Automated Discovery and Visualization of Discriminating Structural Markers from MRI using Transport-Based Morphometry

Submitted in partial fulfillment of the requirements for

the degree of

Doctor of Philosophy

in

Biomedical Engineering

Shinjini Kundu

B.S., Electrical Engineering, Stanford University

M.S., Electrical Engineering, Stanford University

Carnegie Mellon University

Pittsburgh, PA

August 2016

Dedicated to the loving memory of the two people

who were always the proudest of me:

My grandfather, Pulak Ranjan Sinha and

My grandmother, Purnima Sinha

Acknowledgements

I owe my deepest gratitude to my advisor Prof. Gustavo K. Rohde for giving me the opportunity to pursue this thesis under his guidance. I am grateful for his support and mentorship over the past three years, including the many hours he spent in teaching me. His insightful advice and patience sustained me through many obstacles, and his enthusiasm motivated me to reach higher.

Next, I would like to express my gratitude to my dissertation committee members: Prof. Ken Urish, Prof. Kirk Erickson, and Prof. John Galeotti. I am thankful for their involvement and invaluable feedback throughout the dissertation process. Their knowledgeable input has helped improve this body of work.

I am indebted to my collaborators and paper co-authors: Soheil Kolouri, Marcel Just, Hoby Hetherington, T. Kevin Hitchens, Lesley Foley, Pratik Mukherjee, Julia Owen, David Creswell, Ken Urish, Kirk Erickson, Anish Ghodadra, Lea Alhilali, and Saeed Fakhran for their contributions, interesting new perspectives, and generosity in sharing datasets. I owe my thanks to several post-doctoral researchers who were resourceful mentors: Soheil Kolouri and Yue Huang, and my graduate student peers Serim Park, Chi Liu, Yang Zou, and Ligong Han.

I owe my gratitude to the Dowd family for their generous support of my research through the Dowd-ICES graduate fellowship during the 2014-2015 academic year.

I would like to thank my family for their unconditional love and support: my sister - Shohini Kundu, my mother - Deblina Kundu, and my father - Prof. Sandip Kundu. Their faith in me gives me the courage to pursue my highest ambitions. My grandparents, many aunts, uncles and cousins have sent their love and encouragement across continents and seas: Prof. Pulak Ranjan

Sinha, Purnima Sinha, Subhoranjan Sinha & Shilpi Sinha, Srija Sinha, Priyaranjan Sinha & Shipra Sinha, Dr. Smarajit Jana & Madhulina Jana, Samaita Jana & Sambit Jana, Tarun Kanti Pramanik & Syamalina Pramanik, Tamoghno Pramanik; Prof. Hari Mohan Kundu & Prabhati Kundu, Dr. Saikat Kundu & Debbani Kundu, Satyaki Kundu & Satwik Kundu. Last, but not least, I owe thanks to Dr. George Leef for reminding me to stay focused.

This thesis is based on work that will be submitted for scientific publication. This work was supported in part by the Department of Biomedical Engineering and grants from the National Science Foundation CCF 1421502, the National Institute of Health R01 GM090033.

Abstract

Imaging studies are a vital part of accurate medical care for every part of the body. Recent advances in magnetic resonance imaging (MRI) technology hold promise to uncover structural changes underlying diseases commonly considered medical mysteries. Unfortunately, the changes are often subtle, spatially distributed and complex, escaping detection by visual inspection. The key challenge is to disambiguate meaningful information from heterogeneous normal variations. The assistance of computer-aided techniques is sought to answer the following questions: are there significant morphologic differences that differentiate these patients? If so, what is the nature of these differences?

Traditional computer-aided techniques that extract a set of pre-specified descriptors (i.e. volume, thickness, etc.) or compare images pixelwise can test directed hypotheses about regional differences, but do not consider important spatially diffuse effects. Furthermore, current approaches for assessing statistical differences do not allow recovery or visualization of the images underlying the associations, hindering physical interpretation. The goal of this research is to develop a new framework for automated discovery and visualization of discriminating structural markers from MRI data. We utilize the mathematics of optimal mass transport (OT) to quantify tissue spatial distribution and transform data to enhance separability of images. Our approach for Transport-Based Morphometry (TBM) defines a fully invertible transformation, allowing statistically significant discriminant differences identified in the transform domain to be visualized as morphology changes through inverse transformation. This thesis extends the mathematics of TBM to enable its first application to 3D radiology data, and applies the new TBM framework for classification and

regression tasks in a variety of open clinical and research problems.

