** and **temporal-lobe** distance produced accuracy at around 75%. The **hippocampal** distance was again the worst among the selected distance metrics in both spectral embedding and MDS embedding, indicating less deformation detected in the corresponding areas.

In the above five experiments when comparing the classification accuracy using the same distance metric in each dataset, we found that most of the distances work better with spectral embedding. We speculate that the spectral embedding method strengthens the similarity within the underlying sub-group and the dissimilarity between sub-groups, and thus helps classification.

In addition, we ranked the regional distances according to the classification performance based on them (Table 2.8). Among the five distance metrics, the **WM** distance and **GM** distance were shown to have the most stable performance for discrimination, no matter which database is used (ADNI or OASIS) and no matter which stage the diseased subjects are in (sever AD, mild AD, MCI or even potential MCI). The **ventricle** distance was ranked slightly lower than the WM and GM distance. It is shown in Table 2.8 that the ventricle distance worked much better in distinguishing pre-AD cases, such as in experiments on the ADNI NC *vs* cMCI dataset, OASIS NC *vs* pMCI dataset and ADNI sMCI *vs* cMCI dataset.

Table 2.8: Ranking of different regional distances

COMBINATIONS		ADNI NC-AD	ADNI NC-cMCI	OASIS NC-mAD	OASIS NC-pMCI	ADNI sMCI-cMCI	mean
Spectral embedding	GM	2	2	1	1	4	2.0
	WM	2	2	2	3	1	2.0
	Vent	5	1	3	1	1	2.2
	Hippo	2	5	5	3	5	4.0
	Temp	1	2	3	5	3	2.8
MDS embedding	GM	2	3	2	1	4	2.4
	WM	2	2	1	1	2	1.6
	Vent	5	1	4	1	1	2.4
	Hippo	4	5	5	1	3	3.6
	Temp	1	4	3	5	5	3.6

GM=Gray Matter; WM=White Matter; Vent=Ventricle; Hippo=Hippocampus; Tempo=Temporal.

Since white matter is adjacent to both gray matter and ventricles, any changes that occur in gray matter and ventricles result in changes in white matter. Therefore, WM distance performed the best among the five listed distance metrics. The **temporal-lobe** distance performed considerably well on the ADNI, especially on the NC *vs* AD dataset, but not as good on the OASIS datasets. The **hippocampal** distance was the worst among the five. The reason might be that information obtained from a smaller region could be easier affected by noise, registration error and inter-individual variability.

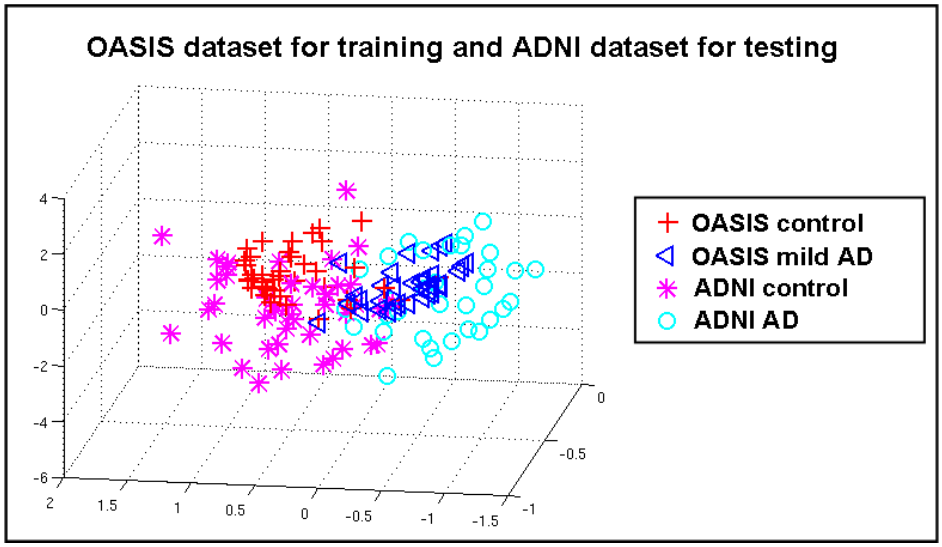
Experiment 6: Generalization test

From experiment 1 to 5, classification was performed within each dataset. However, we are also interested in how the classification works on discriminating brains scanned from different communities using different scanners with different imaging protocols. In addition, a disadvantage of the leave-one-out cross validation is that all the data has been used for training, so this does not give a true independent assessment of the quality of the model. Thus, we are interested in how the proposed algorithm works on classifying subjects never used for training.

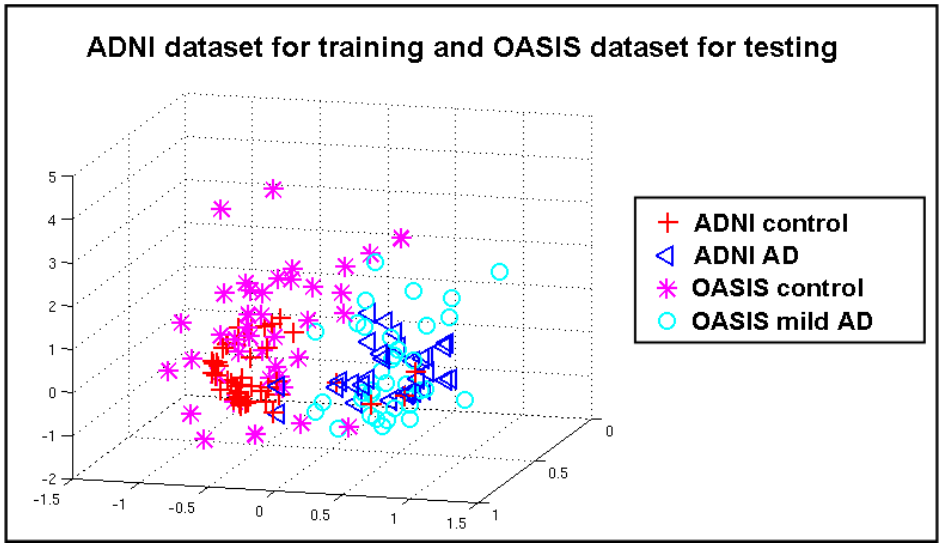
Two datasets were selected in this experiment: the OASIS (NC *vs* mAD) and the ADNI (NC *vs* AD) dataset. The GM distance was calculated and the spectral embedding along with the linear discriminant was used for classification.

We first treated the subjects in the OASIS (NC *vs* mAD) dataset as training samples and the subjects in the ADNI (NC *vs* AD) dataset as testing samples. The embedded images

in the $\mathbb{R}^3$ space are shown in Fig. 2.11-(a). Subjects from the ADNI (NC *vs* AD) dataset were correctly assigned to the appropriate diagnostic group in 92.21% of trials (sensitivity 90.91%, specificity 93.18%).



(a)



(b)

Figure 2.11: Train on the OASIS/ADNI dataset and classification on the ADNI/OASIS dataset. (a) OASIS (NC *vs* mAD) for training and ADNI (NC *vs* AD) for testing; (b) ADNI (NC *vs* AD) for training and OASIS (NC *vs* mAD) for testing.

When the training dataset and the testing dataset were exchanged, that is, training on the ADNI (NC *vs* AD) dataset and classifying the OASIS (NC *vs* mAD) dataset, 84% of subjects were assigned to the appropriate group (sensitivity 77.14%, specificity 90%). The embedded images are shown in Fig. 2.11-(b).

2.4 Discussion

2.4.1 Comparison with recently published methods

We compared the classification accuracy of our method with those reported in a recently published work by Cuignet et al. [77]. In [77], ten methods were evaluated on the same subsets of the ADNI database. These ten methods were grouped into three categories: voxel-based volumetric approaches, vertex-level surface-based approaches and hippocampus-based approaches. Cuignet et al. also performed experiments to classify three different datasets: control *vs* AD patient, control *vs* converting MCI, stable MCI *vs* converting MCI. We compared our results with the best classification rates from each category of the methods assessed by Cuignet et al. (as shown in Table 2.9).

In addition, the whole point of a diagnostic test is to use it to make a diagnosis, so we need to know the probability that the test will give the correct diagnosis. The sensitivity and specificity do not give us this information. Instead we must approach the data from the direction of the test results, using predictive values. Thus the positive predictive value (PPV) and negative predictive value (NPV) are also compared in this section.

Comparing the sensitivity and specificity in Table 2.9 on the ADNI datasets, our method showed better performance than those approaches evaluated in the paper of Cuignet et al. in discriminating both AD and MCI subjects. In separating stable MCI and converting MCI subjects, our method yielded comparable sensitivity and specificity rates as the other approaches did using GM. However, if we go back to Fig. 2.10, we find that GM did not work so well in distinguishing stable MCI and converting MCI compared with WM and ventricle. Using the WM and ventricle distance, the accuracy of grouping is obtained as high as 92.11%, which is much better than the best classification rate ever reported in the same case [108]. Moreover, there is no way to complete this challenging task clinically since no criteria can be made to separate stable MCI and converting MCI. Thus the accuracy of 92.11% is considerably promising. Comparing the positive and negative predictive values, we

Table 2.9: Comparison between our method and other published methods

	Our method				Voxel-based method		Vertex-based method		Hippo-based method	
	ADNI (using GM)		ADNI (using Hippo)		ADNI (using GM)		ADNI (using GM)		ADNI (using Hippo)	
	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE	SEN	SPE
Normal-AD	97%	95%	95%	97%	81%	95%	79%	90%	69%	84%
Normal-cMCI	95%	91%	85%	89%	68%	95%	65%	94%	73%	74%
sMCI-cMCI	68%	84%	89%	63%	62%	67%	32%	91%	62%	69%

	Our method				Voxel-based method		Vertex-based method		Hippo-based method	
	ADNI (using GM)		ADNI (using Hippo)		ADNI (using GM)		ADNI (using GM)		ADNI (using Hippo)	
	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV	PPV	NPV
Normal-AD	94%	98%	94%	98%	93%	86%	87%	84%	78%	76%
Normal-cMCI	90%	95%	87%	86%	86%	87%	83%	85%	56%	86%
sMCI-cMCI	81%	73%	71%	86%	51%	76%	67%	71%	52%	77%

GM=gray matter; cMCI=converting Mild Cognitive Impairment; sMCI=stable Mild Cognitive Impairment; Hippo=Hippocampus; SEN=Sensitivity; SPE=Specificity; PPV=Positive Predictive Value; NPV=Negative Predictive Value. The evaluation method used in the paper of Cuignet et al. is leave-one-out cross validation. The datasets used in that paper are also selected from ADNI database but with a larger size.

observe that our results also outperform the others, indicating that the reliability of positive and negative test of our method is better than in those methods used for comparison.

The advantage of our method may lie in the fact that we calculated more precise distance between images and used as much information as possible in dimensional reduction. We compute the pair-wise registration instead of aligning subjects to an atlas or a constructed template, resulting in more informative embedding and consequently resulting in better classification. This will be further discussed in the following section.

2.4.2 Why our algorithm improves classification results

We have mentioned that many algorithms about classifying AD or MCI brains have been proposed. Among all the existing approaches, the algorithm developed by Kloppel [13] is the most popular one used for comparison. In Kloppel's method, features are defined at the level of MRI voxel and they are directly classified with a support vector machine (SVM). Given a set of MR images $\mathcal{I} = \{I_1, I_2, \ldots, I_N\}$, they first segmented images into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Then, GM segments were further normalized to the population template generated from all the images in the dataset $\mathcal{I}$. A separate 'modulation' step was used to ensure that the overall amount of each tissue class remained constant after normalization [22]. The voxels on the GM segment were treated as coordinates of a high dimensional space and their location was determined by the intensity value at each

voxel.

We implemented Kloppel's algorithm and applied it to our ADNI NC *vs* AD dataset, which has already been described and used in Section 2.3 Experiment 1. Using GM, the classification accuracy of Kloppel's algorithm is 75.32% (with sensitivity[9]of 69.70% and specificity of 79.55%). This result is much worse than ours (with sensitivity of 100%, specificity of 95.45%, and accuracy of 97.40%).

The reasons may be that (a) using features generated from deformations is better than using registration residuals, (b) using pair-wise registration and dimension reduction is better than using template-based registration and performing classification in a high dimensional space.

(a). Different feature generation scheme

The ways of feature generation are different in Kloppel's method and our method (as shown in Fig.2.12). Kloppel et al. utilized the residual anatomic differences, whereas we analyzed the deformation fields or the spatial derivatives of deformation fields.

In Fig. 2.12 although the warped images ((I-i) and (II-i)) are similar, that is, the variations between the registration residuals are small, the deformation fields of the two source images differ much, especially around the hole area. Thus we speculated that the difference between deformation fields could better represent the difference between the corresponding source images.

Then, still using the ADNI NC *vs* AD dataset, we generated features from the Jacobian determinant of deformation fields and classified the features in the high dimensional space. Accuracy was now obtained with the rate of 90.91% (sensitivity 90.91%, specificity 90.91%).

(b). Different classification space and different registration strategy

In Kloppel's method, features were classified in the high dimensional space. The number of voxels in GM segments was the dimension of the classification space. In our method, we performed a dimension reduction procedure, which improves the efficiency and accuracy of image grouping.

The reasons for doing dimension reduction can be explained in two respects: statistical and machine learning based [78]. From a statistical point of view it is desirable that the number of subjects in the training set should significantly exceed the number of features used to describe those subjects. In practice this is not the case as real high-dimension data will on-

[9]Considering a correctly identified AD case as a true positive.

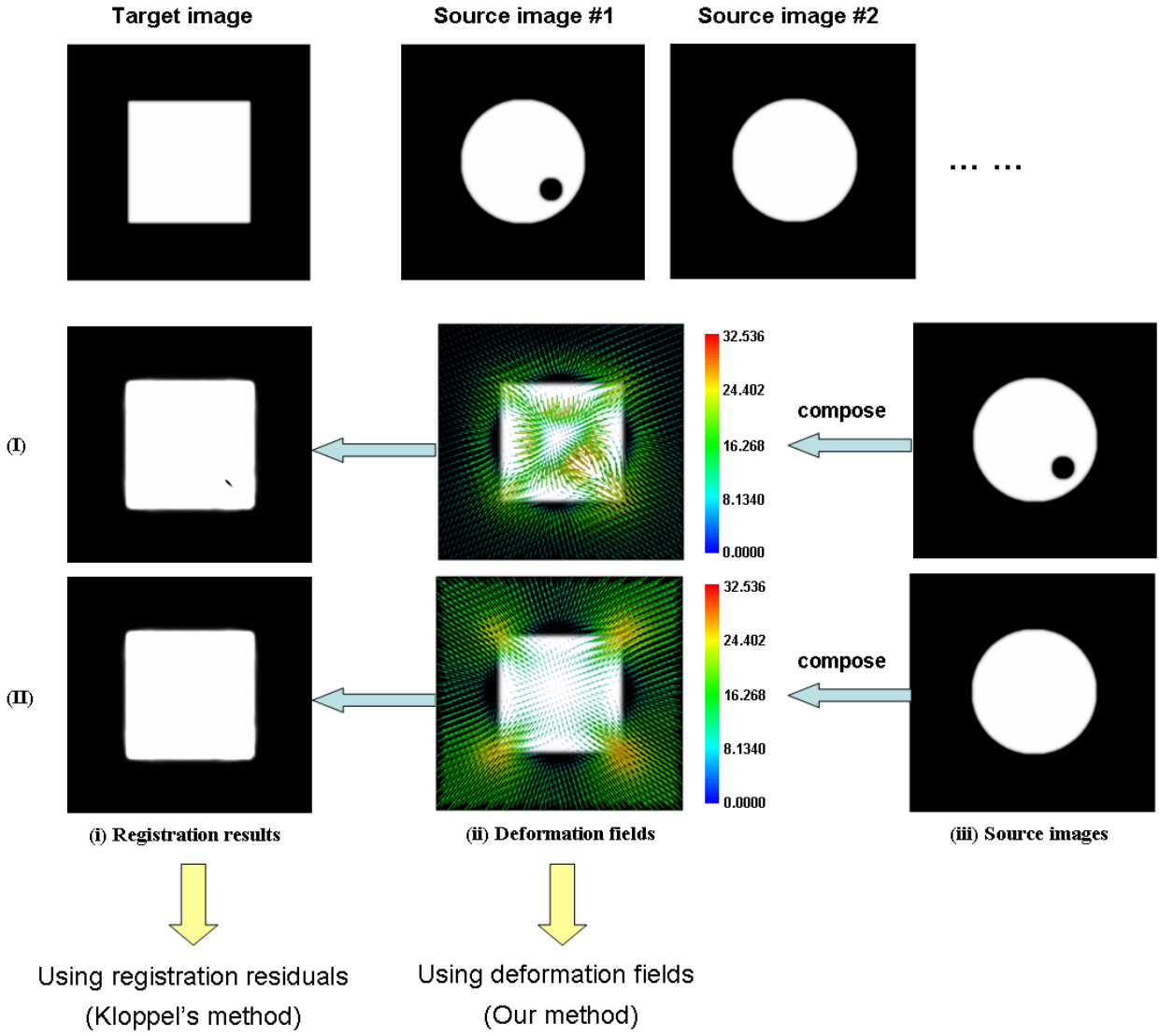
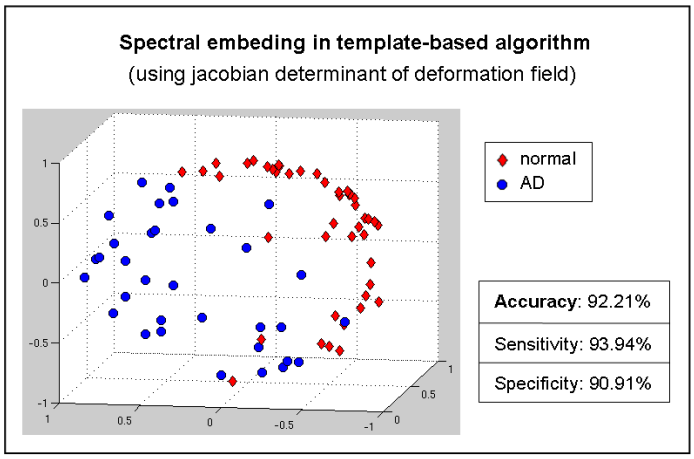


Figure 2.12: Feature generation. Features are generated from registration residuals in Kloppel's method, while they are generated from deformation fields in our method.

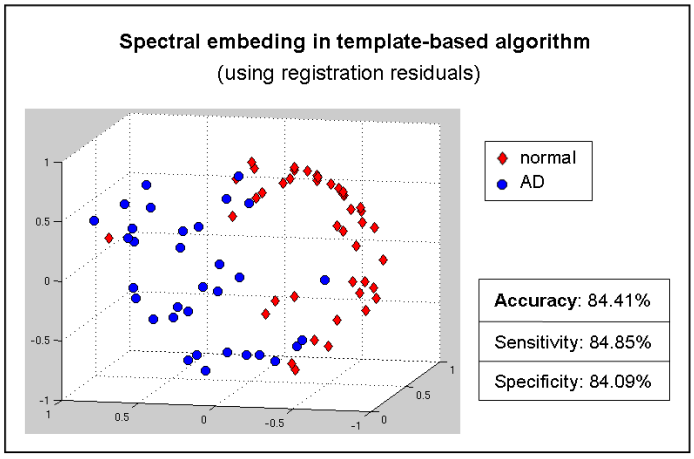
ly occupy a manifold in the input space so the implicit dimension of the data will be lower than the number of features. From the viewpoint of machine learning and pattern recognition, traditional algorithms are often susceptible to the well-known problem of the *curse of dimensionality*, which refers to the degradation in the performance of a given learning algorithm as the number of features increases. To deal with this issue, dimension reduction techniques are often applied as a data pre-processing step or as part of the data analysis to

simplify the data model. By working with this reduced representation, tasks such as classification or clustering can often yield more accurate and readily interpretable results, while computational costs may also be significantly reduced.

Therefore, we reduced the dimension of the features used in grouping the ADNI NC *vs* AD dataset and performed classification for comparison. The spectral embedding algorithm was employed here, and the dimension of the embedded space was chosen 3. Using registration residuals, the accuracy for correctly distinguishing AD subjects increased to 84.41% (sensitivity 84.85%, specificity 84.09%). Using deformation fields, the accuracy increased up to 92.21% (sensitivity 93.94%, specificity 90.91%). The results are shown in Fig. 2.13.



(a)



(b)

Figure 2.13: Spectral embedding based on template-based registration. (a) Using registration residuals, (b) Using deformation fields.

Although the classification accuracy have been improved after using the spectral embedding, it is still not as good compared with our experimental results shown in Section 2.3 Experiment 1. This may be due to the difference in calculating distances. The distance used in existing methods and our method are both correlated to image registration, however, other algorithms chose template-based registration while we chose pair-wise registration (as shown in Fig. 2.14).

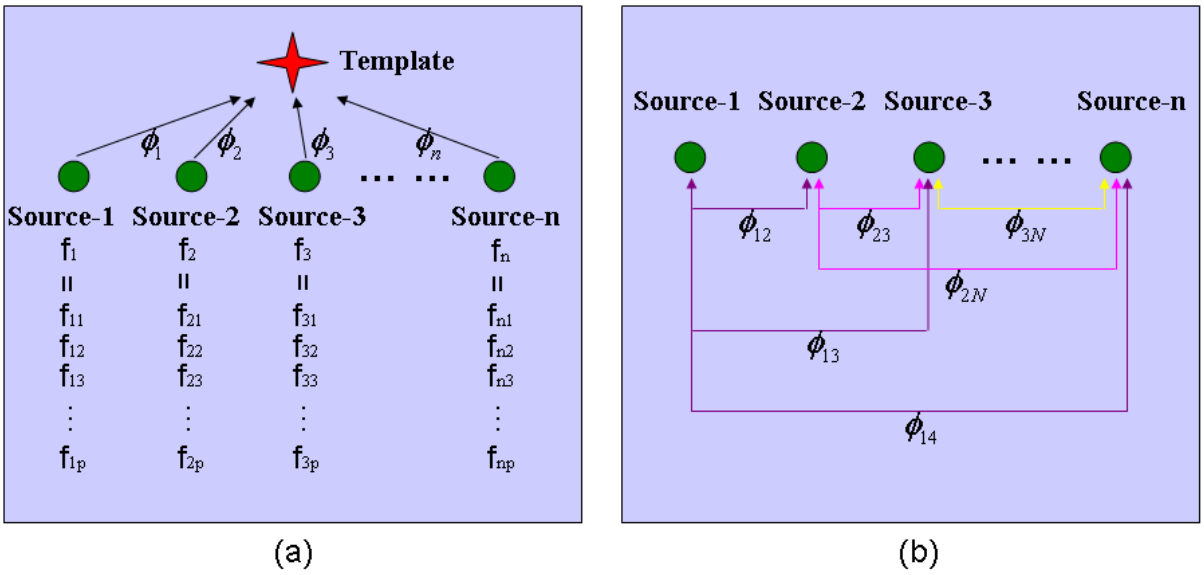


Figure 2.14: Distance calculation on template-based registration and pair-wise registration. (a) Template-based registration and features, (b) Pair-wise registration.