Our results based on knee cartilage images indicate that TBM enables detection of osteoarthritis with 86.2% accuracy three years in advance of symptoms. TBM also enables identification of 16p11.2 duplication, deletion, and control carriers with 96% accuracy based on structural appearance of the brain alone. Patterns of injury underlying post-concussive cognitive deficits are identified with the TBM technique. The new technique also finds that aerobic fitness significantly affects brain tissue distribution in older adults. These are significant improvements over prior detection methods. Moreover, unlike previous approaches, our method allows structural markers to be visualized.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview and Goals

On a frosty winter's day in Germany, Wilhelm Roentgen experienced a curious phenomenon. He witnessed a peculiar glow emanating from the barium platinocyanide screen in his lab. He paused for a moment, wondering whether it was an optical illusion. The glimmer, however, was unmistakable. In response to x-rays generated in his lab, he was observing fluorescence of the screen several feet away. Cautious about the implications of his discovery, Roentgen at first told no one what he saw. Then one day, six weeks later, he invited his wife into the laboratory and took a picture of the bones in her hand. This would mark the day that medicine would be forever changed. The ability to visualize its inner workings without needing even a single incision by a scalpel was a remarkable advance. News of the discovery spread rapidly across the globe and Roentgen was awarded a Nobel Prize for his discovery [10]. The rest, of course, is history.

Since then, new modalities in imaging have enabled us to see not only the bones, but the organs, soft tissues, blood vessels and cells that comprise the human body. Imaging studies today are a vital part of accurate medical diagnosis and treatment for every part of the body. Reduced scan times, increasing resolution, and proliferation of new modalities have provided us with a look inside the human body at an unprecedented level of precision and accuracy. Cardiologists

use echocardiography to assess the heart's function in real-time. Neurologists use diffusion tensor imaging (DTI) to assess brain damage acutely during active stroke. Orthopedic surgeons use computed tomography (CT) to plan their surgeries. Pharmacologists use positron emission tomography (PET) to monitor drug uptake in the body. Neuroscience researchers use functional magnetic resonance imaging (fMRI) to understand functional connectivity in the brains of autistic patients. As we learn more about the details of its inner workings, our enchantment with the human body and awe of its complexity have only amplified.

IBM Watson estimates that up to 90% of all medical data today is in the form of images [95]. As the proliferation in image modalities, volume of studies ordered, and increase in resolution continue, new studies are able to detect subtle, complex and spatially distributed changes. However, many of these changes escape human visual detection. In Figure 1.1a, we see MRIs of knee cartilage MRIs belonging to healthy subjects. Following these subjects over a period of three years, we know that only one of these groups will go on to develop clinical osteoarthirits with pain and bone degeneration in the knee. However, there appears to be no common feature identifiable by eye that differentiates these two groups. In Figure 1.1b, we see fractional anisotropy maps of patients who have suffered mild traumatic brain injury. Following the injury, many of the patients exhibited cognitive deficits, manifested in their reaction times. However, the differences between the two groups are not visually discernible.

Relevant questions include:

- Is there a characteristic difference in morphology that differentiates the two groups?
- If so, what is it and how can we visualize it?
- What features should we measure that are most informative for discriminating the groups?
- Which anatomic regions are most important for discrimination?
- How do the regions interact to produce the clinical symptoms?

- How can we best model and visualize the relationship between disease state and anatomy?

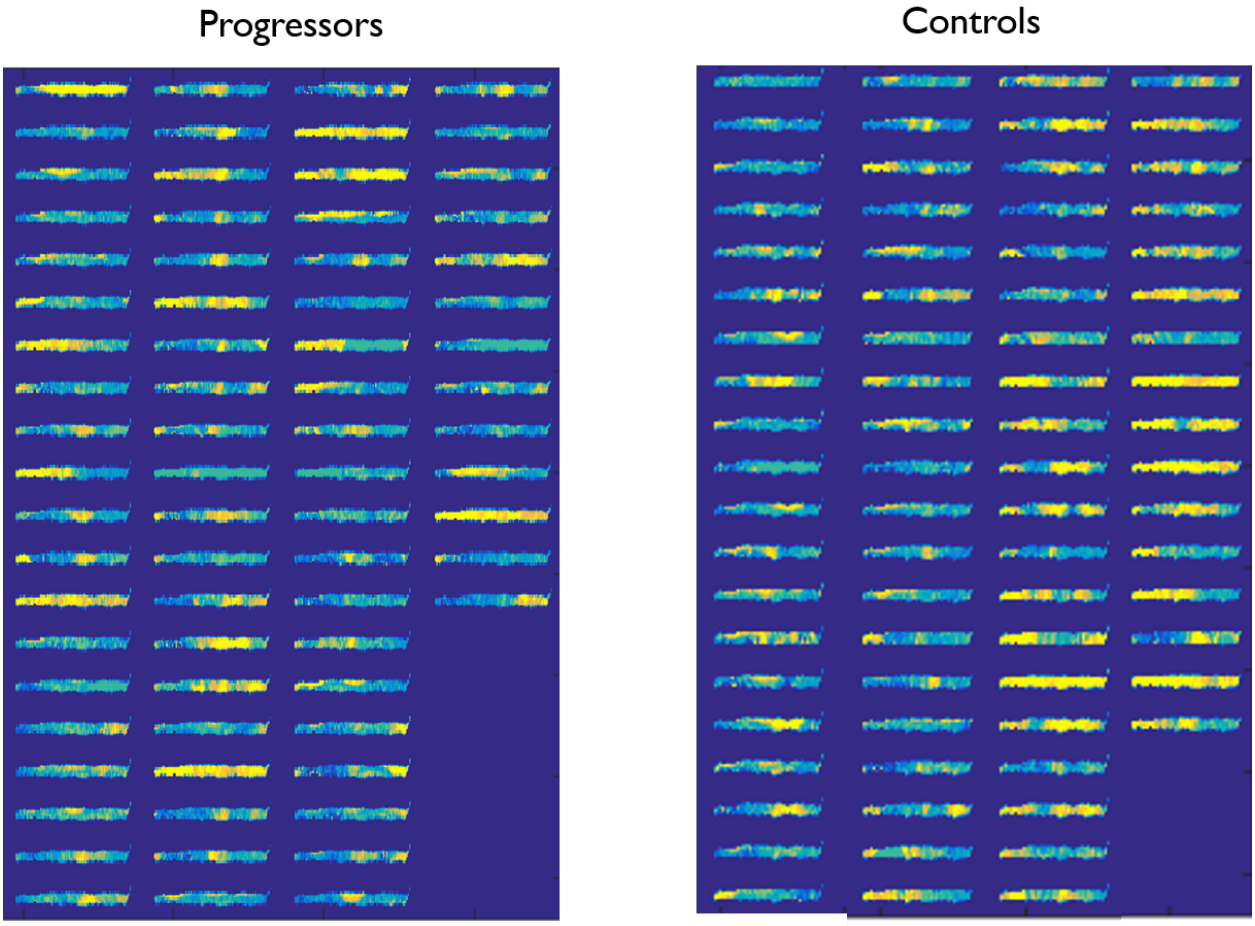
The help of computer-aided detection (CAD) is sought to help answer the questions such as the ones above. In addition to high-throughput analysis, the goal of computer-assisted techniques is to extract patterns, trends, and correlations missed by human visual inspection alone. One area where CAD tools have already been developed for commercial use is in automated analysis of PAP smears [18].

However, data mining from medical images and automated inference are complex tasks fraught with many obstacles and challenges. In this chapter, we will first review several of the current techniques and challenges. Then, we will state the specific contributions of this thesis. Finally, we will present an outline for the remainder of the thesis.