There are two reasons why we used pair-wise registration instead of template-based registration. The first involves the template selection problem as a template would affect the registration and consequently affect the distance calculation a lot. The existing template-based methods either used a common atlas or constructed a template based on their own data. Some issues may be introduced into this procedure: as an atlas or a template is a mean image of a population, it might be blurred or biased to some specific subject. The second reason is because of the non-ideal registration problem. In template-based method, each distance is calculated from two deformation fields or two registration residuals, which might suffer larger registration errors than using pair-wise alignment. For example, calculating the distance between source 1 and source 2 in the template-based scheme involves calculating a mapping ϕ_1 from source 1 to a template, a mapping ϕ_2 from source 2 to a template and a distance between ϕ_1 and ϕ_2, whereas calculating the same distance in the pair-wise scheme

is only concerned with a mapping ϕ_{12} between source 1 and source 2.

The improving procedure is also demonstrated in Fig. 2.15.

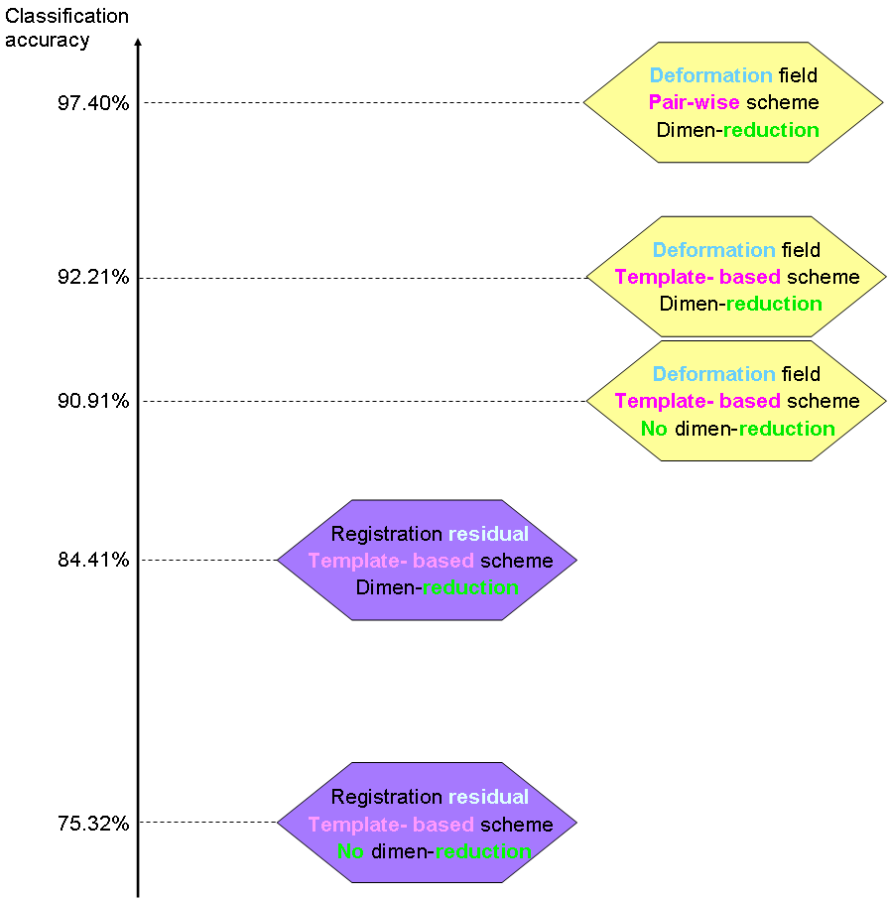


Figure 2.15: The step-by-step improvement of classification accuracy. Our proposed method is on the top, with deformation field, pair-wise scheme and dimensional reduction.

2.4.3 Disease progression

In this dissertation, experiments are designed to distinguish between normal subjects and different level of AD patients, including potential MCI, converting MCI, mild AD and severe AD. Another experiment is designed to discriminate stable MCI and converting MCI subjects. From the classification results and the ranking shown in Table 2.8, we can briefly summarize the progression of Alzheimer's disease.

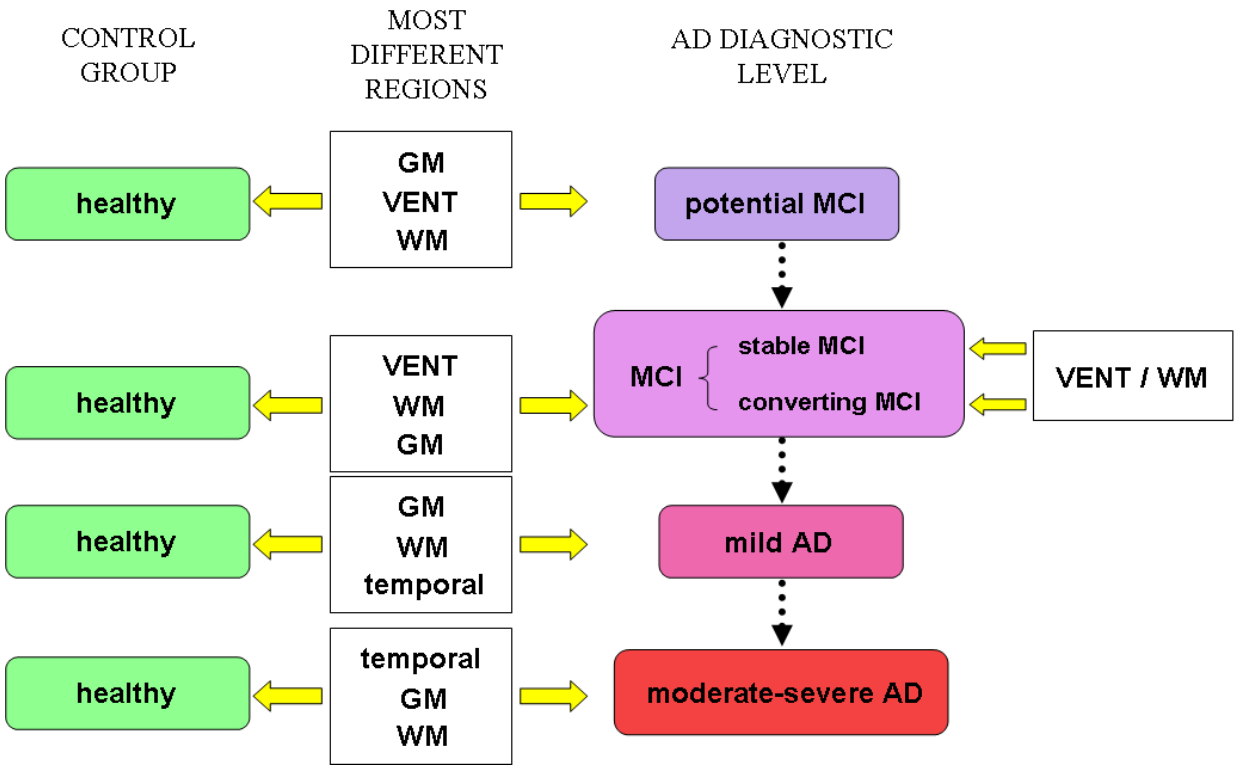


Figure 2.16: The most different regions found between groups.

In pre-AD (potential MCI and MCI) levels, GM, WM and ventricle present the largest differences between healthy subjects and pre-AD patients. Deformation occurs in almost all the brain regions instead of only in specific structures such as temporal lobe and hippocampus. In post-AD (mild AD and moderate-severe AD) levels, GM and WM still show significant deviations between healthy subjects and AD patients. However, temporal lobe becomes a stronger biomarker to discriminate AD, especially to discriminate severe AD, indicating that although shrinkage occurs on the whole gray matter region, the most significant changes appear on the temporal lobe as the disease deteriorates.

For the classification between stable MCI and converting MCI, the performance is low using GM, hippocampus and temporal lobe. However, WM and ventricles still present much detectable differences, with high classification accuracy rates around 92%. This observation is consistent with our previous deduction that measuring the WM or ventricles is advantageous in detecting pre-AD cases.

Chapter 3

Pathology localization and comparison with existing methods

3.1 Background

Creating maps of pathology distribution in AD is important for establishing structure-deficit correlations. These maps may reveal AD-related regional atrophy at an early time point, or at a specific stage of the disease. Moreover, they provide insight into longitudinal disease progression, and may reveal subtle brain changes associated with prognosis or treatment, i.e. an AD brain phenotype.

Some of the widely used approaches to detect alterations in brain structures between groups include cortical surface-based comparisons of gray matter thickness, whole-brain voxel-wise morphometric analysis and region-of-interest (ROI) based techniques. Cortical thickness [15, 16, 18, 17] offers a direct quantitative index of cortical atrophy that can be applied to single subjects and to group analysis. Each subject image is first pre-processed with intensity normalization, affinely aligned to a standard space, skull stripped and the tissue types segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Then a 3D map is computed to quantify the distance from the cortical gray matter voxels or vertices to the gray/white matter interface. Thickness values are mapped onto each point of a cortical surface model extracted from the segmentation. The surface models are then registered to an atlas space via non-linear surface-based matching. A statistical model can be applied to the thickness measures from all subjects at each surface point to establish

differences. However, one disadvantage of cortical thickness measurement is that it doesn't take into consideration subcortical structures, which are also observed to be highly associated with psychiatric and neurodegenerative diseases [52, 53, 54].

Statistical models are also used in whole-brain voxel-based morphometric analyses [19, 20, 21, 22, 23]. This class of technique can be divided into two sub-categories [22]: one deals with differences in brain shape (such as Deformation-Based Morphometry, DBM, and Tensor-Based Morphometry, TBM) and another deals with differences in the local composition of brain tissue after macroscopic differences in shape have been discounted (such as Voxel-Based Morphometry, VBM). In voxel-wise morphometry, subjects are all linearly and non-linearly registered to a standard space. While DBM and TBM use the deformation fields or the spatial derivatives of deformation fields, VBM compares residual anatomic differences after registration. Compared with cortical thickness measurement, DBM, TBM and VBM assess the entire brain. However, the use of voxel-wise hypothesis tests introduces the multiple comparison problem.

The last class of approaches are ROI-based techniques. Although ROI-based methods have gained general acceptance, it is limited by the fact that the selection of regions depends on *a-priori* knowledge or a hypothesis and leaves large areas of the brain unexplored. However, such studies generally have better statistical power than voxel-wise methods.

In this chapter, we examined a classification-based algorithm to detect regional deviations in the brain. Although it is, at its core, an ROI-based method, it improves upon the classic approach by allowing hierarchical searching among the regions of interest, with no user bias or *a-priori* region selection (as shown in Fig. 3.1).

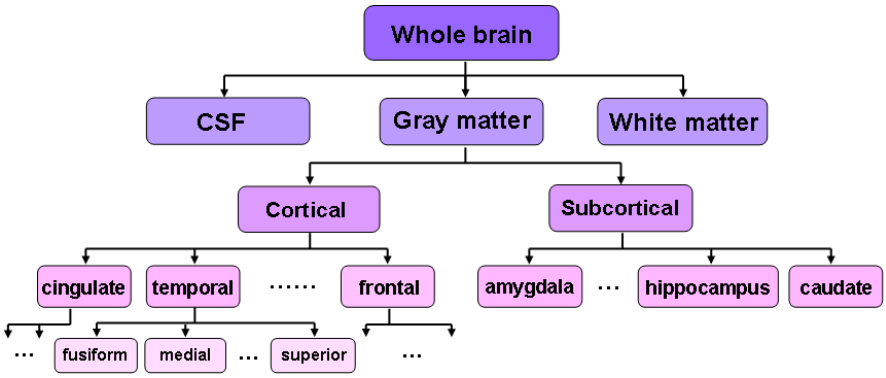


Figure 3.1: Hierarchical segmentation of brain.

In our method, a brain image is hierarchically segmented into subregions using the *FreeSurfer* [34] software. The classification procedure based on each subregion was described in Chapter 2: pair-wise registration is employed between whole-brain images and deformation on each selected region is quantified as Riemannian distance; regional distances embedded brain images into a low-dimensional Euclidean space; labeled brains are used to train a discriminant hyperplane and test brains are classified according to the computed hyperplane. For each subregion, classification accuracy along with sensitivity and specificity are calculated. Those regions yielding higher correct classification rates are thought to be the better metrics separating the underlying groups, and consequently they are regarded as the regions with significant morphological changes in AD pathology. In comparison with voxel-wise morphometry for pathology exploration, our region-based method is more powerful since the distance for brain embedding is computed by considering the overall deformation of the target structure. Compared with cortical thickness measurement, both cortical regions and subcortical structures are considered in the hierarchical search, detecting regions associated with the disease across the whole brain.

3.2 Method

In pattern recognition, features extracted from a sample are used for classification. Different features result in different classification performance. Conversely, different classification results reflect the different power of features to group samples. If a feature gives a higher classification accuracy than another feature does, the former is regarded as a better metric for partitioning. This is called feature selection, and has been studied extensively [110, 111].

In brain morphometric analysis, we use this feature-classification relationship to solve the problem of pathology localization, that is, we treat it as a feature selection problem. Features here are generated from displacement fields within different regions of the brain, such as cortical lobes and subcortical structures. The Multi-Dimensional Scaling (MDS) algorithm is employed to reduce the feature dimension from $\mathbb{R}^n$ to $\mathbb{R}^d$, where n is the number of voxels in the selected region and d is manually chosen, $d << n$. In the embedded $\mathbb{R}^d$ space, the leave-one-out cross validation is applied to evaluate the classification performance. Once the classification accuracy with different features is obtained, the features are ranked according to their power of discrimination. Different features are generated from different brain regions, including cortical and subcortical regions. Segments were obtained with the *FreeSurfer* software.

3.2.1 FreeSurfer segmentation

FreeSurfer[1] is a free software package developed at the Martinos Center for Biomedical Imaging used for three main purposes: (1) creation of computerized models of the brain from MRI data, (2) measuring various morphometric properties of the brain including cortical thickness, curvature characteristics, and cortical and sub-cortical regional volumes, (3) inter-subject cortical spatial normalization based on aligning individual's cortical folding patterns to a population averaged cortical folding pattern to establish correspondence between homologous anatomical regions.

Starting with a high quality T1-weighted MR image volume, the first step in the *FreeSurfer* processing is to perform an affine transformation (12 degrees of freedom) to the Talairach image space, non-uniform intensity normalization for intensity inhomogeneity correction, and removal of non-brain tissues. The remaining brain image volume is intensity normalized to match the *FreeSurfer* atlas image intensity histogram [34], which is followed by a non-linear warping of the atlas brain image to subject brain image [112, 113]. The warped atlas brain image, in the subject image space, is used for atlas-based tissue segmentation, as well as labeling the subcortical structures, brain stem, cerebellum, and cerebral cortex [114]. The next step in *FreeSurfer* is to generate a topologically correct cortical surface representation per hemisphere [112, 113, 115]. The cortical surface lies either at the WM/GM tissue interface or at the GM/CSF tissue interface. Each hemisphere's cortical surface representation is mapped automatically to a standard spherical coordinate system [94]. Key components of the surface mapping include surface inflation with minimal metric distortion, projection to spherical coordinates, topology correction, and surface-based warping to align anatomically homologous points. Mapping to the standard spherical coordinate system defined by *FreeSurfer* allows for automated anatomical parcellation of the cortex into gyral regions [116, 117]. Surface parcellation is then extended to the GM volume, yielding parcellation of the GM tissue sheet and regional cortical volumes.

Cortical and subcortical segmentation are shown in Fig. 3.2. Although the *FreeSurfer* processing framework allows for manual edits to improve the segmentation results, given the large number of subjects in a dataset, to minimize the processing time and to eliminate the inter-rate variability from manual edits, the fully-automated approach was used in our experiments. A general review of the *FreeSurfer* results however showed no large discrepancy.

[1] http://surfer.nmr.mgh.harvard.edu/ The software version used in our work is freesurfer-Linux-centos4_x86_64-stable-pub-v5.0.0

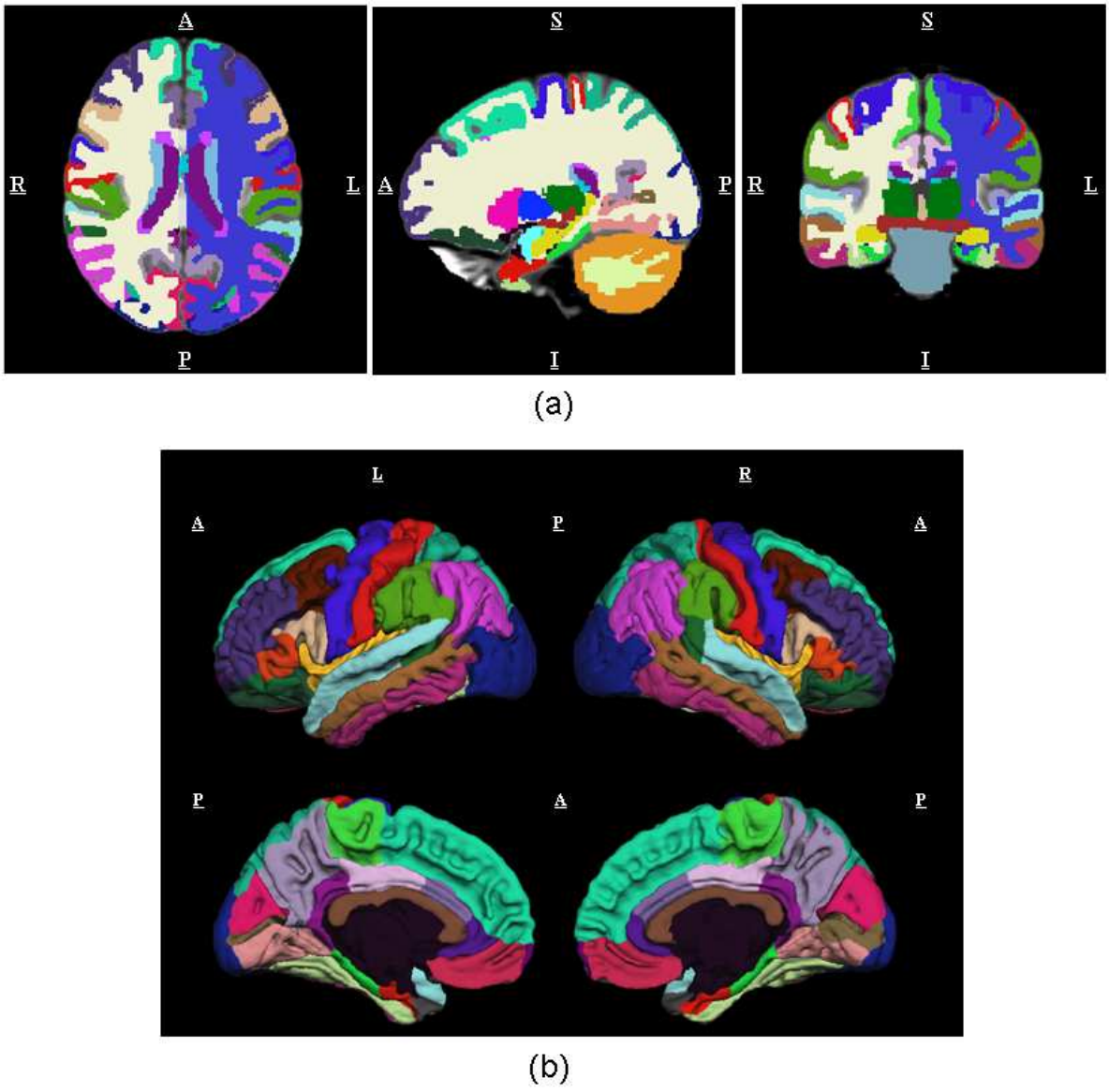


Figure 3.2: *FreeSurfer* segmentation. (a) Cortical and subcortical segmentation overlapped on MR volume, (b) Cortical segmentation overlapped on the white matter surface.

3.2.2 Algorithm summation

The classification procedure based on each region of interest was introduced in Chapter 2. In general, the classification-based pathology localization algorithm is summarized as follows.

Algorithm (Classification-based pathology localization algorithm)

Given a set of images $\mathcal{I} = \{I_1, \ldots, I_N\}$,

1. *Compute and choose segmentations* $\Omega^1, \Omega^2, \ldots$, *representing different cortical regions or subcortical structures, for each subject using the FreeSurfer;*
2. *Select a region of interest* Ω^i*;*
3. *Compute the non-linear registration for each pair of images over the dataset* $\mathcal{I}$ *using the symmetric log-domain diffeomorphic demons algorithm;*
4. *Compute the pair-wise regional distance, based on the diffeomorphism computed from registration along with the segmentation using Eq. 2.14, and construct a distance matrix;*
5. *Perform the multi-dimensional scaling to reduce the feature dimension;*
6. *Perform the leave-one-out cross validation to calculate the classification accuracy* γ^i*;*
7. *Repeat Step2∼6 until all selected regions of interest have been analyzed;*
8. *Compare and rank all the regions of interest according to the classification accuracy to localize the regions associated with the pathology;*

3.3 Other existing methods

3.3.1 Cortical thickness measurement

The assessment of cerebral cortical thickness (the thickness of the cortical grey matter ribbon) are clinically important in investigations such as longitudinal tracking on brain maturation and disease progression. Studies have suggested that various diseases, i.e. Alzheimer's, Schizophrenia and AIDS, may affect the cortical thickness [15, 109].

For comparison to classification-based localization of AD pathology, the cortical thickness measurement is performed on the same data using the *FreeSurfer* software (http://surfer.nmr.

mgh.harvard.edu/fswiki/Download/freesurfer-Linux-centos4_x86_64-stable-pub-v5.0.0). The detailed procedure for cortical thickness measurement has been described in previous publications [34, 55, 95]. Only a rough overview will be given here (Fig. 3.3).