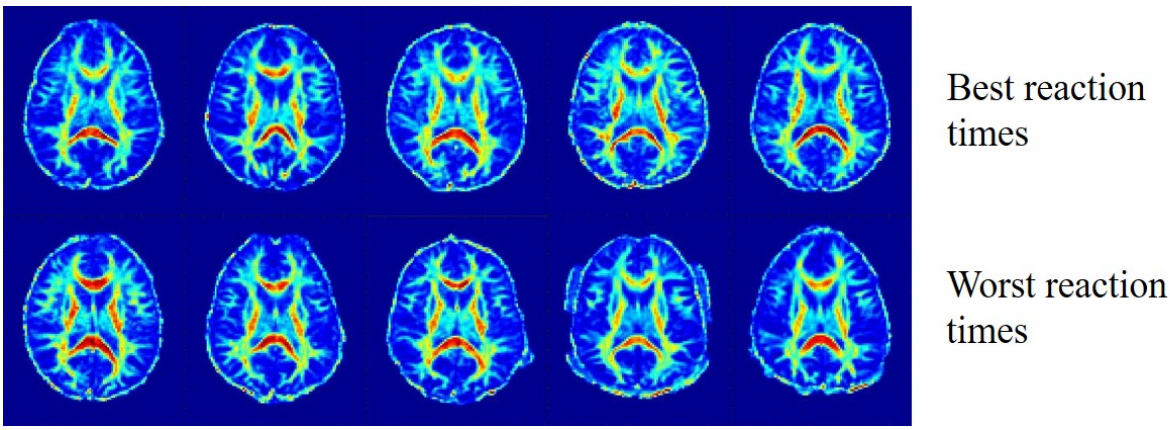
Understanding the changes that are as of yet hidden from the human eye would have far reaching impact on early diagnosis, understanding diseases commonly considered *medical mysteries*, and contributing objective measures to help clinical assessment.

1.2 Challenges

Challenges exist at every stage of the CAD pipeline, as Figure 1.2 illustrates. First, feature extraction from biomedical images is a challenge. What quantities to measure from the images is an open question in many problems, and one that often requires human input. The second stage of the pipeline is modeling. What is the best way to incorporate the features measured into a model that will produce accurate clinical assessment? Finally, in order to draw new clinical insights, we must answer, how does the model enable reliable clinical assessment based on imaging? What do the structural correlates of observed behaviors look like?



(a)



(b)

Figure 1.1: (a) Knee cartilage MRIs of healthy subjects. Three years in the future, one of these groups will go on to develop osteoarthritis (Progressors), (b) Fractional anisotropy maps computed from diffusion tensor images of 10 different post-concussive patients.

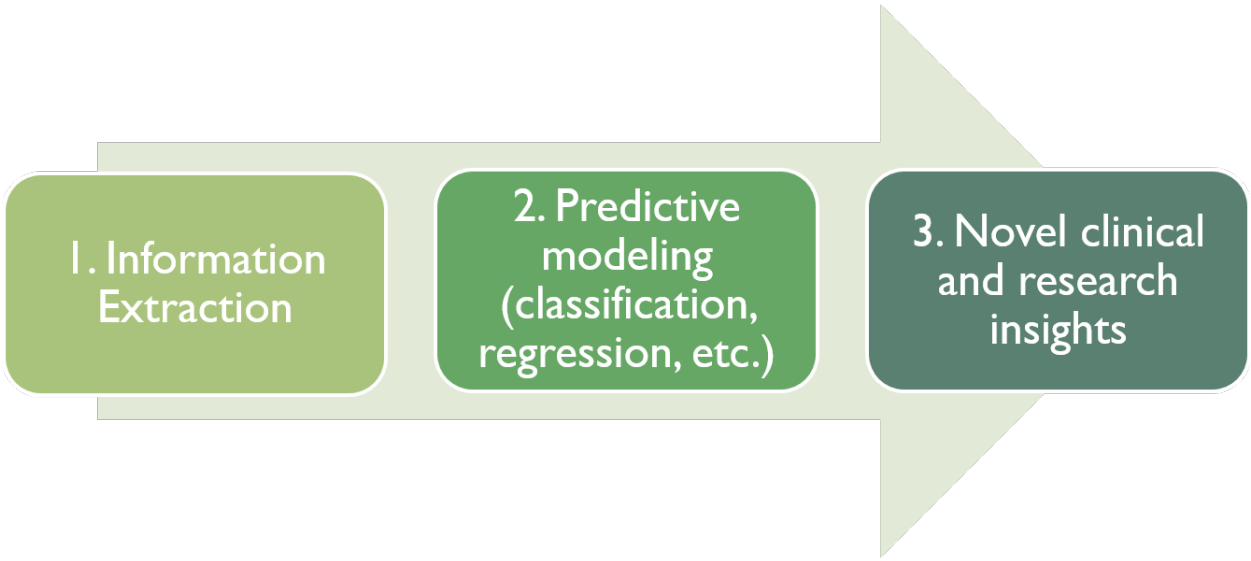


Figure 1.2: Challenges exist at every stage of the pipeline from data mining to automated inference. Credit: [122]

1.2.1 Feature extraction

In their book *Computer Vision*, Ballard and Brown [11] draw attention to the importance of picking good features. They state "picking good features is the essence of pattern recognition. No elaborate formalism will work well for bad features."

Often, the features chosen are pre-specified descriptors. For example, some features quantified in the commercial devices developed for automated PAP analysis are nuclear area, nuclear density, cytoplasmic area, etc. [18]. These features are selected based on human input and map the original image to a set of numbers for statistical analysis.

However, often our prior knowledge of the problems under study is so limited that it is unclear what features to measure or extract in the first place.

Another challenge is that most feature sets are *non-generative*; although an arbitrary image can be mapped to a numerical feature set, images cannot be recovered or visualized given an arbitrary set of numerical features. The latter poses an impediment in visualizing and inferring physical meaning from statistical models constructed in the feature space based on numerical descriptors

[18].

1.2.2 The problem with off-the-shelf machine learning techniques

Machine learning techniques are currently employed in ”big data” applications such as internet retail and media, but these methods cannot generally be applied off-the-shelf to biomedical images. There are several reasons why. First, there are often far fewer data samples, or patient images, than dimensionality of the images. Typical MRI research studies include several tens of subjects, perhaps a hundred, whereas the images may contain millions of voxels. In such high-dimensionality, low-sample size (HDLSS) data, the sampling of the underlying signal space is sparse; every data sample appears to be very different, or ”far”, from every other data sample. Even if the correct features were selected, in a situation with few training samples, classifiers are often prone to over-fitting the data. As a classifier becomes more complex, its variance increases [89]. Machine learning approaches based on deep learning require many training samples in order to construct a representative model - many more training samples than we have often have available. On the other hand, Occam's razor reminds us that simpler explanations are preferred over more complex ones because they require fewer assumptions. Thus, given that the correct features are chosen the next step is to select simpler models that can predict the data.

1.2.3 Nonlinear modeling

Another complicating factor is that the data is nonlinear. This means that the sum of two images belonging to Class 1 does not necessarily produce an image that also belongs to Class 1. Figure 1.3 illustrates the nonlinear nature of structural brain images.

Trying to interpolate between an older adult brain I_2 and a middle-aged adult brain I_1, we see that the linear sum of pixels (Eulerian in PDE parlance) leads to an interpolated brain has significant artifacts in the ventricles. However, the image interpolated using a nonlinear transport-