Using the standard options of the software, we compute cortical thickness measures of each MR scan. Thickness measurements and surface registration can be obtained completely automatically in about 24 hours per subject using current readily available computer hardware, with no major errors or inconsistencies [66]. The tool is called "recon-all"; it is a batch program and runs more than 30 steps.

The first stage of the "recon-all" processing involves several intensity normalization steps, followed by skull stripping using a watershed technique [67, 68]. The watershed algorithm is a general technique for segmenting images into contiguous regions. The hemispheres are separated with a midsagittal cut, and all surface analyses are performed one hemisphere at a time.

The next stage of "recon-all" processing pipeline is the generation of the surface mesh representations using a deformable surface algorithm (active surface) [69]. The volume is first separated into gray matter and white matter, and the inner white matter mass is filled and topologically corrected for connectivity. The initial tessellation is formed on the faces of the voxels on the surface of this white matter mass, and then smoothed using a complex algorithm to form the white-gray matter surface. Then this surface is expanded to the gray matter-CSF surface using the deformable surface algorithm. In both stages of surface adaptation, checks are made to avoid self-intersection. Finally, the topology of the surface is corrected using a calculation of the Euler number of the surface.

Since the outer cortical surface is generated by expanding the inner surface, there is a one-to-one correspondence between mesh nodes. The thickness map is computed by measuring the closest distance from the cortical white-gray matter interface to the cortical surface (gray-CSF interface) using both intensity and continuity information [55]. The generated surfaces are mapped onto a spherical coordinate system and "inflated" [95]. The spherically-mapped surfaces can be registered onto other spherically-mapped surfaces, including template average surfaces, using a procedure of deforming the surface coordinates to maximize correlation between the surface convexity measures. This has the effect of attempting to match the pattern of sulci and gyri between brains. Thickness data are mapped to each vertex on the inflated surface, allowing visualization of data across the entire cortical surface. All cortical thickness maps are smoothed with a 10 mm full-width at half-maximum (FWHM) Gaussian

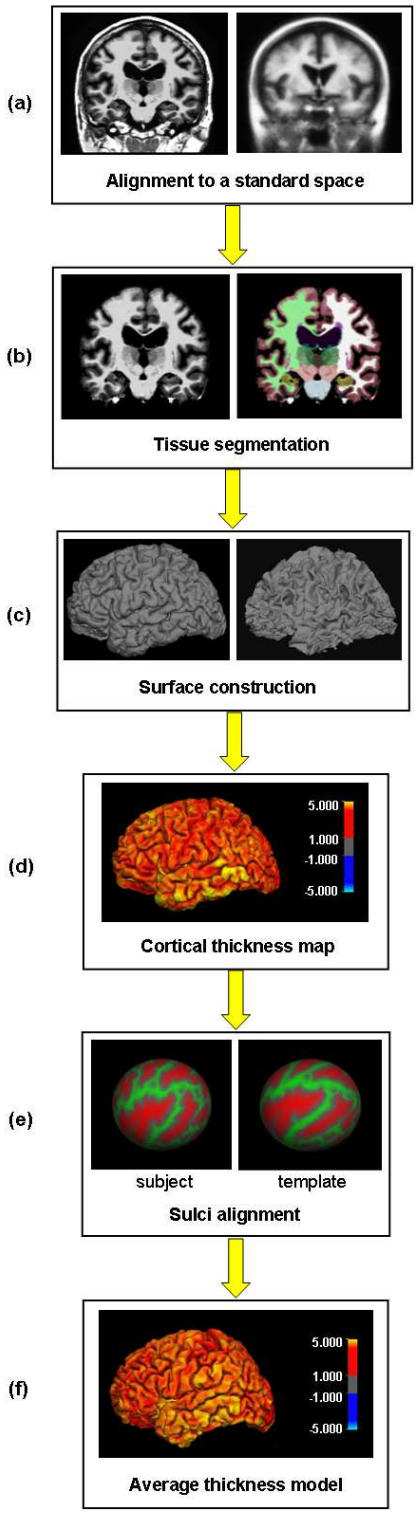


Figure 3.3: Cortical thickness measurement pipeline.

kernel to reduce local variations in the measurements [70, 71]. This improves the quality of comparison and adjust for minor registration errors. Gray matter thickness is then compared across subjects and averaged at each cortical surface location. Statistics are then defined to identify brain structural similarities/differences within/between groups. The regional cortical thickness variations between the control group and patient group were assessed using a vertex-by-vertex general linear model (GLM) with the *Qdec* software (http://surfer.nmr.mgh.harvard.edu/fswiki/Qdec). The statistical analysis used here is a surface-based voxel/vertex-wise hypothesis test. Multiple-comparison correction is performed to control false discovery rate (FDR) at a 0.05 level of significance (FDR control will be described in Chapter 3.3.2).

General linear model (GLM)

A linear model describes the observed data as a linear combination of explanatory factors plus noise, and determines how well that description explains the data being analyzed [74, 75]. The notation is defined here as:

$$\underset{(n\times d)}{\boldsymbol{Y}} = \boldsymbol{W} * \underset{(n\times p)}{\boldsymbol{X}} * \underset{(p\times d)}{\boldsymbol{B}} + \underset{(n\times d)}{\boldsymbol{U}}, \tag{3.1}$$

where $\boldsymbol{Y}$ is the matrix of n independent observations on d response variables (e.g. thicknesses for n subjects at d voxels/vertices), $\boldsymbol{X}$ contains the $n \times p$ explanatory variables (the design matrix, e.g. gender, age, diagnosis), $\boldsymbol{B}$ is the matrix of $p \times d$ unknown regression parameters and $\boldsymbol{W}$ is a diagonal weighing matrix. The errors $\boldsymbol{U}$ are assumed to be independent and identically distributed (*i.i.d.*) d-dimensional normal random variables with mean 0 and covariance Σ, i.e. $\boldsymbol{U} \overset{i.i.d.}{\sim} N_d(0, \Sigma)$.

During the estimation stage, the forward model is inverted to solve for $\boldsymbol{B}$:

$$\boldsymbol{B} = inv(\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{X})\boldsymbol{X}'\boldsymbol{W}'\boldsymbol{Y}. \tag{3.2}$$

The signal estimate is computed as

$$\hat{\boldsymbol{Y}} = \boldsymbol{B} * \boldsymbol{X}. \tag{3.3}$$

The residual error is computed as

$$e = \boldsymbol{Y} - \hat{\boldsymbol{Y}}. \tag{3.4}$$

The noise variance estimate r is computed as the sum of the squares of the residual error

divided by the degrees of freedom (DOF). The DOF equals the number of rows of $\boldsymbol{X}$ minus the number of columns. A contrast matrix $\boldsymbol{C}$ has j rows and as many columns as columns of $\boldsymbol{X}$. The contrast is then computed as:

$$\boldsymbol{G} = \boldsymbol{C} * \boldsymbol{B}. \tag{3.5}$$

The F-ratio for the contrast is then given by:

$$\boldsymbol{F} = \frac{\boldsymbol{G}' * inv\,[\boldsymbol{C} * inv(\boldsymbol{X}'\,\boldsymbol{W}'\,\boldsymbol{W}\boldsymbol{X}) * \boldsymbol{C}'] * \boldsymbol{G}}{j * r}. \tag{3.6}$$

The $\boldsymbol{F}$ is then used to compute a p-value. Note that when $j = 1$, this reduces to a two-tailed t-test.

3.3.2 Tensor-based morphometry (TBM)

While cortical thickness measurement provides maps of changes on the cortex, tensor-based morphometry (TBM) can track volumetric changes throughout the brain. For cross-sectional analysis, TBM can detect and automatically quantify the patterns of brain atrophy in diseases [56, 57]. For longitudinal studies, TBM creates maps of brain changes, visualizing the 3D profile and rates of tissue growth or atrophy [58].

a. Jacobian map creation

Tensor-based morphometry is one method of studying brain shape that is based on deformation fields obtained by non-linear registration of brain images. TBM computes the spatial derivatives of the deformation fields that match a set of brain images to a common template (Fig. 3.4) [21].

Given a set of cross-sectional MR images $\mathcal{S} = \{\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N\}$, the key point in TBM is to non-linearly deform all the images to a common space $\boldsymbol{T}$, which acts as a template, and the displacement vector $\vec{u}(\vec{r})$ is computed such that $\boldsymbol{T}(\vec{r} - \vec{u})$ corresponds with $\boldsymbol{S}(\vec{r})$, where $\vec{r}$

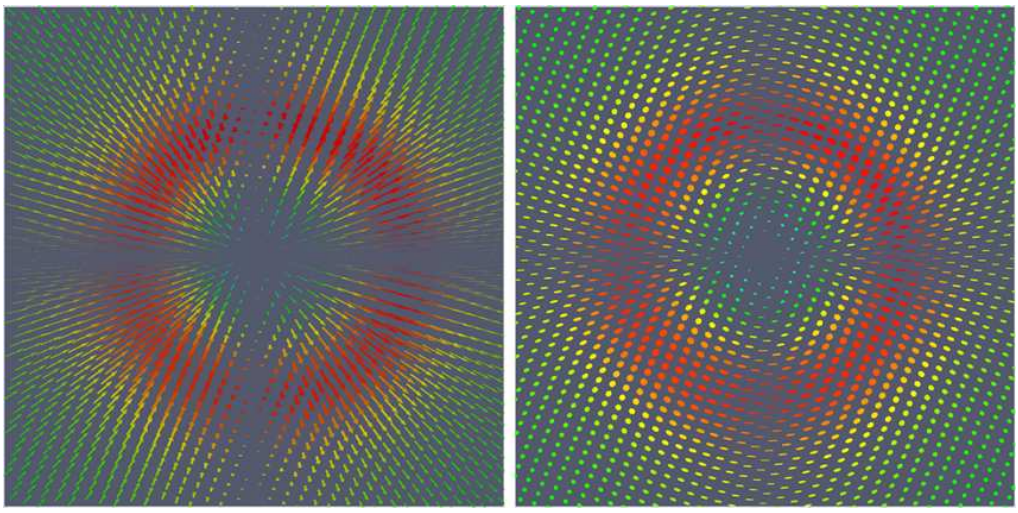


Figure 3.4: Currently the main application of TBM involves using the Jacobian determinants to examine the relative volumes of different structures. However, there are other features of the Jacobian matrices that could be used, such as those representing elongation and contraction in different directions. Left: The arrows in the image show absolute displacements after making a global correction for scalings and translations. Right: the ellipses show how the same circles would be distorted in different parts of the image.

denotes the voxel location. The Jacobian matrix of the deformation field is defined by

$$\boldsymbol{J} = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(x-u_x)/\partial y & \partial(x-u_x)/\partial z \\ \partial(y-u_y)/\partial x & \partial(y-u_y)/\partial y & \partial(y-u_y)/\partial z \\ \partial(z-u_z)/\partial x & \partial(z-u_z)/\partial y & \partial(z-u_z)/\partial z \end{pmatrix}.$$

Then the Jacobian determinant of the deformation field at each voxel is used to measure the local volume differences (i.e. expansion or contraction) between the individual images and the template. A value $\det J(\vec{r}) > 1$ implies that the neighborhood adjacent to $\vec{r}$ in the study was stretched to match the template (i.e. local volumetric expansion), while $\det J(\vec{r}) < 1$ is associated with local shrinkage. When brain images are aligned to the same standard template or atlas, maps of Jacobians can be computed in the atlas coordinate system and used in statistical analysis to identify group differences or localized atrophy at the voxel level. To avoid assuming a normal distribution for the observations at each voxel, a voxel-wise permutation test is performed.

b. Permutation test

Permutation tests are one type of nonparametric test. They were proposed in the early twentieth century, but have only recently become popular with the availability of inexpensive, powerful computers to perform the computations involved.

In permutation tests, weak distributional assumptions are made, which embody the degree of exchangeability [72]. For an experiment with cross-sectional subjects, we might typically assume the following. For a particular voxel, normal and AD brain MR scans within a given block have a distribution with the same shape, though possibly different means. The null hypothesis asserts that the distributions for the normal and AD brain scans have the same mean, and hence are the same. Then the scans are arbitrarily relabeled. For each of the possible relabeling, the statistic of interest is computed; for relabeling i, call this statistic t_i. Under the null hypothesis each of the t_i are equally likely, so the p-value is the proportion of the t_i greater than or equal to the statistic t_0 corresponding to the correctly labeled data.

Single voxel example

Consider assessing the evidence of volume changes at a single voxel of a cross-sectional experiment with six subjects and consequently six MR scans, three of which are AD brains (denoted as A) and other three are normal brains (denoted as B). Suppose that they are nonlinearly aligned to a normal template brain, and the Jacobian determinants at this voxel are $\{1.0300, 0.9993, 0.9976, 0.9048, 0.8783, 0.9606\}$, corresponding to the order of $AAABBB$. For simplicity of illustration we use the "mean difference" as the statistic of interest, i.e., $t_0 = \frac{1}{3}\sum_{j=1}^{3} A_j - \frac{1}{3}\sum_{j=1}^{3} B_j$, where A_j and B_j indicate the value of the jth scan at the particular voxel of interest. We observe the statistic $t_0 = 0.0944$. Suppose that the condition presentation order was randomized, the actual ordering is $AAABBB$. According to combinatorial theory, there are twenty possible outcomes in this randomization scheme. Under the null hypothesis $\mathcal{H}_0$:"The scans would have been the same whatever the experimental condition, A or B", the labels are exchangeable, and the statistic corresponding to the 20 possible labeling are equally likely. The 20 possible labeling are:

(1). $AAABBB$	(2). $AABABB$	(3). $AABBAB$	(4). $AABBBA$	(5). $ABAABB$
(6). $ABABAB$	(7). $ABABBA$	(8). $ABBAAB$	(9). $ABBABA$	(10). $ABBBAA$
(11). $BAAABB$	(12). $BAABAB$	(13). $BAABBA$	(14). $BABAAB$	(15). $BABABA$
(16). $BABBAA$	(17). $BBAAAB$	(18). $BBAABA$	(19). $BBABAA$	(20). $BBBAAA$.

Let t_i be the mean difference for labeling i, t_i for each of the 20 relabeling are computed:

$$
\begin{array}{lllll}
t_1 = +0.0944 & t_2 = +0.0325 & t_3 = +0.0149 & t_4 = +0.0697 & t_5 = +0.0314 \\
t_6 = +0.0137 & t_7 = +0.0686 & t_8 = -0.0481 & t_9 = +0.0067 & t_{10} = -0.0109 \\
t_{11} = +0.0109 & t_{12} = -0.0067 & t_{13} = +0.0481 & t_{14} = -0.0686 & t_{15} = -0.0137 \\
t_{16} = -0.0314 & t_{17} = -0.0697 & t_{18} = -0.0149 & t_{19} = -0.0325 & t_{20} = -0.0944.
\end{array}
$$

Under the null hypothesis the statistics corresponding to these labeling are equally likely. The p-value is the proportion of the permutation distribution greater than or equal to t_0. Here the actual labeling (No. 1 with $t_1 = +0.0944$) gives the largest mean difference of all the possible labeling, so the p-value is $1/20 = 0.05$. For a test at a given α level, we reject the null hypothesis if the p-value is less than α, so we conclude that there is significant evidence against the null hypothesis of no volume changes at this voxel at level $\alpha = 0.05$.

c. Multiple comparisons

Using permutation test, for each voxel we can produce a p-value, p^k, for the null hypothesis $\mathcal{H}_0^k$, where the superscript k indexes the voxels. In the language of multiple comparison, these p-values are *uncorrected* p-values. The overall significance can be assessed by false discovery rate (FDR) although other alternatives exist (i.e. Bonferroni correction). The following simple procedure to control FDR at level α was proposed by Benjamini and Hochberg [73]. For m hypothesis tests, Table 3.1 defines some random variables:

Table 3.1: Random variables in hypothesis test

	# of declared non-significant	# of declared significant	Total
# of true null hypotheses	**U**	**V**	m_0
# of non-true null hypotheses	**T**	**S**	$m - m_0$
Total	$m - R$	R	m

where $\mathcal{H}_0^1, \ldots, \mathcal{H}_0^m$ denote the null hypotheses being tested, m_0 is the number of true null hypotheses, $m - m_0$ is the number of false null hypotheses, **U** is the number of true negatives, **V** is the number of false positives, **T** is the number of false negatives, **S** is the number of true positives. In m hypothesis tests of which m_0 are true null hypotheses, R is an observable random variable, and **S**, **T**, **U**, and **V** are unobservable random variables. The false discovery rate is given by $E\left[\frac{\mathbf{V}}{\mathbf{V}+\mathbf{S}}\right] = E\left[\frac{\mathbf{V}}{R}\right]$ ($\frac{\mathbf{V}}{R}$ is defined to be 0 when $R = 0$) and one wants to keep this value below a threshold γ. When the m hypothesis tests are independent, we rank the p-values in ascending order, and denote by $\mathcal{H}_0^{(i)}$ the null hypothesis corresponding to $P_{(i)}$.

For a given γ, find the largest k such that $P_{(k)} \leq \frac{k}{m}\gamma$. Then reject (i.e. declare positive) all $\mathcal{H}_0^{(i)}$ for $i = 1, \ldots, k$.

In this chapter, the labels of patients and controls are randomly reassigned in 5000 permutations, and the p-values are compared. The FDR is controlled on the p-value map. The FDR-corrected p-values less than 0.05 are accepted, indicating 95% confidence of no false positives in the image. This represents a typical voxel-wise testing approach.

3.3.3 Voxel-based morphometry (VBM)

The second class of voxel-wise morphometry is voxel-based morphometry (VBM) [22]. Instead of studying brain shapes based on deformation fields, VBM is applied to some scalar function of the normalized image, performing statistical comparison of gray matter partitions. The image processing in VBM involves spatially normalizing all the images to a standard space, extracting the gray matter from the normalized image and smoothing. A statistical analysis is finally performed to make inference about group differences.

The VBM was performed with *FSL-VBM* (http://www.fmrib.ox.ac.uk/fsl/fslvbm/index.html), a voxel-based morphometry style analysis [22] carried out with the *FSL* tool [59]. The images are first skull-stripped using *BET* [60] and tissues are segmented using *FAST4* [61]. The gray matter segmentations are then affinely and non-linearly aligned to the MNI152 standard space[2] using registration tools *FLIRT* [62, 63] and *FNIRT* [64, 65], respectively. The resulting images were averaged to create a study-specific template, to which the original gray matter images were non-linearly re-registered. The registered gray matter images were then modulated by dividing by the Jacobian of the warped field. A Gaussian kernel with a sigma of 3 mm was used to smooth the modulated images. The voxel-wise general linear model (GLM) was finally applied using permutation testing. Multiple-comparison correction is again performed to control the family-wise error (FWE) rate at a 0.05 level. The family-wise error rate is the probability of making one or more false discoveries, or type I errors among all the hypotheses when performing multiple pair-wise tests [73, 76]. Using the same random variables defined in Table 3.1, $FWE = Pr(\mathbf{V} \geq 1) = 1 - Pr(\mathbf{V} = 0)$.

[2]MNI152 is a canonical template which is constructed by the Montreal Neurological Institute using 152 normal MRI scans. Details of MNI152 can be found at http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach.

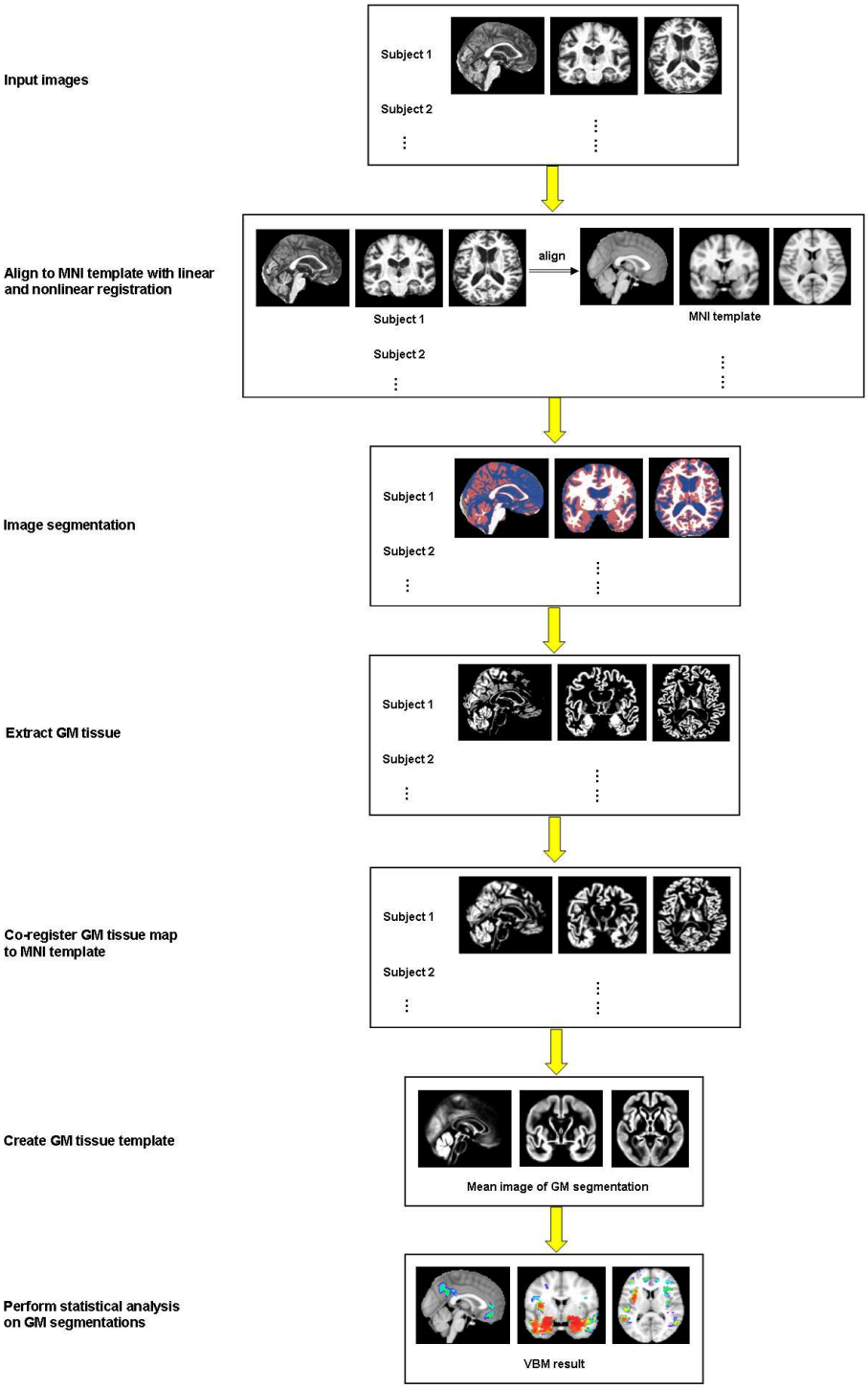


Figure 3.5: Voxel-based morphometry pipeline.