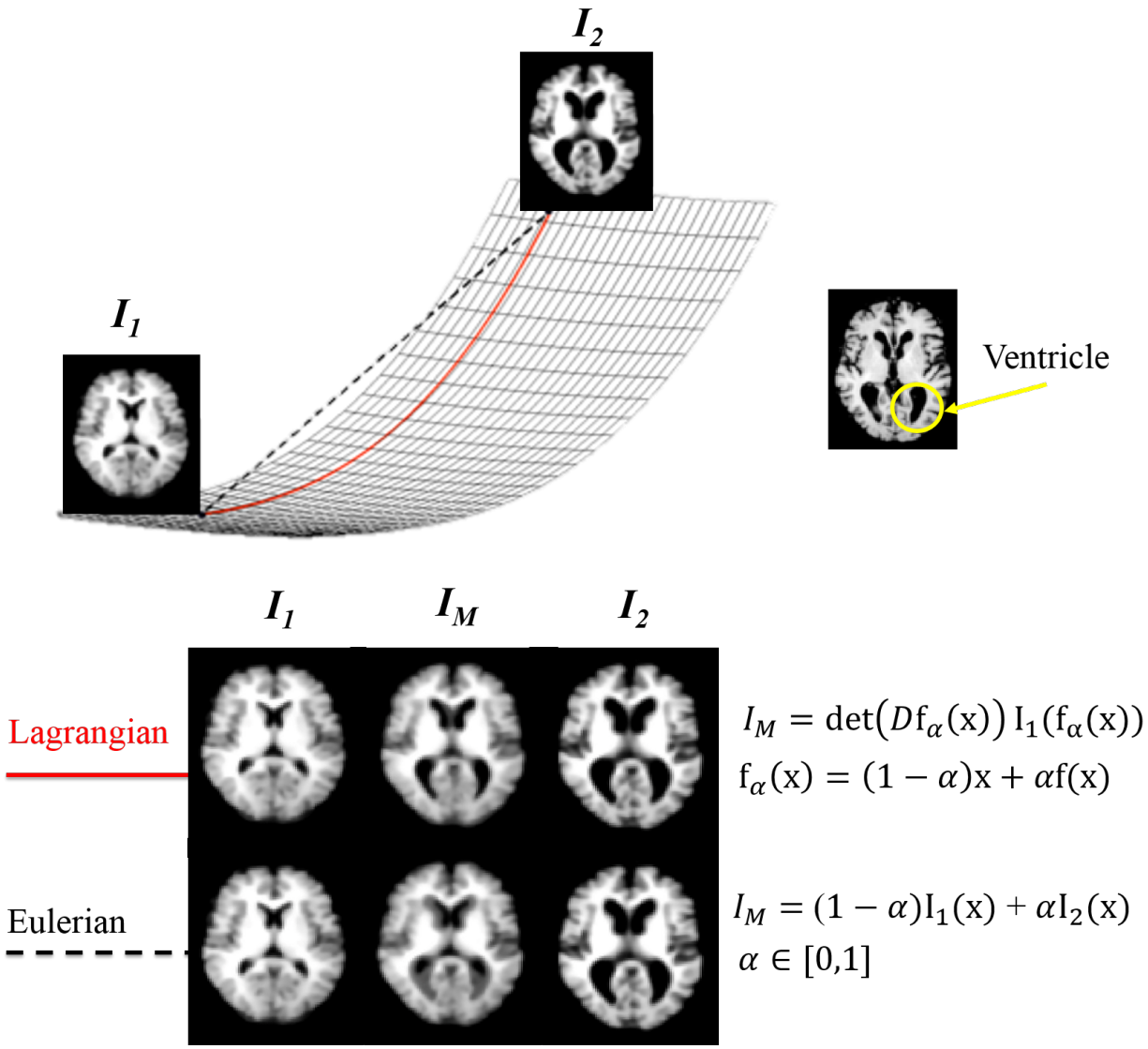


Figure 1.3: I_M is generated using MRIs of a middle-aged (I_1) and healthy older subject (I_2). I_1 and I_2 were affinely coregistered before interpolation.

based model (Lagrangian in PDE parlance) seems to better capture aging-related changes such as tissue thinning and expansion of ventricles. In this case, a linear combination of normal brain images does not result in a normal brain image. Thus, the data appear to lie in metric spaces defined by nonlinear measures of distance.

The need for nonlinear modeling raises the issue of how to define a geometrically meaningful measure of similarity between samples in the dataset, and furthermore, how to learn the nonlinear structure of the data, especially given its HDLSS nature.