3.4 Experiments and results

In this dissertation, the proposed regional classification-based method along with the cortical thickness measurement, TBM and VBM are applied to clinical MR brain images selected from the ADNI database (www.loni.ucla.edu/ADNI) and the OASIS database [49]. The images in the ADNI database are geometry distortion corrected, intensity nonuniformity corrected, and histogram peak sharpened. The images in the OASIS are provided skull-stripped, bias corrected, and affinely registered to the Talairach atlas space.

The subjects in the datasets are clinically diagnosed using the Clinical Dementia Rating (CDR) scale, which combines the information from mental examination, patient history and interviews with family members or caregivers [50]. This score is used to characterize and to track a patient's level of impairment: 0 = Normal, 0.5 = Very Mild Dementia, 1 = Mild Dementia, 2 = Moderate Dementia, 3 = Severe Dementia.

In our classification-based method, the segmentations of gray matter (GM), white matter (WM), CSF, cortical regions and subcortical structures are obtained using the *FreeSurfer* software (http://surfer.nmr.mgh.harvard.edu/). Pair-wise registration is done by using the symmetric log-domain diffeomorphic demons algorithm with default parameters (the parameter setting have been present in Chapter 2.2.3), providing deformation fields and logarithms of deformation fields (using "log fields" for shorthand as described in Chapter 2). Riemannian regional distance between images is measured from the log field using the method introduced in Chapter 2.2.1. In our experiments, the dimension of the embedded space is selected as 3. Thus all images are projected onto an $\mathbb{R}^3$ Euclidean space, using the MDS embedding. A linear discriminant hyperplane is then trained on the embedded space using the labeled subjects. A test unlabeled subject is projected onto the constructed space by out-of-sample embedding. The leave-one-out cross validation is performed to calculate the classification accuracy with different regional distances.

The classification tests using the distance on the whole brain, gray matter, white matter and CSF have been already shown in our previous paper [123] and in Chapter 2. The gray matter has been observed to be one of the regions with most significant structural changes. In this chapter, we don't repeat the same testing again, but mainly focus on analyzing the sub-regions of interest within gray matter, including the cortical regions and subcortical structures (Fig. 3.6). The cortical regions involve frontal, central, parietal, occipital, temporal lobes, cingulate, and insula cortex (more subtle segmentations of the cortex is shown in Appendix 1). The selected subcortical structures are hippocampus, amygdala, thalamus,

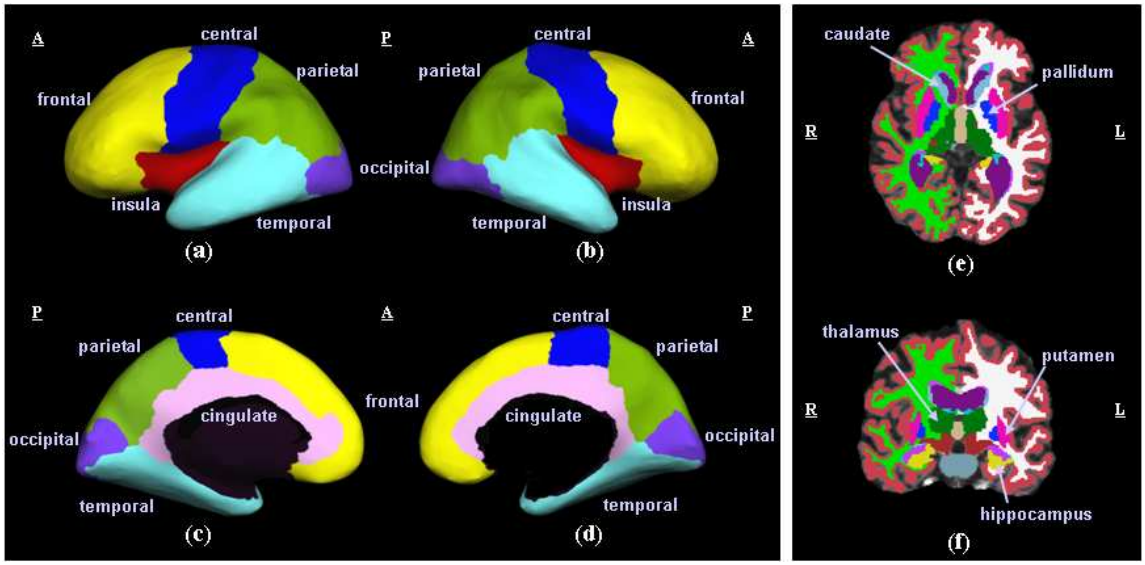


Figure 3.6: Cortical segmentation and subcortical segmentation. (a)∼(d) Inflated cortical representations of the regions of interest in both hemispheres. (a) The lateral view of the left hemisphere; (b) The lateral view of the right hemisphere; (c) The medial view of the left hemisphere; (d) The medial view of the right hemisphere. (e)∼(f) Subcortical representations of the regions of interest. (e) Axial view; (f) Coronal view.

caudate, putamen, and pallidum.

3.4.1 Experiment 1 on the ADNI normal *vs* MCI converter (NC-cMCI) dataset

The first dataset consists of 83 MR scans selected from the ADNI database, with 39 brains from MCI subjects (aged 56.25∼87.81) which later had progressed to AD and 44 age-matched normal controls (aged 59.98∼89.67). The diagnosis characteristics is listed in Table 3.2.

Table 3.2: Age and diagnosis characteristics of the ADNI (NC-cMCI) dataset

Groups	# of subjs	Age–mean (range)	MMSE–mean (range)	CDR at scanning	latter CDR
Normal	44	76.06 (59.98∼89.67)	27.98 (24∼30)	0	0
Patient	39	75.48 (56.25∼87.81)	20.97 (5∼30)	0.5	1∼2

NC=Normal Control; cMCI=converting Mild Cognitive Impairment (MCI subjects which later progress to AD); MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

The classification accuracy along with sensitivity and specificity using distances from the

selected subcortical structures and cortical regions are shown in Table 3.3 and Fig. 3.7. The highest classification rates among all regions of interest were around 87%, indicating moderate differences between the control group and the patient group appearing on the corresponding areas: hippocampus, putamen, temporal lobe and cingulate cortex. Differences are also found in frontal, central and parietal cortex between the two groups.

Table 3.3: Results of classification using distance metrics on subcortical structures and cortical regions in Experiment 1

Regions	***Hippocampus***	***Amygdala***	***Thalamus***	***Caudate***	***Putamen***	***Pallidum***
Accuracy	0.86	0.80	0.66	0.81	0.86	0.77
Sensitivity*	0.87	0.77	0.69	0.74	0.82	0.72
Specificity	0.84	0.82	0.64	0.86	0.89	0.82
Regions	***Frontal***	***Central***	***Parietal***	***Occipital***	***Temporal***	***Cingulate***	***Insula***
Accuracy	0.84	0.86	0.83	0.77	0.88	0.87	0.72
Sensitivity*	0.87	0.85	0.79	0.74	0.85	0.90	0.72
Specificity	0.82	0.86	0.86	0.80	0.91	0.84	0.73

∗ Considering a correctly identified diseased case as a true positive.

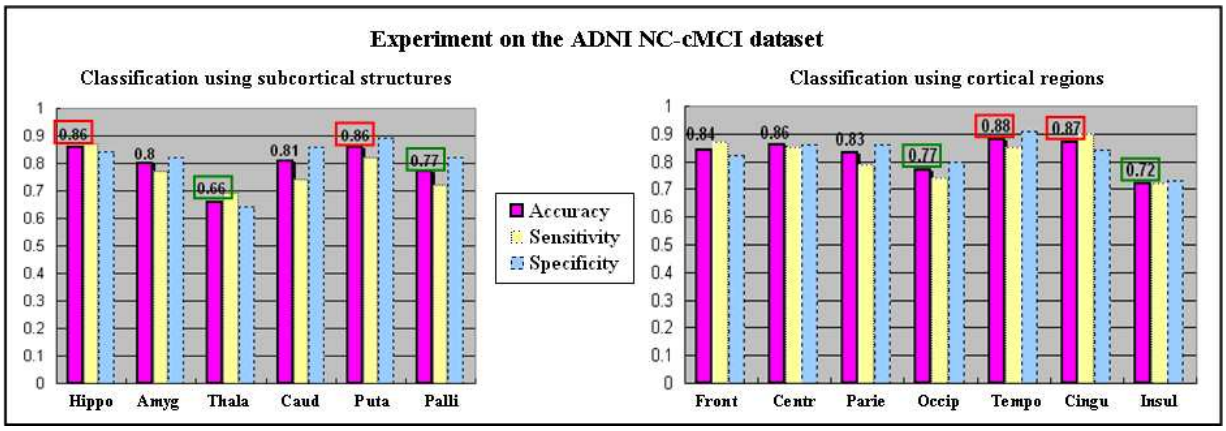


Figure 3.7: Classification with different regions of interest on the ADNI (NC-cMCI) dataset. Hippo=hippocampus; Amyg=amygdala; Thala=thalamus; Caud=caudate; Puta =putamen; Palli=Pallidum; Front=frontal; Centr=central; Parie=parietal; Occip=occipital; Tempo=temporal; Cingu=cingulate; Insul=insula.

We further segmented the temporal lobe into eight sub-regions (Fig. 3.8): fusiform, inferior temporal, lingual region, medial temporal, parahippocampal/entorhinal cortex, superior temporal, temporal pole and transverse temporal, to explore pathology distribution in more subtle areas of the cortex. The classification results are shown in Table 3.4 and Fig. 3.9. The

classification accuracy with parahippocampal/entorhinal cortex, superior and medial temporal lobe were the highest three among the eight segmented subregions, indicating larger changes occur within these areas.

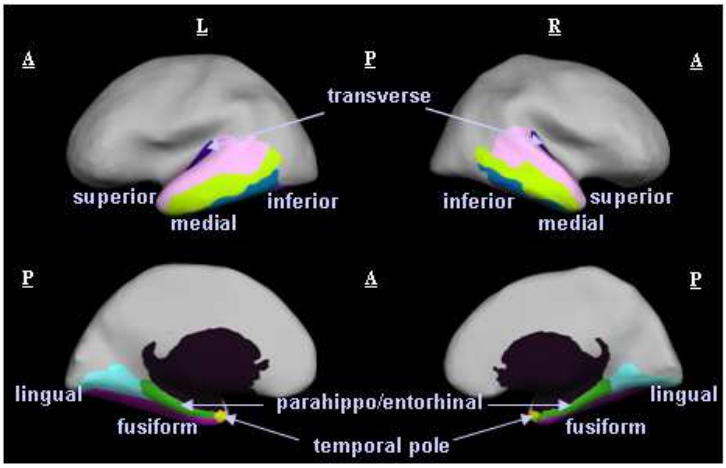


Figure 3.8: Further segmentations in temporal lobe. A=Anterior; P=Posterior; L=Left; R=Right.

Table 3.4: Results of classification using distance metrics on temporal sub-regions

Regions	*Fusiform*	*Inferior*	*Parahippo/Entorhinal*	*Lingual*	*Medial*	*Pole*	*Superior*	*Transverse*
Accuracy	0.80	0.83	0.90	0.80	0.87	0.77	0.89	0.76
Sensitivity*	0.82	0.77	0.90	0.82	0.85	0.72	0.87	0.74
Specificity	0.77	0.89	0.91	0.77	0.89	0.82	0.91	0.77

∗ Considering a correctly identified diseased case as a true positive.

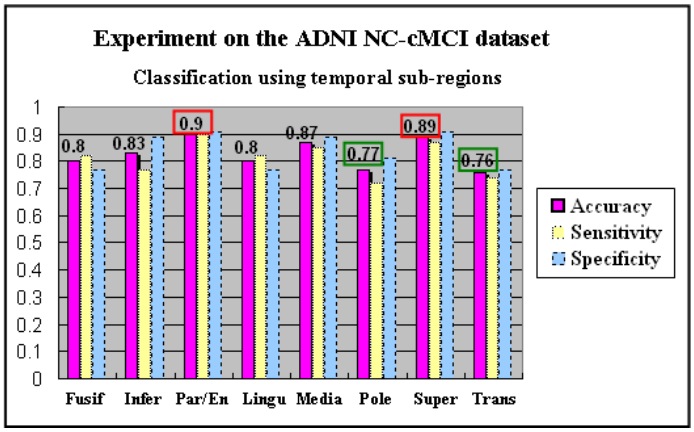


Figure 3.9: Classification with different sub-regions of temporal lobe on the ADNI (NC-cMCI) dataset. Fusif=fusiform; Infer=inferior; Par/En=parahippocampal/ entorhinal; Lingu=lingual; Media=Medial; Super=superior; Trans=transverse.

We reproduced the cortical thickness measurement, TBM and VBM on this dataset for comparison. In Fig. 3.10, the thickness difference between the patient group and the control group was analyzed. The magnitude represents $|log(p)|$, the positive sign denotes expansion in the patient group and the negative sign denotes shrinkage in the patient group. Significant cortical thinning was found in temporal, frontal, parietal lobe and cingulate cortex, with the most pronounced changes occurring in the medial temporal, parahippocampal, entorhinal and isthmus/precuneus region (readers can refer to Appendix 1 for region correspondence). This is consistent with other studies [118, 119, 120].

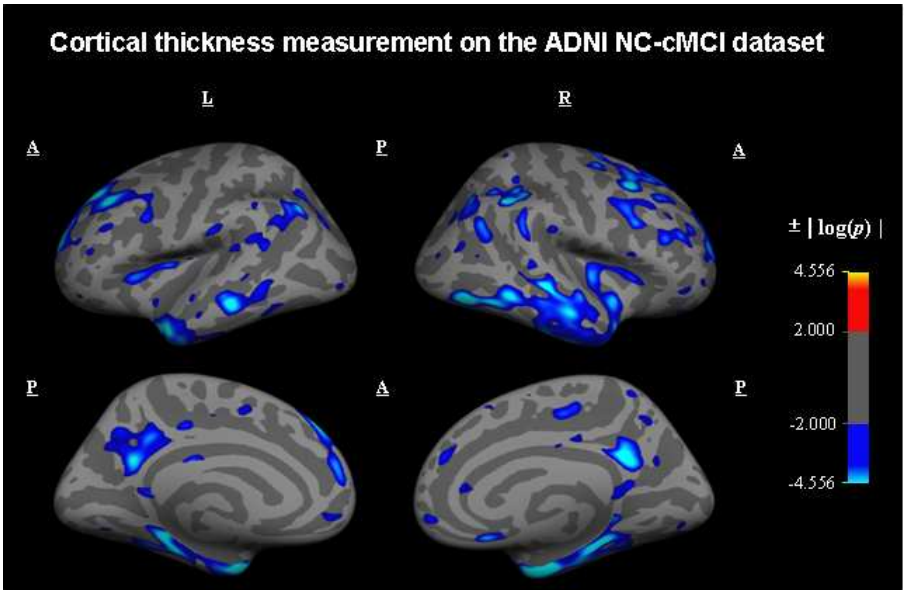


Figure 3.10: Cortical thickness measurement on the ADNI NC-cMCI dataset. A=Anterior; P=Posterior; L=Left; R=Right. The magnitude represents $|log(p)|$, the positive sign denotes significant expansion in the cMCI patient group and the negative sign denotes significant shrinkage in the cMCI patient group. $|log(p)| > 2 \Leftrightarrow p < 0.01$, false discovery rate is corrected.

The TBM and VBM results are shown in Fig. 3.11. In TBM, we observed significant changes in thalamus, putamen, hippocampus and parahippocampal cortex. In VBM, volumetric reduction was also found between the two groups, located in putamen, amygdala, medial temporal, hippocampal, parahippocampal and entorhinal cortex.

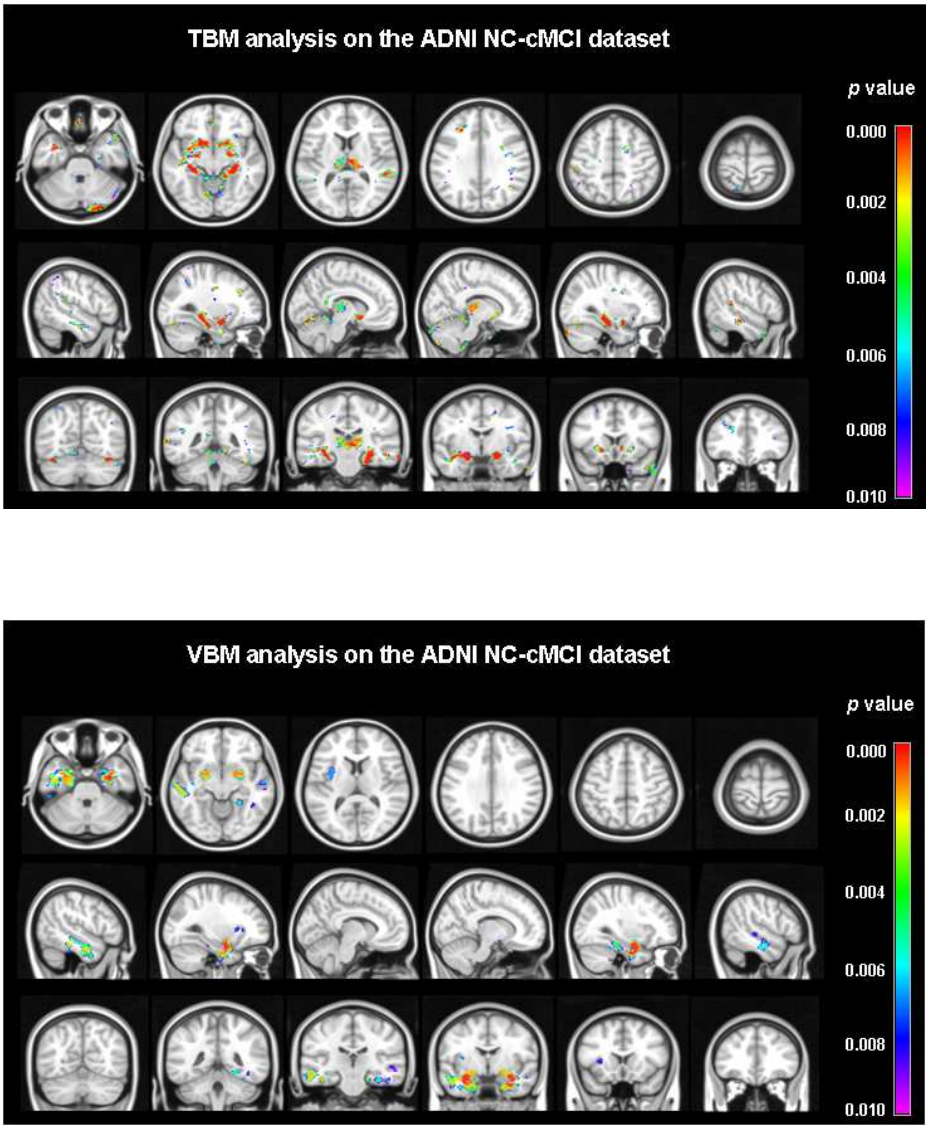


Figure 3.11: TBM and VBM analysis on the ADNI NC-cMCI dataset. $p < 0.01$, false discovery rate is corrected in TBM and family-wise error rate is corrected in VBM.

3.4.2 Experiment 2 on the ADNI cMCI-AD dataset

The dataset used in this experiment consists of 72 MR scans also selected from the ADNI database, with 39 brains from MCI subjects (aged 56.25∼87.81) which later had progressed to AD and 33 age-matched AD patients (aged 58.75∼88.84). The diagnosis characteristics

is listed in Table 3.5. This experiment is designed to identify the regional differences before and after AD pathology sets in.

Table 3.5: Age and diagnosis characteristics of the ADNI (cMCI-AD) dataset

Groups	# of subjs	Age–mean (range)	MMSE–mean (range)	CDR at scanning	latter CDR
MCI converter	39	75.48 (56.25∼87.81)	20.97 (5∼30)	0.5	1∼2
AD patient	33	77.66 (58.75∼88.84)	15.64 (2∼26)	1∼3	1∼3

cMCI=converting Mild Cognitive Impairment (MCI subjects which later progress to AD); AD=Alzheimer's disease; MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

Table 3.6: Results of classification using distance metrics on subcortical structures and cortical regions in Experiment 2

Regions	***Hippocampus***	***Amygdala***	***Thalamus***	***Caudate***	***Putamen***	***Pallidum***
Accuracy	0.93	0.94	0.85	0.79	0.89	0.88
Sensitivity*	0.91	0.94	0.85	0.76	0.88	0.91
Specificity	0.95	0.95	0.85	0.82	0.90	0.85

Regions	***Frontal***	***Central***	***Parietal***	***Occipital***	***Temporal***	***Cingulate***	***Insula***
Accuracy	0.88	0.93	0.83	0.88	0.97	0.86	0.86
Sensitivity*	0.91	0.97	0.85	0.88	1.00	0.88	0.82
Specificity	0.85	0.90	0.82	0.87	0.95	0.85	0.90

∗ Considering a correctly identified AD case as a true positive.

The classification accuracy along with sensitivity and specificity are shown in Table 3.6 and Fig. 3.12. From the results, we observed hippocampus, amygdala, temporal and cingulate cortex producing higher classification accuracy, with correct rates higher than 93%, indicating more significant changes occurring in these areas. Accuracy rates using other regions of interest were almost all greater than 80%, showing that deviations appear in most of the gray matter region.

Cortical thickness measurement, TBM and VBM were also performed on this dataset for comparison. In Fig. 3.13 the resulting map of atrophy showed region specificity of thickness decline in AD. The magnitude represents $|log(p)|$, the positive sign denotes expansion in the AD group and the negative sign denotes shrinkage in the AD group. The most significant changes were found in temporal, frontal, cingulate and insula cortex. Most of these effects were found bilaterally.

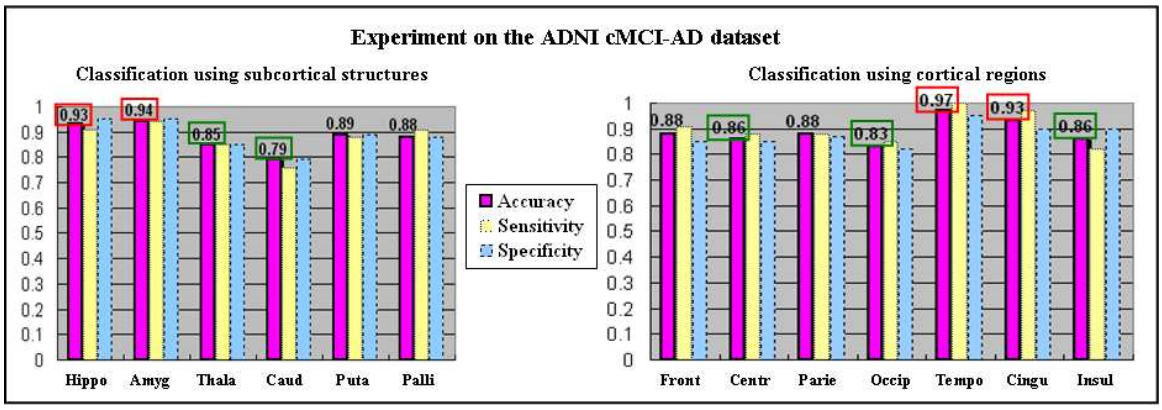


Figure 3.12: Classification with different regions of interest on the ADNI (cMCI-AD) dataset. Hippo=hippocampus; Amyg=amygdala; Thala=thalamus; Caud=caudate; Puta =putamen; Palli=Pallidum; Front=frontal; Centr=central; Parie=parietal; Occip=occipital; Tempo=temporal; Cingu=cingulate; Insul=insula.

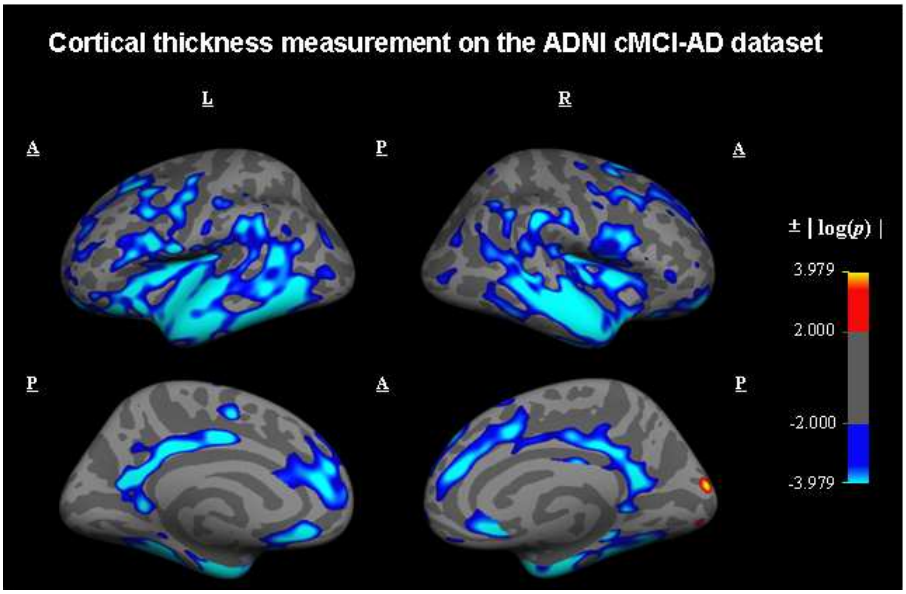


Figure 3.13: Cortical thickness measurement on the ADNI cMCI-AD dataset. A=Anterior; P=Posterior; L=Left; R=Right. The magnitude represents $|log(p)|$, the positive sign denotes significant expansion in the AD patient group and the negative sign denotes significant shrinkage in the AD patient group. $|log(p)| > 2 \Leftrightarrow p < 0.01$, false discovery rate is corrected.

In TBM result, striking differences appeared around the ventricles, and hippocampus-amygdala regions. Compared with TBM, VBM found more changes, located in temporal, frontal, cingulate and insula areas. TBM and VBM results are shown in Fig. 3.14.

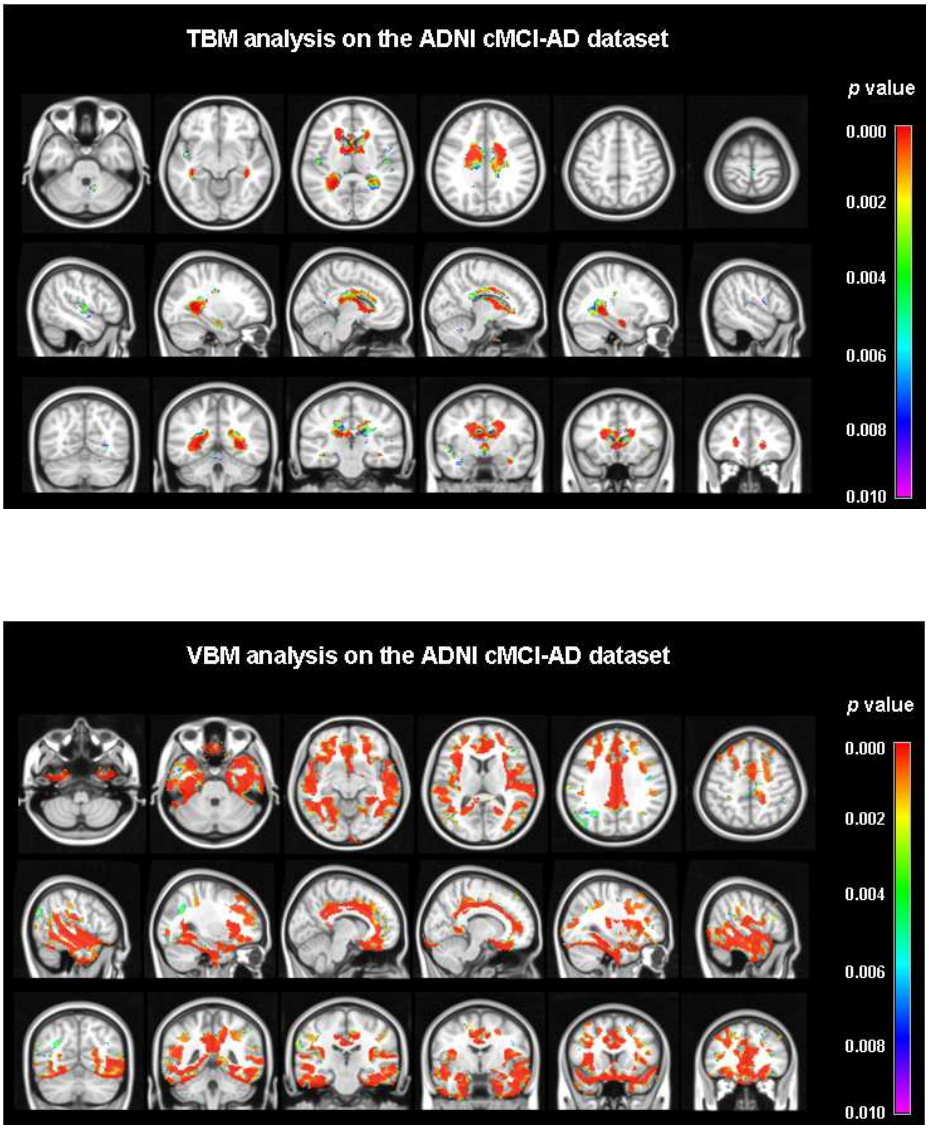


Figure 3.14: TBM and VBM analysis on the ADNI cMCI-AD dataset. $p < 0.01$, false discovery rate is corrected in TBM and family-wise error rate is corrected in VBM.

3.4.3 Experiment 3 on the OASIS normal *vs* mild AD (NC-mAD) dataset

The dataset used in the third experiment consists of 75 MR scans selected from the OASIS database, with 35 brains from subjects (aged 69∼96) with very mild to mild dementia and 40 age-matched normal controls (aged 67∼91). The diagnosis characteristics is listed in Table 3.7.

Table 3.7: Age and diagnosis characteristics of the OASIS (NC-mAD) dataset

Groups	# of subjs	Age–mean (range)	MMSE–mean (range)	CDR at scanning	latter CDR
Normal	40	77.28 (67∼91)	28.98 (26∼30)	0	0
AD patient	35	80.37 (69∼96)	22.63 (15∼29)	0.5∼1	0.5∼1

NC=Normal Control; mAD=mild Alzheimer's disease; MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

The classification results are listed in Table 3.8 and Fig. 3.15. Among all the regions of interest used for classification, amygdala, thalamus, parietal lobe and cingulate cortex were observed to better separate the AD patients and normal aging subjects (with accuracy up to 95%), indicating that significant deformations between AD brains and normal brains can be found in these regions.

Table 3.8: Results of classification using distance metrics on subcortical structures and cortical regions in Experiment 3

Regions	***Hippocampus***	***Amygdala***	***Thalamus***	***Caudate***	***Putamen***	***Pallidum***
Accuracy	0.83	0.93	0.89	0.83	0.75	0.75
Sensitivity*	0.83	0.94	0.86	0.86	0.80	0.66
Specificity	0.83	0.93	0.93	0.80	0.70	0.83

Regions	***Frontal***	***Central***	***Parietal***	***Occipital***	***Temporal***	***Cingulate***	***Insula***
Accuracy	0.88	0.88	0.93	0.91	0.91	0.95	0.77
Sensitivity*	0.94	0.83	0.91	0.97	0.94	0.94	0.80
Specificity	0.83	0.93	0.95	0.85	0.88	0.95	0.75

∗ Considering a correctly identified diseased case as a true positive.

Classification on temporal lobe and frontal lobe also gave relatively good discrimination, with correct classification rate around 90%. However, the central lobe and insula region was relatively preserved in AD.

We further segmented the parietal lobe and cingulate cortex to find pathology distribution

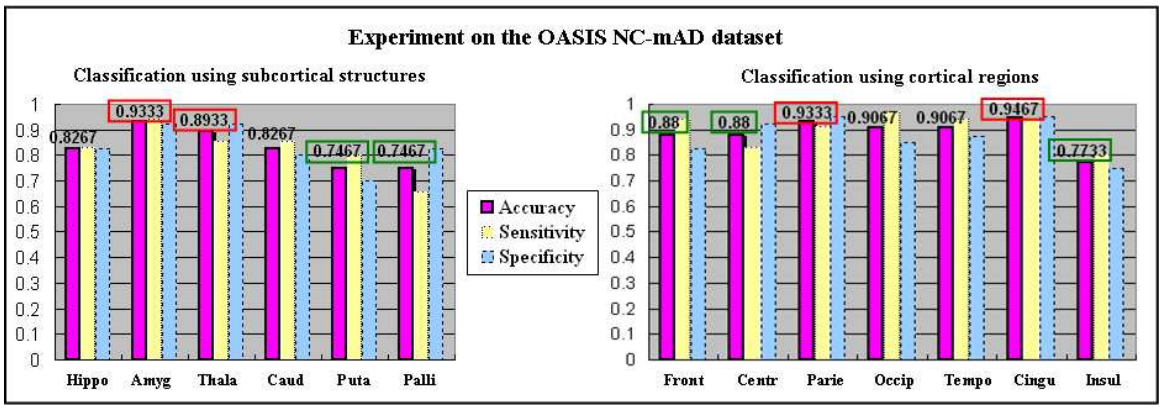
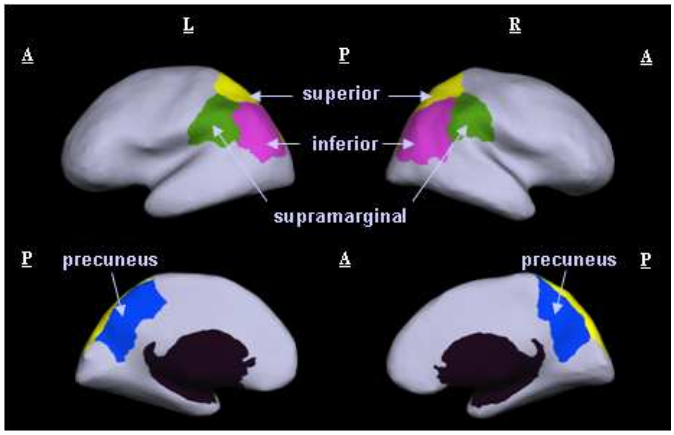


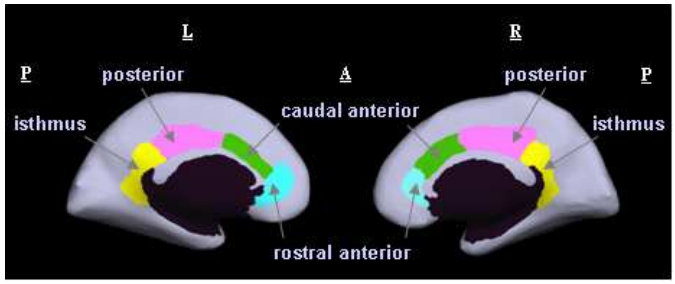
Figure 3.15: Classification with different regions of interest on the OASIS (NC-mAD) dataset. Hippo=hippocampus; Amyg=amygdala; Thala=thalamus; Caud= caudate; Puta=putamen; Palli=Pallidum; Front=frontal; Centr=central; Parie =parietal; Occip=occipital; Tempo=temporal; Cingu=cingulate; Insul=insula.

in more subtle areas. The parietal lobe was divided into inferior parietal, precuneus, supramarginal and superior parietal (as shown in Fig. 3.16(a)). The cingulate cortex was split into caudal anterior (CA) cingulate, isthmus, posterior cingulate and rostral anterior (RA) cingulate (as shown in Fig. 3.16(b)). The classification results with parietal sub-regions and cingulate sub-regions are shown in Fig. 3.17. The accuracy using parietal sub-regions were all higher than 88%. The most significant changes were found in precuneus and supramarginal region. Within cingulate cortex, we observed most pronounced deformations in the isthmus area (as shown in Fig. 3.17).

Cortical thickness measurement, TBM and VBM are again performed for comparison. In cortical thickness measurement, shrinkage was found almost throughout the cortex, especially in the posterior regions including parietal, superior temporal, parahippocampus/entorhinal, right occipital and isthmus region (as shown in Fig. 3.18). Gray matter changes were also detected in TBM and VBM study. In Fig. 3.19, thalamus, putamen, temporal and parietal changes were observed in both TBM and VBM results. In addition, deviations in amygdala and other cortical regions were found in VBM.



(a)



(b)

Figure 3.16: Further segmentations in parietal lobe and cingulate cortex. A=Anterior; P=Posterior; L=Left; R=Right.

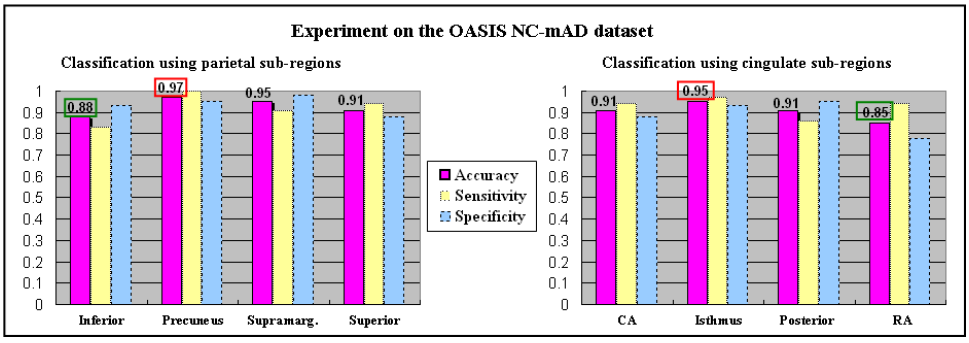


Figure 3.17: Classification with different sub-regions of parietal lobe and cingulate cortex on the OASIS (NC-mAD) dataset. Supramarg=Supramarginal; CA=Caudal Anterior; RA=Rostral Anterior.

Table 3.9: Results of classification using distance metrics on parietal and cingulate sub-regions

Regions	*Inferior*	*Precuneus*	*SupraMarginal*	*Superior*
Accuracy	0.88	0.97	0.95	0.91
Sensitivity*	0.83	1.00	0.91	0.94
Specificity	0.93	0.95	0.98	0.88

Regions	*Caudal Anterior*	*Isthmus*	*Posterior*	*Rostral Anterior*
Accuracy	0.91	0.95	0.91	0.85
Sensitivity*	0.94	0.97	0.86	0.94
Specificity	0.88	0.93	0.95	0.78

$*$ Considering a correctly identified diseased case as a true positive.

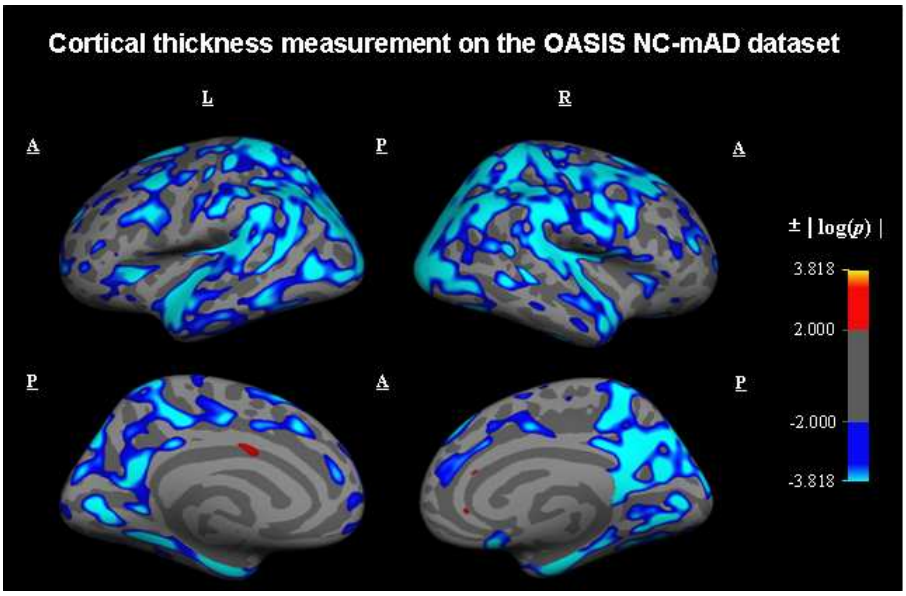


Figure 3.18: Cortical thickness measurement on the OASIS NC-mAD dataset. A=Anterior; P=Posterior; L=Left; R=Right. The magnitude represents $|log(p)|$, the positive sign denotes significant expansion in the mAD patient group and the negative sign denotes significant shrinkage in the mAD patient group. $|log(p)| > 2 \Leftrightarrow p < 0.01$, false discovery rate is corrected.

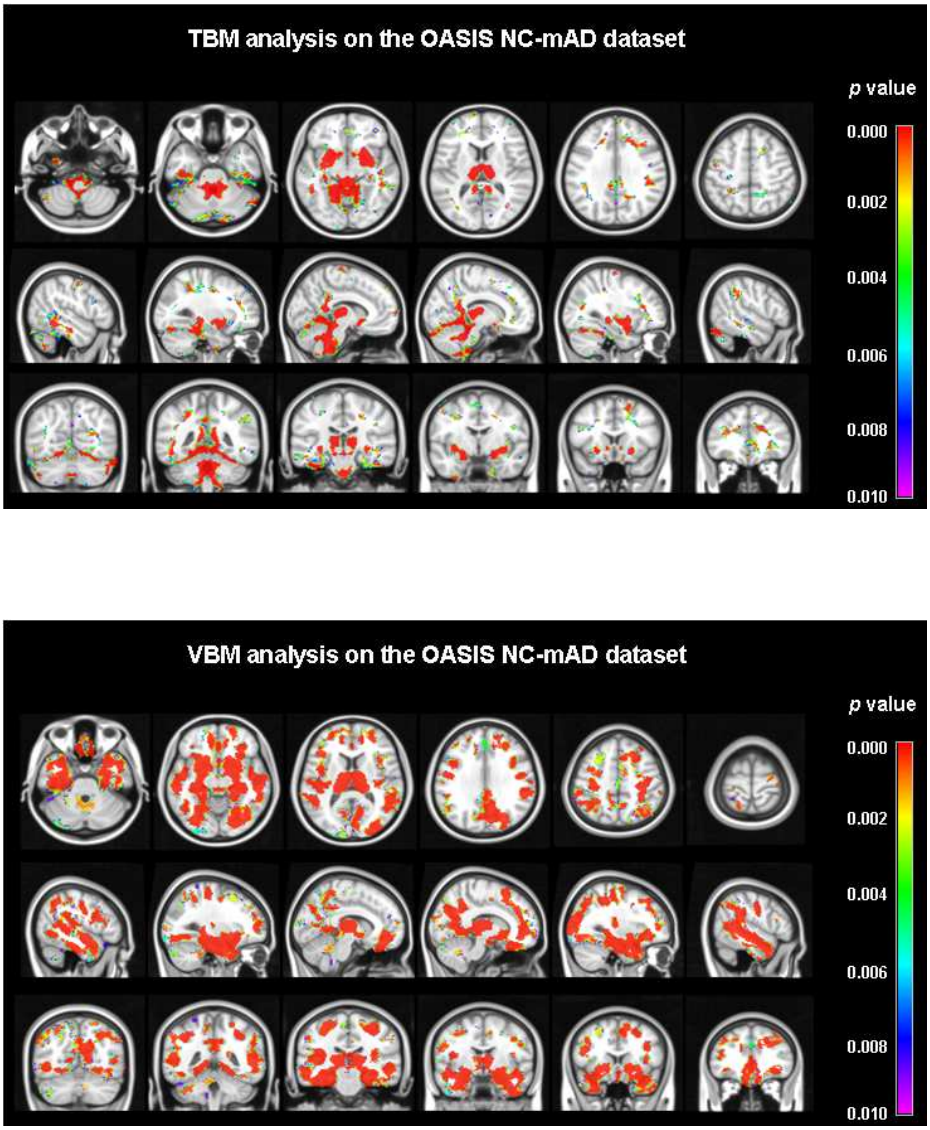


Figure 3.19: TBM and VBM analysis on the OASIS NC-mAD dataset. $p < 0.01$, false discovery rate is corrected in TBM and family-wise error rate is corrected in VBM.

3.4.4 Experiment 4 on the OASIS normal *vs* potential MCI (NC-pMCI) dataset

The dataset used in the fourth experiment consists of 17 MR scans selected from the OASIS database, with 10 normal controls (aged 61∼80) and 7 brains from potential diseased subjects (aged 69∼87) which were healthy at the time of scanning but had converted to MCI after 4±1.41 years. The diagnosis characteristics is listed in Table 3.10.

Table 3.10: Age and diagnosis characteristics of the OASIS (NC-pMCI) dataset

Groups	# of subjs	Age–mean (range)	MMSE–mean (range)	CDR at scanning	latter CDR
Normal	10	73.60 (61∼80)	29.30 (28∼30)	0	0
AD patient	7	78.43 (69∼87)	29.28 (27∼30)	0	0.5

NC=Normal Control; pMCI=potential Mild Cognitive Impairment (normal subjects which later progress to MCI); MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

Our classification-based pathology localization, cortical thickness measurement, TBM and VBM were performed on this dataset. Although no significant changes were detected in VBM analysis, the classification accuracy using different regions of interest differs. The results are shown in Table 3.11 and Fig. 3.20. We observed relatively higher discrimination accuracy between the normal controls and potential MCI subjects in regions such as hippocampus, caudate, thalamus, frontal lobe, central lobe and parietal lobe. Small regions with gray matter thinning were also found by cortical thickness measurement (as shown in Fig. 3.21), located in frontal, central and parietal area. In TBM, changes were observed around ventricles, and also in hippocampus, frontal and parietal regions (as shown in Fig. 3.22).

Table 3.11: Results of classification using distance metrics on subcortical structures and cortical regions in Experiment 4

Regions	***Hippocampus***	***Amygdala***	***Thalamus***	***Caudate***	***Putamen***	***Pallidum***
Accuracy	0.94	0.59	0.82	0.88	0.41	0.29
Sensitivity*	0.86	0.57	0.57	0.86	0.43	0.57
Specificity	1.00	0.60	1.00	0.90	0.40	0.10

Regions	***Frontal***	***Central***	***Parietal***	***Occipital***	***Temporal***	***Cingulate***	***Insula***
Accuracy	0.82	0.82	0.82	0.76	0.71	0.64	0.47
Sensitivity*	0.86	0.71	0.86	1.00	0.57	0.57	0.43
Specificity	0.80	0.90	0.80	0.60	0.80	0.70	0.50

∗ Considering a correctly identified diseased case as a true positive.

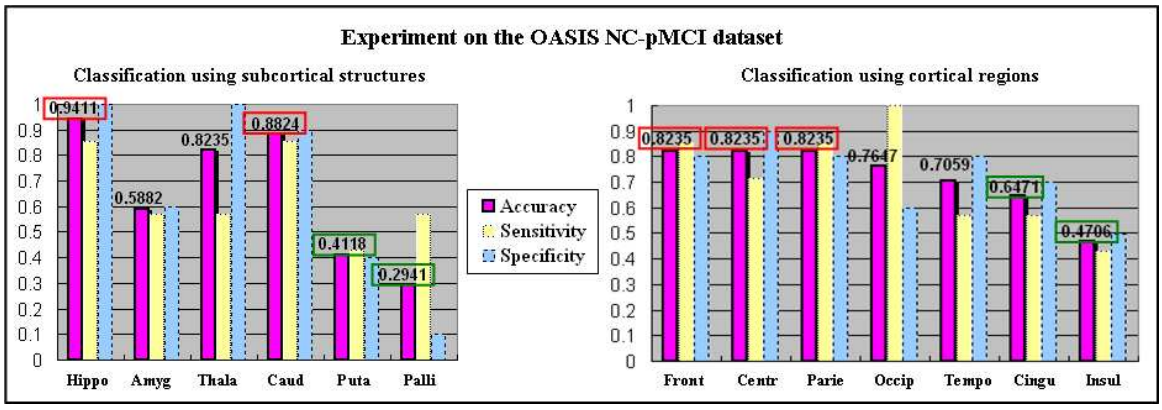


Figure 3.20: Classification with different regions of interest on the OASIS (NC-pMCI) dataset. Hippo=hippocampus; Amyg=amygdala; Thala=thalamus; Caud= caudate; Puta=putamen; Palli=Pallidum; Front=frontal; Centr=central; Parie =parietal; Occip=occipital; Tempo=temporal; Cingu=cingulate; Insul=insula.

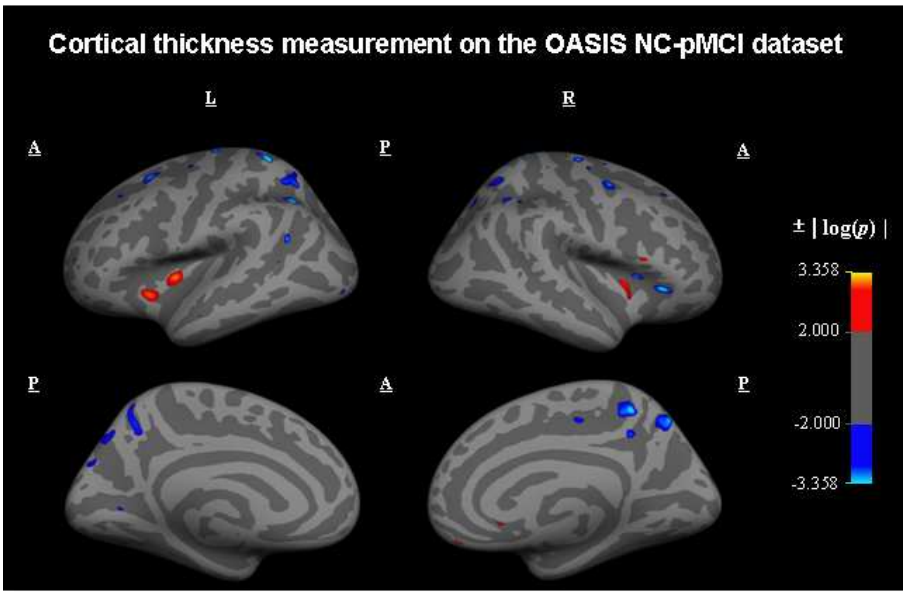


Figure 3.21: Cortical thickness measurement on the OASIS NC-pMCI dataset. A=Anterior; P=Posterior; L=Left; R=Right. The magnitude represents $|log(p)|$, the positive sign denotes significant expansion in the pMCI patient group and the negative sign denotes significant shrinkage in the pMCI patient group. $|log(p)| > 2 \Leftrightarrow p < 0.01$, false discovery rate is corrected.

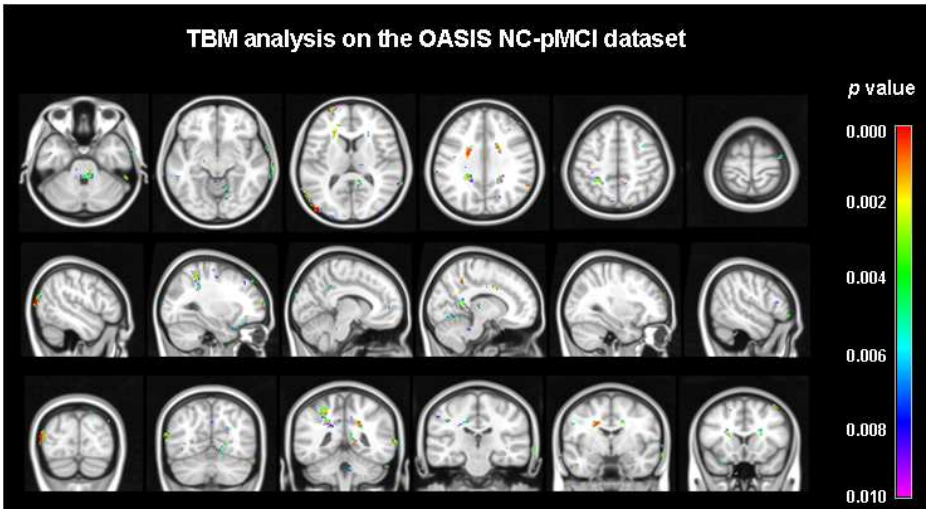


Figure 3.22: TBM and VBM analysis on the OASIS NC-pMCI dataset. $p < 0.01$, false discovery rate is corrected.

3.4.5 Experiment 5 on the ADNI stable MCI *vs* converting MCI (sMCI-cMCI) dataset

The dataset used in this experiment consists of 38 MR scans selected from the ADNI database, with 19 stable MCI subjects (aged 57.92∼87.81) who had not converted to AD within 2.81±0.61 years and 19 brains from converting MCI subjects (aged 56.25∼87.24) who had converted to AD in 1.88±0.62 years. The diagnosis characteristics is listed in Table 3.12. This experiment is designed to identify the regional differences between MCI subjects converting and non-converting to AD.

Table 3.12: Age and diagnosis characteristics of the ADNI (sMCI-cMCI) dataset

Groups	# of subjs	Age–mean (range)	MMSE–mean (range)	CDR at scanning	latter CDR
stable MCI	19	75.16 (57.92∼87.81)	26.95 (24∼29)	0.5	0.5
converting MCI	19	77.36 (56.25∼87.24)	26.05 (24∼29)	0.5	1∼2

sMCI=stable Mild Cognitive Impairment; cMCI = converting Mild Cognitive Impairment (subjects with mild cognitive impairment, which latter progress to AD); MMSE=Mini Mental State Examination; CDR=Clinical Dementia Rating.

As in the previous experiments, region-based classification, cortical thickness measurement, TBM and VBM were performed on this dataset. VBM again cannot detect significant changes between the two groups. Region-based classification results are shown in Table 3.13

and Fig. 3.23. The accuracy was not as high as in the above four experiments, indicating smaller deviation appears between stable MCI group and converting MCI group. The same observations were found in both cortical thickness measurement (as shown in Fig. 3.24, with positive sign denoting expansion in stable MCI group and negative sign denoting shrinkage in sMCI group), and TBM analysis (as shown in Fig. 3.25). However, the consistent result in classification, cortical thickness measurement and TBM was that parietal and temporal lobe showed most pronounced differences between the two groups. Changes around hippocampus were also detected in our classification and TBM.

Table 3.13: Results of classification using distance metrics on subcortical structures and cortical regions in Experiment 5

Regions	***Hippocampus***	***Amygdala***	***Thalamus***	***Caudate***	***Putamen***	***Pallidum***
Accuracy	0.79	0.68	0.68	0.47	0.68	0.68
Sensitivity*	0.95	0.74	0.68	0.63	0.74	0.63
Specificity	0.63	0.63	0.68	0.32	0.63	0.74

Regions	***Frontal***	***Central***	***Parietal***	***Occipital***	***Temporal***	***Cingulate***	***Insula***
Accuracy	0.53	0.61	0.82	0.71	0.74	0.68	0.53
Sensitivity*	0.53	0.63	0.84	0.74	0.84	0.63	0.47
Specificity	0.53	0.58	0.79	0.68	0.63	0.74	0.58

∗ Considering a correctly identified diseased case as a true positive.

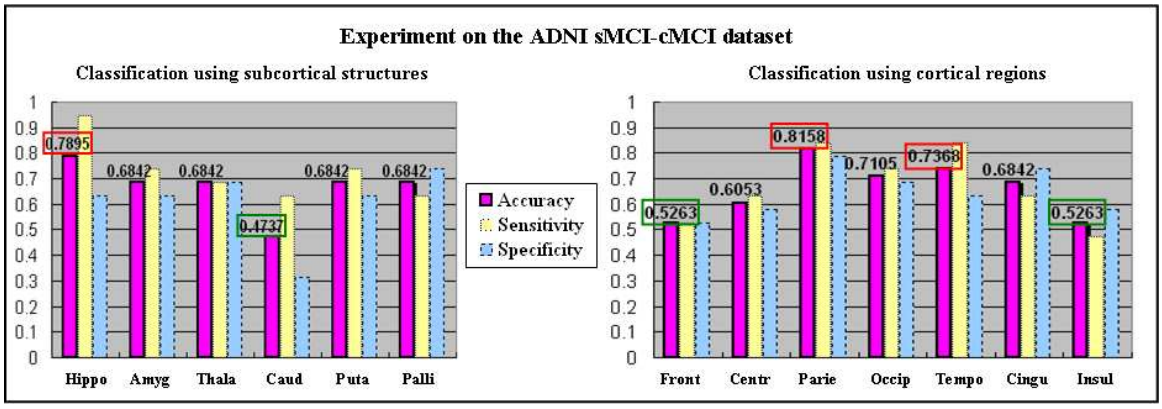


Figure 3.23: Classification with different regions of interest on the ADNI (sMCI-cMCI) dataset. Hippo=hippocampus; Amyg=amygdala; Thala=thalamus; Caud= caudate; Puta=putamen; Palli=Pallidum; Front=frontal; Centr=central; Parie =parietal; Occip=occipital; Tempo=temporal; Cingu=cingulate; Insul=insula.

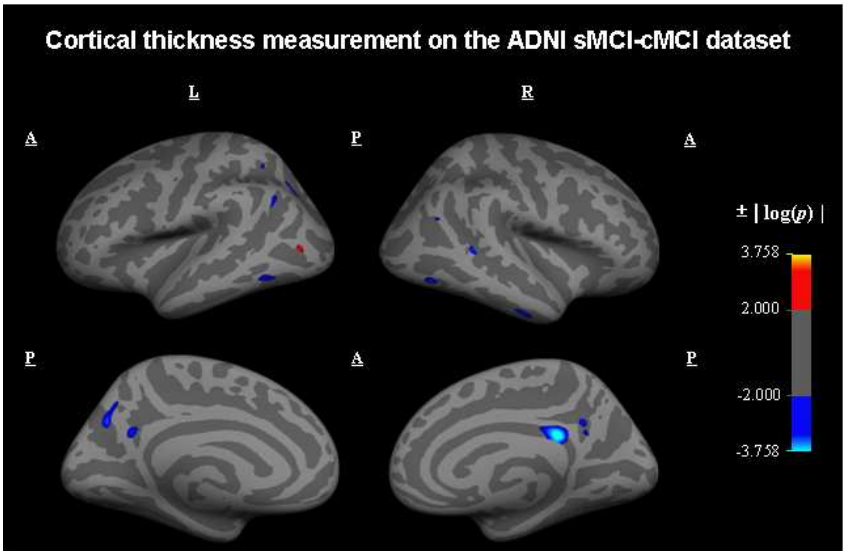


Figure 3.24: Cortical thickness measurement on the ADNI sMCI-cMCI dataset. A=Anterior; P=Posterior; L=Left; R=Right. The magnitude represents $|log(p)|$, the positive sign denotes significant expansion in the converting MCI group and the negative sign denotes significant shrinkage in the converting MCI group. $|log(p)| > 2 \Leftrightarrow p < 0.01$, false discovery rate is corrected.

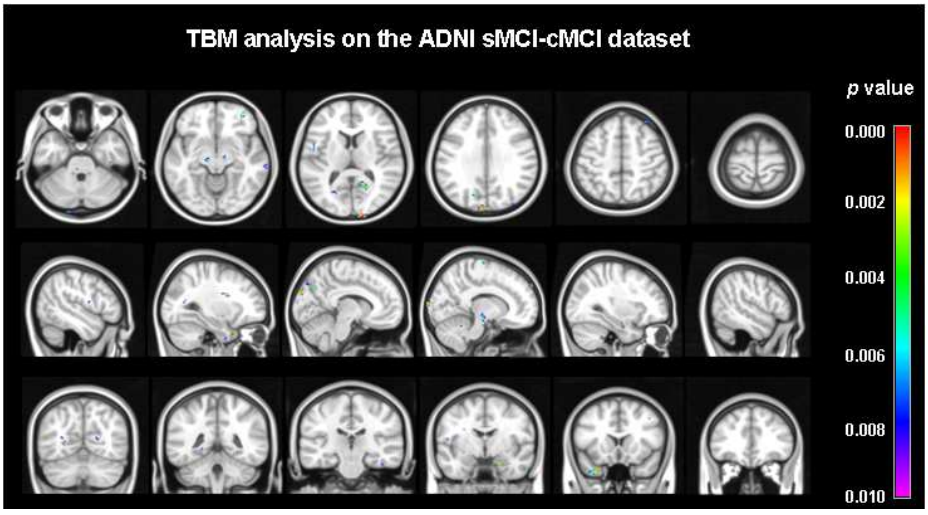


Figure 3.25: TBM analysis on the ADNI sMCI-cMCI dataset. $p < 0.01$, false discovery rate is corrected.

3.5 Discussion

From the above experiments and comparison, our classification-based pathology localization results are observed to be consistent with cortical thickness measurement, TBM and VBM analysis. Significant changes in hippocampus and temporal lobe, especially in parahippocampal/entorhinal and medial temporal cortex, occur at almost all AD stages. Frontal, parietal lobe, cingulate, and some other subcortical structures, such as amygdala and thalamus, are also observed closely related to AD pathology.

In longitudinal aspects, while frontal, central, parietal thinning, and large hippocampal deformations are found in pre-MCI subjects, there are no much changes detected in temporal or cingulate cortex. However, alterations in these two areas are observed between normal subjects and confirmed MCI patients, who later had progressed to AD. As the disease worsens, the temporal, parietal, frontal, as well as cingulate regions degenerate and the insula shrinkage appears.

Compared with cortical thickness measurement, TBM and VBM, our classification-based method can detect subtle differences in MCI/early-AD or even pre-MCI cases, which may be missed in other existing approaches, over the whole brain including cortex and subcortical regions. Although our method is an ROI-based method, the hierarchical search scheme can help obtain pathology distribution on the whole brain regardless of whether *a-priori* knowledge is known.

Chapter 4

Conclusion and future work

4.1 Conclusion

The goal of this dissertation was to develop an algorithm which could be applied for both disease diagnosis and prognosis, that improved upon previous work in terms of sensitivity, specificity and accuracy. A secondary goal was to use this method for localizing pathology.

In the developed algorithm, the symmetric log-domain diffeomorphic demons algorithm was first used to compute the pair-wise registration between whole-brain images. All the brain volumes were automatically segmented into gray matter, white matter, and CSF. The whole-brain diffeomorphism or regional diffeomorphisms, which are defined as the overlapped region between the whole-brain diffeomorphism and segmentations, were quantified as Riemannian distances. The matrix constructed by distance between each pair of subjects was embedded onto a low-dimensional space. In the embedded space, labeled images were used to train the discriminant hyperplane separating the underlying groups. Then unlabeled subjects were projected onto the constructed embedded space using out-of-sample embedding algorithm and assigned to appropriate groups based on the computed hyperplane. Correct classification rate, sensitivity, and specificity were calculated to evaluate the performance of classification with different regional distances.

In our classification experiments, the algorithm showed improved or equivalent performance to existing methods on both diagnosis and prediction of AD and MCI. It was also shown that classification using the entire white matter or gray matter region yielded higher discriminative accuracy than that using other ROIs to distinguish between underlying groups, since it

captures the relationships of morphological features across the whole brain. This result is consistent with observations in other research [13, 51]. Although large deformations on the temporal lobe were detected in the patients with medium-late AD, the whole gray matter deformation was found in mild AD or MCI subjects. In addition, if we don't have the *a-priori* knowledge on the disease, the gray matter region may be the better choice for disease detection and there is no problem of multiple comparisons in our algorithm.

Based on the classification, the application of the proposed algorithm is extended to explore the distribution of pathology. Comparing the classification results, those regions of interest (ROI) yielding higher classification accuracy are regarded as pathological regions. Although the original ROI approach has the problem that the selection of regions depends on the researcher's *a-priori* knowledge or hypothesis and leaves large areas of the brain unexplored, we used a hierarchical searching strategy which avoids the above two issues, with no user bias or *a-priori* hypothesis. In our experiments, we observed significant changes caused by gray matter shrinkage in thalamus, putamen, amygdala, hippocampus, parahippocampal/entorhinal cortex, medial/superior temporal lobe, precuneus/isthmus region, middle frontal lobe, and insula in diseased brains at different stages. The results were also observed to be consistent with those of other popular computational anatomic methods, such as cortical thickness measurement, VBM, and TBM.

Our algorithm is fully automated, and it reduces experimenters' bias, generates reproducible results, and makes the method less labor-intensive. Moreover, the high accuracy on distinguishing stable MCI and converting MCI shows its great prediction power which is important in future clinical practice. It is known that most studies of drugs for the treatment of AD have focused on early (mild) AD in the belief that disease progression will limit efficacy of the drugs [121]. However, it may be that pre-AD prevention is more efficacious than early treatment of the disease; the loss of neurons that occurs early in the disease may limit the action of some types of drugs. Thus, if the prediction of a conversion from MCI to AD can be made clinically, this disorder may actually be preventable, or at least its onset may be significantly delayed. Furthermore, this algorithm can be used for assessing treatments by longitudinal analysis, presenting an opportunity to track therapeutic response.

4.2 Constributions

Our contributions in this work are summarized as follows.

First, we compute the pair-wise registration and pair-wise distance between each pair of images to obtain more informative embedding. Since pair-wise registration captures more precise similarity or dissimilarity between images, the pair-wise information used for embedding minimizes the error between the true locations of images and the locations we calculated, and consequently increases the classification accuracy.

Second, we calculate a Riemannian distance between each pair of images. Riemmanian distance calculates the shortest length between two points in a manifold and it quantifies the transformation between images. This distance represents the shape displacement more precisely than using voxel-wise comparisons in a template space.

4.3 Future work

4.3.1 Classification with an advanced registration method

The goal of image registration is to align one image to match another. Ideally, medical images of the same anatomical structure using the same modality should present similar structural characteristics, such as homogeneity in local regions or the same topology between individuals. However, due to normal variation in anatomy, image artifacts and the presence of pathology (e.g. white matter lesions and meningioma), topological changes occur often in medical images, which complicates inter-subject registration. In such a case, the deformation field may not represent the true shape displacement between images. This issue may affect those deformation-based algorithms, including our method, deformation-based morphometry and tensor-based morphometry. In AD-related diagnosis or pathology localization analysis, topological changes such as white matter lesions might not be associated with the disease. For example, they might appear in both healthy brains and diseased brains (as shown in Fig. 4.1 (i-a), (i-d), (ii-b) and (ii-c)). Using registration algorithms which could not handle topology change problem may produce undesired displacement vectors in those regions. An example is shown in Fig. 4.2. White matter lesions appear in the source image (Fig. 4.2 (i-b)) around ventricles. However, no lesions are found in the target image (Fig. 4.2 (i-a)). Fig. 4.2 (i-c) and (ii-c) are respectively the aligned source image and the corresponding deformation field using the diffeomorphic demons algorithm. In the lesion area, the differences between the target and source images are from the intensity rather than the shape. Nevertheless, the resulting deformation field (Fig. 4.2 (ii-c)) presents a large number of vectors around the lesion area.

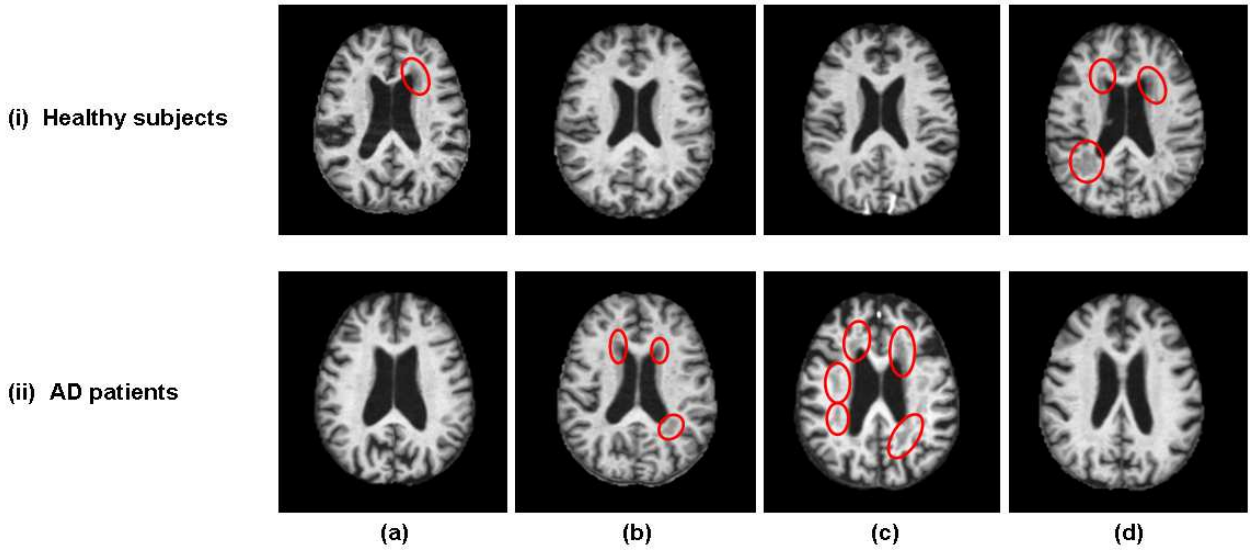


Figure 4.1: Source images. (i) Healthy subjects; (ii) AD patients.

Recently, Li [102] proposed a novel algorithm that embeds $3D$ images from an $\mathbb{R}^3$ Euclidean space to an $\mathbb{R}^4$ Riemannian space and treat the registration procedure from volumetric to surface-based. In the $\mathbb{R}^4$ space, the first 3 dimensions of the final deformation field correspond to the spatial displacement in the $\mathbb{R}^3$ Euclidean space, and the 4th dimension corresponds to the intensity displacement. Therefore, undesired deformation is effectively suppressed (as shown in Fig. 4.2 (i-d) and (ii-d)).

In our application, this new registration method may improve the distance calculation which quantifies the deformation field obtained from registration, and consequently may improve the classification and pathology localization result.

4.3.2 Classification with the Nystrom method

In Chapter 2.4.2 we explained why pair-wise registration was used in the algorithm. With pair-wise registration, we obtain good results, some of which are even much better than those recently published. However, a disadvantage of our algorithm is that computing pair-wise registration is computationally expensive if the size of dataset is large. Although the complexity is trivially reduced by parallelization at the level of individual registration, the Nystrom method would be a good approach to reduce computational complexity.

The idea of the Nystrom method is that rather than compute registration between all possible

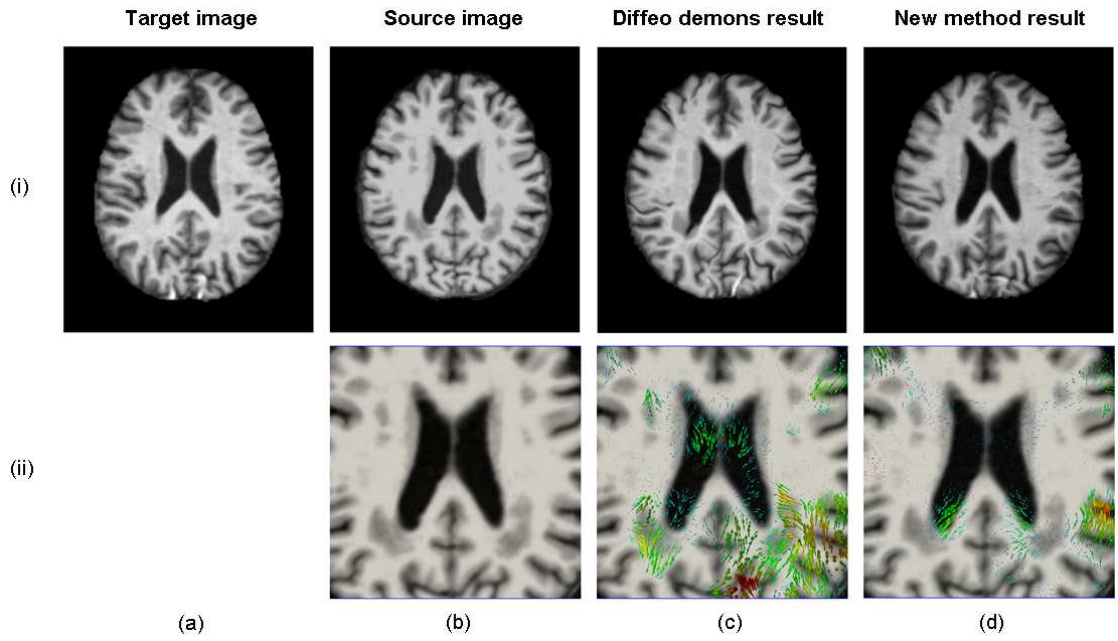


Figure 4.2: An example showing that how intensity differences instead of shape differences affect registration and its resulting deformation field. (i-a) is the target image; (i-b) and (ii-b) are the source image and the zoomed-in view around the lesion region; (i-c) and (ii-c) are the registered source image and the zoomed-in deformation field around the lesion region, using diffeomorphic demons registration; (i-d) and (ii-d) are the registered source image and the zoomed-in deformation field around the lesion region, using a new method proposed by Li.

pairs of images, we align these images to a small subset of randomly chosen data. The approach is simple and has the appealing characteristic that for a given number of sampled images, its complexity scales linearly with the number of subjects in the dataset. This produces two distance matrices, one that holds the sampled images' pair-wise distances ($\mathbf{A}$), and another ($\mathbf{B}$) that holds distances of the rest of images to the sampled images. The layout of the distance matrix is illustrated in Fig. 4.3.

The embedding vector of each image is obtained based on the matrices $\mathbf{A}$ and $\mathbf{B}$. The matrix $\mathbf{C}$ does not need to be calculated. Now the computational complexity reduces from $O(N \times N)$ to $O(N \times n)$, where N denotes the size of dataset and n denotes the number of sampled subjects.

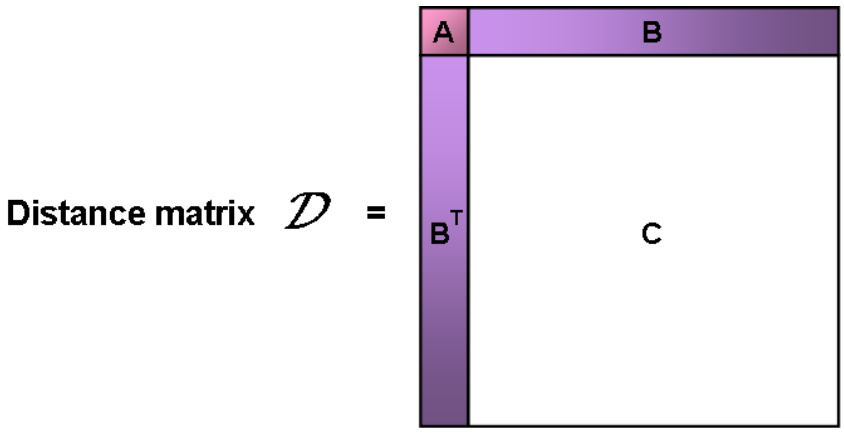


Figure 4.3: The layout of the distance matrix using Nystrom method. The regions within **A** and **B** represent individual subjects; **C** is the part of the matrix that does not need to be calculated.

Appendix A

Full segmentation of cortex

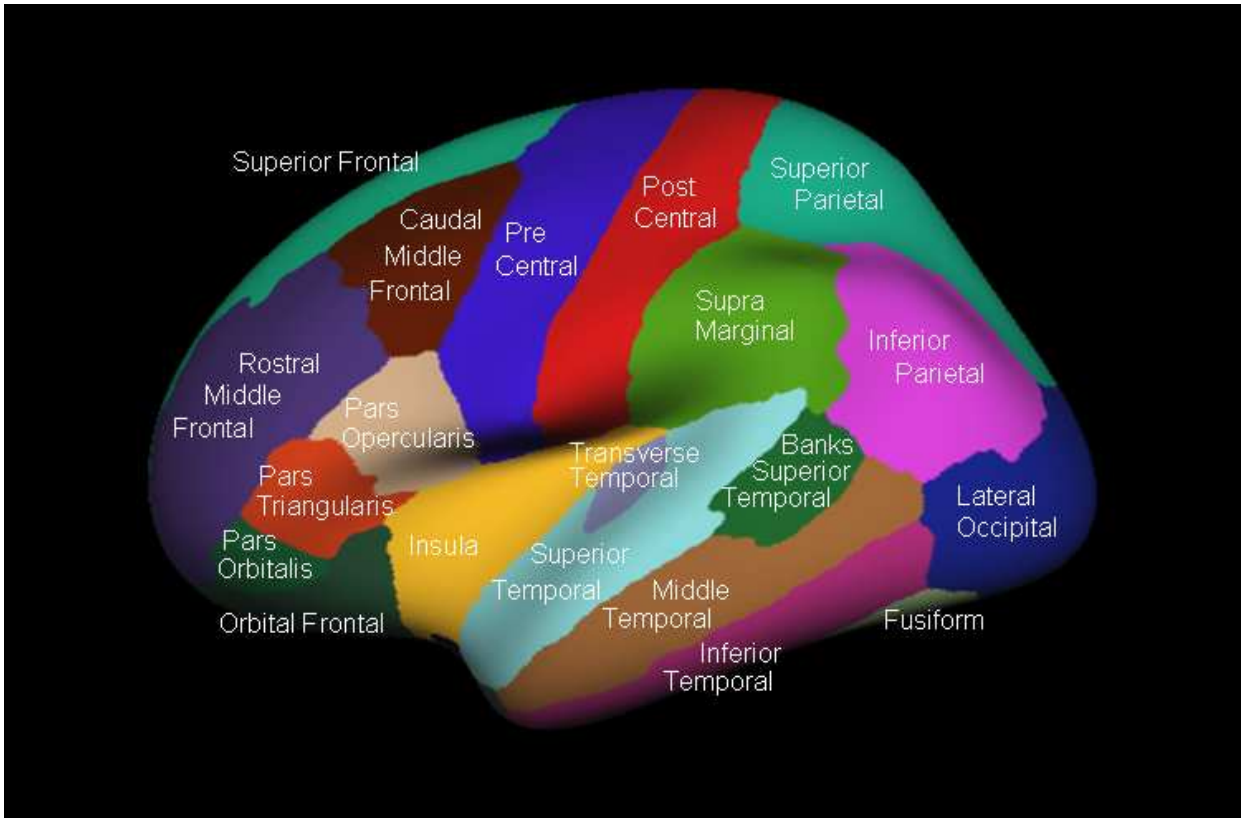


Figure A.1: The inlated cortical representations of the regions of interest at the lateral view of left hemisphere.

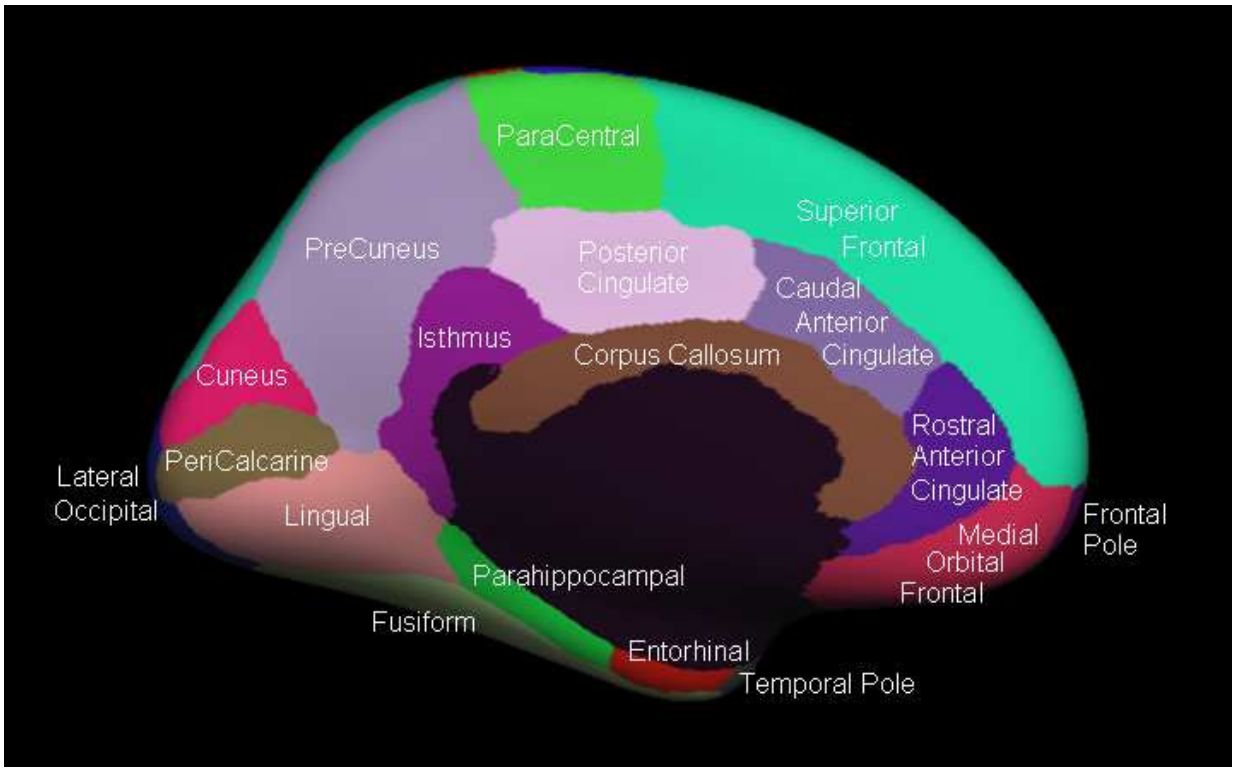


Figure A.2: The inlated cortical representations of the regions of interest at the medial view of left hemisphere.

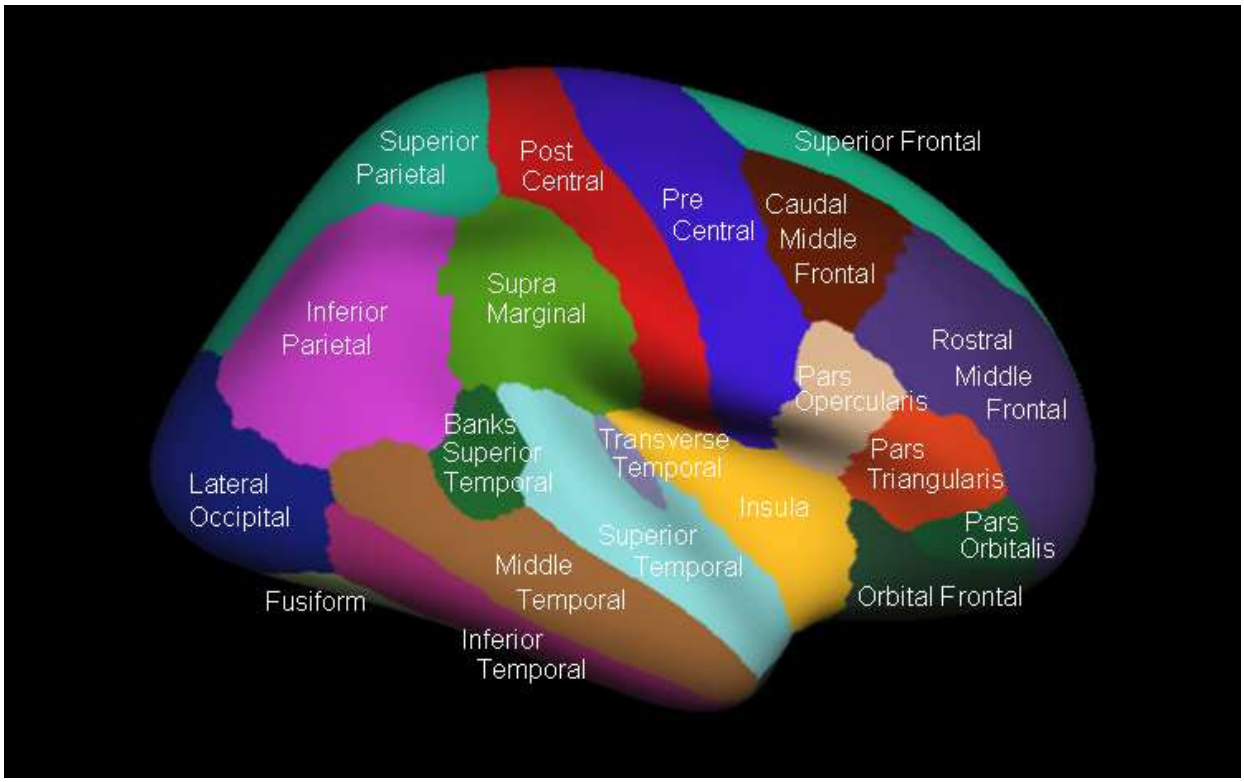


Figure A.3: The inlated cortical representations of the regions of interest at the lateral view of right hemisphere.

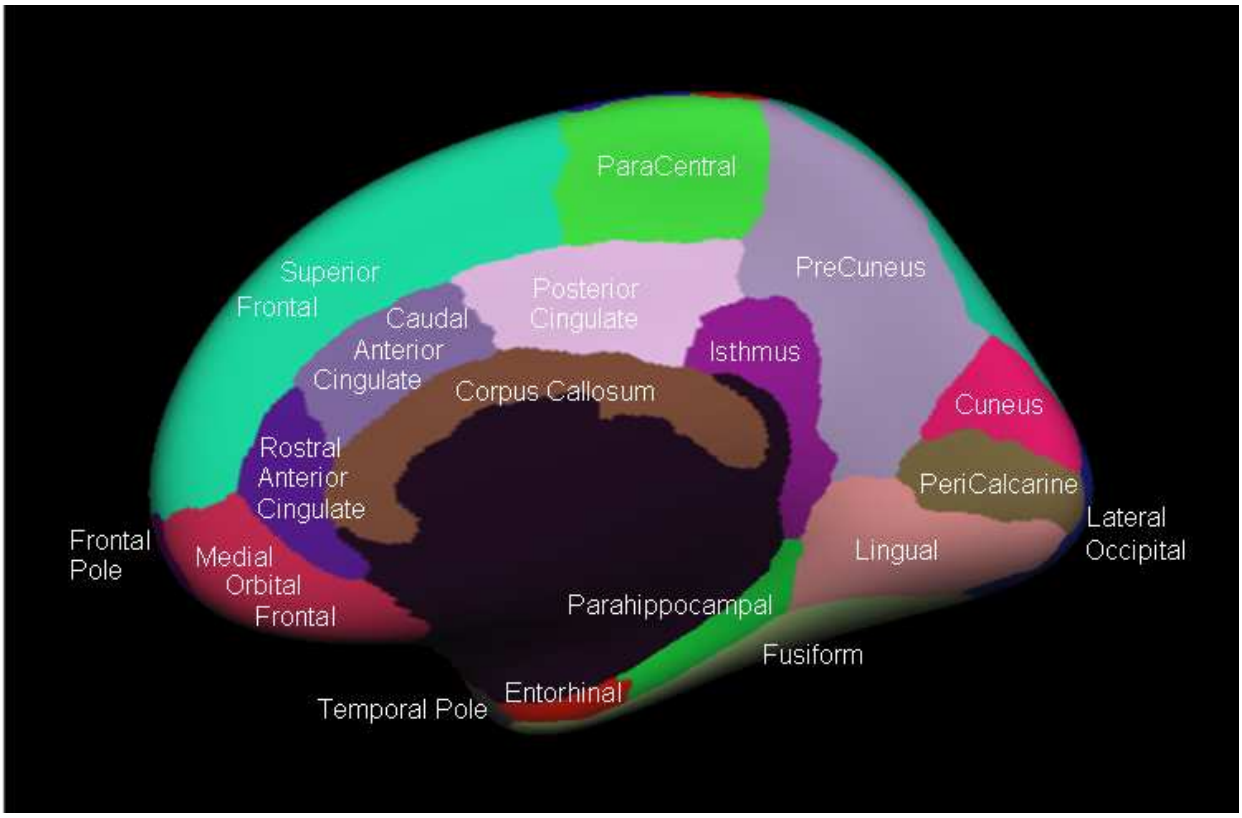


Figure A.4: The inlated cortical representations of the regions of interest at the medial view of right hemisphere.

Bibliography

[1] Thomas, M. and Isaac M., " Alois Alzheimer: a memoir," *Trends Neurosci* **10**, 306-307 (1987).

[2] McKhann, G. and Drachman, D. and Folstein, M. and Katzman, R. and Price, D. and Stadlan, E.M., "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology* **34**, 939-944 (1984).

[3] Braak, H. and Braak, E., "Neuropathological staging of alzheimer-related changes," *Acta Neuropathol* **82**, 239–259 (1991).

[4] Tiraboschi, P., Hansen, L., Thal, L., and Corey-Bloom, J., "The importance of neuritic plaques and tangles to the development and evolution of ad," *Neurology* **62 (11)**, 1984–1989 (2004).

[5] http://alzheimers.about.com/od/diagnosisofalzheimers/tp/neuropsychtests.htm

[6] Kramer, J.H. and Miller, B.L., "Alzheimer's disease and its focal variants," *Semin Neurol* **20**, 447-454 (2000).

[7] Norfray, J.F. and Provenzale, J.M., "Alzheimer's Disease: Neuropathologic Findings and Recent Advances in Imaging," *American Journal of Roentgenology* **182**, 3-13 (2004).

[8] Valcour, V.G., Masaki, K.H., Curb, J.D., Blanchette, P.L., "The detection of dementia in the primary care setting," *Arch Intern Med* **160**, 2964-2968 (2000).

[9] Solomon, P.R., Murphy, C.A., "Should we screen for Alzheimer's disease? A review of the evidence for and against screening Alzheimer's disease in primary care practice," *Geriatrics* **60**, 26-31 (2005).

[10] Fox, N. and Schott, J., "Imaging cerebral atrophy: normal ageing to alzheimer's disease," *Lancet* **363**, 392–394 (2004).

[11] Akgul, C.B., Unay, D., Ekin, A., "Automated Diagnosis of Alzheimer's Disease using Image Similarity and User Feedback," *Proc. of the ACM Intern Conf on Image and Video Retrieval*, doi:10.1145/1646396.1646438 (2009).

[12] Trosset, M., Priebe, C., Park, Y., and Miller, M., "Semisupervised learning from dissimilarity data," *Technical Report, Department of Statistics, Indiana University, Bloomington,* **IN4705** (2007).

[13] Kloppel, S., Stonnington, C., and et al., "Automatic classification of mr scans in alzheimer's disease," *Brain* **131**, 681–689 (2008).

[14] Fan, Y., Shen, D., Gur, R.C., and Davatzikos, A,C., "COMPARE: classification of morphological patterns using adaptive regional elements," *IEEE Trans. Med. Imaging* **26 (1)**, 93-105 (2007).

[15] Thompson, P., Hayashi, K., Sowell, E., Gogtay, N., Giedd, J., Rapoport, J., and et al., "Mapping cortical change in alzheimer's disease, brain development, and schizophrenia," *Neuroimage* **23 suppl 1**, S3–S18 (2004).

[16] Thompson, P. and Apostolova, L., "Computational anatomical methods as applied to ageing and dementia," *Br. J. Radiol* **80(Spec No 2)**, S78–91 (2007).

[17] Querbes, O., Aubry, F., Pariente, J., Lotterie, J., and et al., "Early diagnosis of alzheimer's disease using cortical thickness: impact of cognitive reserve," *Brain* **132(8)**, 2036–2047 (2009).

[18] Lerch, J. P., Pruessner, J. C., Zijdenbos, A., Hampe, H., Teipe, S. J., and Evans, A. C., "Focal decline of cortical thickness in alzheimer's disease identified by computational neuroanatomy," *Cerebral Cortex* **15**, 995–1001 (2005).

[19] Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., and Friston, K., "Identifying global anatomical differences: Deformation-based morphometry," *Human Brain Mapping* **6**, 348–357 (1998).

[20] Chung, M., Worsley, K., Paus, T., Cherif, C., Collins, D., Giedd, J., Rapoport, J., and Evans, A., "A unified statistical approach to deformation-based morphometry," *NeuroImage* **14**, 595–606 (2001).

[21] Lepore, N., Brun, C., Chou, Y., Chiang, N., Dutton, R., Hayashi, K., and et al., "Generalized tensor-based morphometry of hiv/aids using multivariate statistics on deformation tensors," *IEEE Trans Med Imaging* **27(1)**, 129–141 (2008).

[22] Ashburner, J. and Friston, K., "Voxel-based morphometry - the methods," *NeuroImage* **11**, 805–821 (2000).

[23] Baronb, J. C., Chtelat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., and Eustache, F., "In vivo mapping of gray matter loss with voxel-based morphometry in mild alzheimer's disease," *NeuroImage* **14(2)**, 298–309 (2001).

[24] Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G. and Kokmen, E., "Mild cognitive impairment: clinical characterization and outcome," *Arch Neurol* **56**, 303-308(1999).

[25] Blennow, K., de Leon, M., and Zetterberg, H., "Alzheimer's disease," *Lancet* **368**, 387–403 (2006).

[26] Harvey, P., Moriarty, P., Kleinman, L., and et al, "The validation of a caregiver assessment of dementia: the dementia severity scale," *Alzheimer Dis Assoc Disord* **19**, 186–194 (2005).

[27] Jr., C. J., Petersen, R., Xu, Y., O'Brien, P., Smith, G., Ivnik, R., Boeve, B., Waring, S., Tangalos, E., and Kokmen, E., "Prediction of ad with mri-based hippocampal volume in mild cognitive impairment," *Neurology* **52**, 1397–1403 (1999).

[28] Barnes, J., Scahill, R., R.G.Boyes, Frost, C., Lewis, E., Rossor, C., Rossor, M., and Fox, N., "Differentiating ad from aging using semiautomated measurement of hippocampal atrophy rates," *NeuroImage* **23**, 574–581 (2004).

[29] Hirata, Y., Matsuda, H., Nemoto, K., Ohnishi, T., Hirao, K., Yamashita, F., Asada, T., Iwabuchi, S., and Samejima, H., "Voxel-based morphometry to discriminate early alzheimer's disease from controls," *Neurosci Lett* **382**, 269–274 (2005).

[30] Bozzali, M., Filippi, M., Magnani, G., Cercignani, M., Franceschi, M., Schiatti, E., and et al., "The contribution of voxel-based morphometry in staging patients with mild cognitive impairment," *Neurology* **67**, 453–460 (2006).

[31] Gutman, B., Wang, Y., Morra,J., Toga,A., and Thompson, P., "Disease classification with hippocampal shape invariants," *Hippocampus* **19**, 572–578 (2009).

[32] Gerardin, E., Chupin, M., Cuingnet, R., Dubois, B., Lehricy, S., Garnero, L., Colliot, O., and ADNI, "Svm classification of patients with alzheimer's disease and mild cognitive impairment using hippocampal shape features," *NeuroImage* **47(suppl 1)**, 39–41 (2009).

[33] Karas, G., Scheltens, P., Rombouts, S., Visser, P., van Schijndel, R., Fox, N., and et al., "Global and local gray matter loss in mild cognitive impairment and alzheimer's disease," *Neuroimage* **23**, 708–716 (2004).

[34] Dale, A., Fischl, B., and Sereno, M., "Cortical surface-based analysis i: Segmentation and surface reconstruction," *NeuroImage* **9(2)**, 179–194 (1999).

[35] Joshi, S., Davis, B., Jomier, M., and Gerig, G., "Unbiased diffeomorphic atlas construction for computational anatomy," *Neuroimage* **23(suppl 1)**, 151–160 (2004).

[36] Vaillant, M., Miller, M., Younes, L., and Trouve, A., "Statistics on diffeomorphisms via tangent space representation," *Neuroimage* **23(suppl 1)**, 161–169 (2004).

[37] Vercauteren, T. and Pennec, X. and Perchant, A. and Ayache, N., "Non-parametric diffeomorphic image registration with the demons algorithm," *MICCAI'07 LNCS* **4792**, 319-326 (2007).

[38] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N., "Symmetric log-domain diffeomorphic registration: A demons-based approach," *MICCAI* **Part I, LNCS 5241**, 754–761 (2008).

[39] Fletcher, P. T., Lu, C., and Joshi, S., "Statistics of shape via principal component analysis on lie group," *In Proceedings of CVPR* (2003).

[40] Belkin, M. and Niyogi, P., "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation* **15(6)**, 1373–1396 (2003).

[41] Tenenbaum, J., de Silva, V., and Langford, J., "A global geometric framework for nonlinear dimensionality reduction," *Science* **290(5500)**, 2319–2323 (2000).

[42] Shlens, J., "A tutorial on principal component analysis," Institute for Nonlinear Science, UCSD (2005).

[43] Cox, T. F. and Cox, M. A. A., "Multidimensional scaling," Chapman and Hall (2001).

[44] Bengio, Y., Paiement, J. F., Vincent, P., and Delalleau, O., "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering," *Advances in Neural Information Processing Systems* **16** (2004).

[45] Shi, J. and Malik, J., "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22(8)**, 888–905 (2000).

[46] Ng, A., Jordan, M., and Weiss, Y., "On spectral clustering: Analysis and an algorithm," *Advances in Neural Information Processing Systems* **12** (2001).

[47] Abdi, H., "Metric multidimensional scaling (mds): analyzing distance matrices," *Encyclopedia of Measurement and Statistics (Salkind N, ed)* , 1–13 (2007).

[48] Bengio, Y., Vincent, P., Paiement, J. F., Delalleau, O., Ouimet, M., and Roux, N. L., "Learning eigenfunctions links spectral embedding and kernel pca," *Neural Comput* **16**, 2197–2219 (2004).

[49] Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., and Buckner, R., "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience* **19**, 1498–1507 (2007).

[50] Morris, J., "The clinical dementia rating(cdr): Current version and scoring rules," *Neurology* **43**, 2412–2414 (1993).

[51] Fan, Y., Batmanghelich, N., Clark, C., Davatzikos, C., and Initiative, A. D. N., "Spatial patterns of brain atrophy in mci patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *Neuroimage* **39(4)**, 1731–1743 (2008).

[52] Vereecken, T., Vogelsa, O., and Nieuwenhuys, R., "Neuron loss and shrinkage in the amygdala in alzheimer's disease," *Neurobiology of Aging* **15**, 45–54 (1994).

[53] Barber, R., McKeith, I., Ballard, C., and O'Brien, J., "Volumetric mri study of the caudate nucleus in patients with dementia with lewy bodies, alzheimer's disease, and vascular dementia," *J Neurol Neurosurg Psychiatry* **72**, 406–407 (2002).

[54] de Jong, L., van der Hiele, K., Veer, I., Houwing, J., Westendorp, R., Bollen, E., and et al., "Strongly reduced volumes of putamen and thalamus in alzheimer's disease: an mri study," *Brain* **131(12)**, 3277–3285 (2008).

[55] Fischl, B. and Dale, A., "Measuring the thickness of the human cerebral cortex from magnetic resonance images," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11050–11055 (2000).

[56] Chiang, M., Dutton, R., Hayashi, K., Toga, A., Lopez, O., Aizenstein, H., Becker, J., and Thompson, P., "3d pattern of brain atrophy in hiv/aids mapping using tensor-based morphometry," *Neuroimage* **34**, 44–60 (2007).

[57] Hua, X., Leow, A., Lee, S., Klunder, A., Toga, A., Lepore, N., and et al., "3d characterization of brain atrophy in alzheimer's disease and mild cognitive impairment using tensor-based morphometry," *NeuroImage* **41**, 19–34 (2008).

[58] Leow, A., Klunder, A., Jack, C. J., and et al., "Longitudinal stability of mri for mapping brain change using tensor-based morphometry," *Neuroimage* **31**, 627–640 (2006).

[59] Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., and et al., "Advances in functional and structural mr image analysis and implementation as fsl," *NeuroImage* **23(S1)**, 208–219 (2004).

[60] Smith, S., "Fast robust automated brain extraction," *Human Brain Mapping* **17(3)**, 143–155 (2002).

[61] Zhang, Y., Brady, M., and Smith, S., "Segmentation of brain mr images through a hidden markov random field model and the expectation maximization algorithm," *IEEE Trans. on Medical Imaging* **20(1)**, 45–57 (2001).

[62] Jenkinson, M. and Smith, S., "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis* **5(2)**, 143–156 (2001).

[63] Jenkinson, M., Bannister, P., Brady, J., and Smith, S., "Improved optimisation for the robust and accurate linear registration and motion correction of brain images," *NeuroImage* **17(2)**, 825–841 (2002).

[64] Andersson, J., Jenkinson, M., and Smith, S., "Non-linear optimisation," *FMRIB technical report TR07JA1 from www.fmrib.ox.ac.uk/analysis/techrep* .

[65] Andersson, J., Jenkinson, M., and Smith, S., "Non-linear registration, aka spatial normalisation," *FMRIB technical report TR07JA2 from www.fmrib.ox.ac.uk/analysis/techrep* .

[66] Perlman, D., "Cortical Thickness: Practicalities and Comparisons," Statistics 692 project, University of Wisconsin (2007).

[67] Roerdink, J.B.T.M. and Meijster, A., "The watershed transform: definitions, algorithms and parallelization strategies," *Fundamenta Informaticae* **41**, 187-228 (2001).

[68] Sgonne, F., Dale, A.M., Busa, E., Glessner, M., Salat, D., Hahn, H.K.and Fischl, B.A," Hybrid approach to the skull stripping problem in MRI," *NeuroImage* **22**, 1060-1075 (2004).

[69] Bakos, M., "Active contours and their utilization at image segmentation," *Proceedings from SAMI 2007: 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics*, 313-317 (2007).

[70] Hagler, D.J., Sr. and Saygin, A.P. and Sereno, M.I., "Smoothing and cluster thresholding for cortical-surface based group analysis of fMRI data," *NeuroImage* **33**, 1093-1103 (2006).

[71] Chung, M. and Robbins, S.M. and Dalton, K.M. and Davidson, R.J. and Alexander, A.L. and Evans, A.C., "Cortical thickness analysis in autism with heat kernel smoothing," *NeuroImage* **25**, 1256-1265 (2005).

[72] Nichols, T.E. and Holmes, A.P., "Nonparametric permutation tests for functional neuroimaging: A primer with examples," *Human brain mapping* **15**, 1-25 (2001).

[73] Benjamini, Y. and Hochberg, Y., "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B (Methodological)* **57(1)**, 289-300 (1995).

[74] Gaser, C. and Volz, H. and Kiebel, S. and Riehemann, S. and Sauer H., "Detecting structural changes in whole brain based on nonlinear deformations - Application to Schizophrenia research," *NeuroImage* **10**, 107-113 (1999).

[75] "http://surfer.nmr.mgh.harvard.edu/fswiki/mri_glmfit".

[76] Shaffer J.P., "Multiple Hypothesis Testing," *Annual Review of Psychology* **46**, 561-584 (1995).

[77] Cuignet, R., Gerardin, E., Tessieras, J., Auzias, G., and et al., "Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database," *NeuroImage* **doi:10.1016/j.neuroimage.2010.06.013** (2010).

[78] Cunningham, P., "Dimension reduction," *Technical report* **UCD-CSI-2007-7** (2007).

[79] Ashburner, J., Friston, K.J., "Unified segmentation," *NeuroImage* **26 (3)**, 839-851 (2005).

[80] Ashburner, J., "A fast diffeomorphic image registration algorithm," *NeuroImage* **38 (1)**, 95-113 (2007).

[81] Vemuri, P., Gunter,J.L., Senjem,M.L., Whitwell,J.K., Kantarci,K., Knopman,D.S. and et al., "Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies," *NeuroImage* **39 (3)**, 1186-1197 (2008).

[82] Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S. M. and Davatzikos, C., "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *Neuroimage* **21 (1)**, 46-57 (2004).

[83] Magnin, B., Mesrob, L., Kinkingnehun, S., Pelegrini-Issac, M., Colliot, O., Sarazin, M., Dubois, B., Lehericy, S. and Benali, H., "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," *Neuroradiology* **51 (2)**, 73-83 (2009).

[84] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M., "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage* **15**, 273-289 (2002).

[85] Fan, Y., Resnick, S. M., Wu, X. and Davatzikos, C., "Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study," *Neuroimage* **41 (2)**, 277-285 (2008).

[86] Thompson, P. M., Mega, M. S., Woods, R. P., Zoumalan, C. I. and et al., "Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas," *Cereb. Cortex* **11 (1)**, 1-16 (2001).

[87] Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L. and et al., "Dynamics of gray matter loss in Alzheimer's disease," *J. Neurosci.* **23 (3)**, 994-1005 (2003).

[88] Lerch, J. P., Pruessner, J., Zijdenbos, A. P., Collins, D. L., Teipel, S. J., Hampel, H., and Evans, A. C., "Automated cortical thickness measurements from MRI can accurately

separate Alzheimer's patients from normal elderly controls," *Neurobiol. Aging* **29 (1)**, 23-30 (2008).

[89] Bakkour, A., Morris, J. C., and Dickerson, B. C., "The cortical signature of prodromal AD: regional thinning predicts mild AD dementia," *Neurology* **72 (12)**, 1048-1055 (2009).

[90] Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J. and et al., "The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals," *Cereb. Cortex* **19 (3)**, 497-510 (2009).

[91] Hua, X., Lee, S., Yanovsky, I., Leow, A. D., Chou, Y. Y. and et al. and Alzheimer's Disease Neuroimaging Initiative, "Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects," *Neuroimage* **48 (4)**, 668-681 (2009).

[92] McDonald, C. R., McEvoy, L. K., Gharapetian, L., Fennema-Notestine, C., Hagler Jr., D. J., Holland, D., Koyama, A., Brewer, J. B., Dale, A. M. and Alzheimer's Disease Neuroimaging Initiative, "Regional rates of neocortical atrophy from normal aging to early Alzheimer disease," *Neurology* **73 (6)**, 457-465 (2009).

[93] Sled, J. G., Zijdenbos, A. P. and Evans, A. C., "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans. Med. Imaging* **17**, 87-97 (1998).

[94] Fischl, B., Sereno, M. I. and Dale, A. M., "Cortical surface-based analysis ii: Inflation, flattening, and a surface based coordinate system," *NeuroImage* **9(2)**, 195-207 (1999).

[95] Fischl, B., Sereno, M. I., Tootell, R. B. and Dale, A. M., "High-resolution intersubject averaging and a coordinate system for the cortical surface," *Hum. Brain Mapp.* **8 (4)**, 272-284 (1999).

[96] Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., Schmansky, N. J., Greve, D. N., Salat, D. H., Buckner, R. L., Fischl, B. and Alzheimer's Disease Neuroimaging Initiative, "Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease," *Brain* **132 (8)**, 2048-2057 (2009).

[117] Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C. and et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage* **31 (3)**, 968-980 (2006).

[98] Chupin, M., Mukuna-Bantumbakulu, A. R., Hasboun, D., Bardinet, E., Baillet, S. and et al., "Automated segmentation of the hippocampus and the amygdala driven by competition and anatomical priors: Method and validation on healthy subjects and patients with Alzheimer's disease," *Neuroimage* **34**, 996-1019 (2007).

[99] Chupin, M., Hammers, A., Liu, R. S., Colliot, O., Burdett, J., Bardinet, E., Duncan, J. S., Garnero, L. and Lemieux, L., "Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints: method and validation," *Neuroimage* **46 (3)**, 749-761 (2009).

[100] Chupin, M., Gerardin, E., Cuingnet, R., Boutet, C., et al. and Alzheimer's Disease Neuroimaging Initiative, "Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus* **19 (6)**, 579-587 (2009).

[101] Colliot, O., Chetelat, G., Chupin, M., Desgranges, B. and et al., "Discrimination between Alzheimer disease, mild cognitive impairment, and normal aging by using automated segmentation of the hippocampus," *Radiology* **248 (1)**, 194-201 (2008).

[102] Li, X., "Registration of images with varying topology using embedding maps," Unpublished Ph.D. dissertation, Virginia Tech, Blacksburg, VA, (2010).

[103] Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M., "Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging," *Neurobiol Aging* **29**, 514-523 (2008).

[104] Mahony, R., Manton, J.H., "The geometry of the Newton method on non-compact Lie-groups," *J. Global Optim.* **23(3)**, 309-327 (2002).

[105] Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P. and et al., "The Alzheimer's disease Neuroimaging Initiative (ADNI):MRI methods," *J. Magn. Reson. Imaging* **27(4)**, 685-691 (2008).

[106] Gunter, J.L., Bernstein, M.A., Borowski, B.J., Ward, C.P. and et al., "Measurement of MRI scanner performance with the ADNI phantom," *Med. Phys.* **36**, 2193-2205 (2009).

[107] Mortamet, B., Bernstein, M.A., Jack Jr., C.R., Gunter, J.L. and et al., "Automatic quality assessment in structural brain magnetic resonance imaging," *Magn. Reson. Med.* **62**, 365-372 (2009).

[108] Misra, C., Fan, Y. and Davatzikos, C., "Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI," *Neuroimage* **44 (4)**, 1415-1422 (2009).

[109] Thompson, P.M. and et al., "Abnormal cortical complexity and thickness profiles mapped in Williams syndrome," *J. Neuroscience* **25 (16)**, 4146-4152 (2005).

[110] Slotta, D.J., Vergara, J.P., Ramakrishnan, N. and Heath, L.S., "Algorithms for feature selection in rank-order spaces," Technical report TR-05-08, Computer Science, Virginia Tech, 2005.

[111] Brank, J., Grobelnik, M., Milic-Frayling, N. and Mladenic, D., "Feature selection using linear support vector machines," Technical report, Microsoft Research, 2002.

[112] Davatzikos, C., and Bryan, R.N., "Using a Deformable Surface Model to Obtain a Shape Representation of the Cortex," *IEEE Trans. Med. Imaging* **15**, 785-795 (1996).

[113] MacDonald, D., "A Method for Identifying Geometrically Simple Surfaces from Three Dimensional Images," Montreal Neurological Institute, McGill University, Montreal.

[114] Fischl, B., Salat, D.H. and et al., "Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain," *Neuron* **33**, 341-355 (2002).

[115] Dale, A.M., and Sereno, M.I., "Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction:Alinear approach," *J. Cogn. Neurosci.* **5(2)**, 162-176 (1993).

[116] Fischl, B., van der Kouwe, A. and et al., "Automatically Parcellating the Human Cerebral Cortex," *Cerebral Cortex* **14**, 11-22 (2004).

[117] Desikan, R.S., Segonne, F. and et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage* **31(3)**, 968-980 (2006).

[118] Lucignani, G., "The neuroimaging approach to the assessment of mild cognitive impairment," *Eur J Nucl Med Mol Imaging* **33**, 745-748 (2006).

[119] Furukawa, K., Okamura, N. and et al. "Amyloid PET in mild cognitive impairment and Alzheimer's disease with BF-227: comparison to FDG-PET," *Journal of Neurology* **257 (5)**, 721-727 (2009).

[120] Liua, Y., Paajanen, T. and et al., "Combination analysis of neuropsychological tests and structural MRI measures in differentiating AD, MCI and control groups - The AddNeuroMed study," *Neurobiology of Aging*, doi:10.1016/j.neurobiolaging.2009.07.008 (2009).

[121] Bullock, R., "Drug treatment for early Alzheimer's disease," *Advances in Psychiatric Treatment* **4**, 126-134 (1998).

[122] Rao, G., "Remembering the meanings of sensitivity, specificity, and predictive values. - Language of evidence: defining the terms of evidence-based medicine - Predictive value of tests Medical," University of Pittsburgh Medical Center-St. Margaret, Pittsburgh, PA (2004).

[123] Long, X. and Wyatt, C., "An automatic unsupervised classification of mr images in alzheimer's disease," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, doi:10.1109/CVPR.2010.5540031 (2010).