The following sections will review current approaches for quantifying similarity between radiology images. Given that the data lie in a nonlinear space as Figure 1.3 suggests, the Euclidean metric is not typically used directly. It is known that Euclidean geometry is not well-suited for curved spaces and Riemannian geometry is needed [45].

1.3 Current approaches for feature extraction and modeling

What is needed is a distance metric that can adequately capture the level of similarity or difference between two images. As previously mentioned, the simple Euclidean metric is not adequate to model anatomical images. Unfortunately, most current approaches are not *generative*. That is, although the specific images in the sample can be represented by the model, the model is not able to generate what new image samples would look like. The latter is a critical impediment because statistical analyses performed in the feature space (i.e. on model parameters) cannot be translated to visualizable images in order to infer physical meaning of the found associations [53].

1.3.1 Numerical descriptors

One commonly used method to search for morphologic differences between subjects is to quantify and compare a set of pre-specified numerical descriptors in a trial and error manner. Commonly used features for trial and error are numerical descriptors, such as histogram statistics, texture

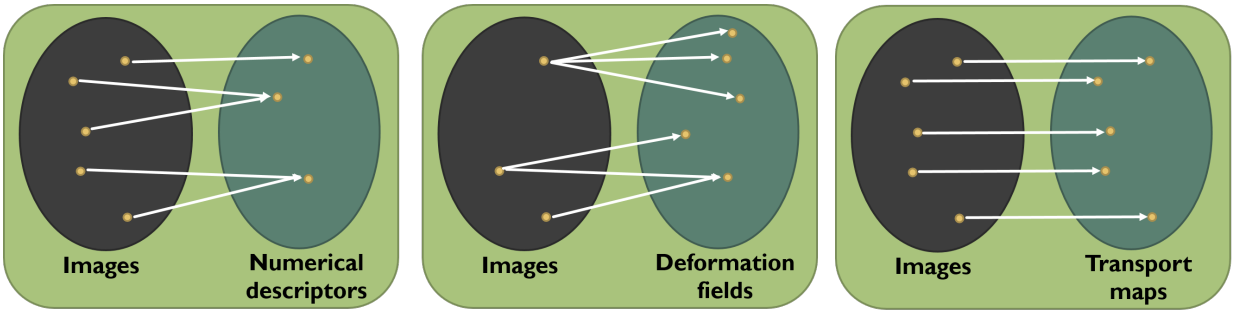


Figure 1.4: Some feature extraction techniques. Only transport-based feature extraction is bijective.

features, gray-level-run-length (GLRL), wavelet features, Zernicke polynomials etc. Then, it is assessed which features are most useful in discrimination either by statistical tests or by building classifiers to determine which features lead to the best accuracy.

An example of a software that utilizes the trial and error approach for biomedical image analysis is WND-CHRM [115], which computes over 2600 features for testing. In building a classifier, features among the initial set are tested and selected for their discriminating ability. The FreeSurfer software [50] for brain imaging in particular samples $\sim$17,000 points from the cortex and computes local features such as thickness, convexity, etc. Again, in later tasks of pattern recognition and discrimination, these features (ex. cortical thickness, convexity) are tested and retained based on discriminating and predictive ability.

There are several drawbacks of the trial and error approach, however. First, mapping an image into a set of numerical descriptors inevitably discards information present in the original image. The information discarded may be potentially important in understanding the phenomenon of interest. Second, numerical descriptors do not take into account *a priori* image models or known anatomy, missing the opportunity to quantify variation of images as compared to a known template. Third, having to choose numerical descriptors enables specific hypotheses to be tested, but does not enable exploratory analysis. Finally, commonly used numerical descriptors like Haralick features or Zernicke polynomials, Tamura texture statistics, histogram statistics, etc. have no di-

rect biological interpretation. Furthermore, as Figure 1.4 illustrates, mapping an image to a set of numerical descriptors is not generative; although an arbitrary image can be mapped to a feature set, an arbitrary feature set cannot necessarily be mapped to an image. Thus, direct visualization of statistical models constructed in the feature domain is often precluded.

1.3.2 Deformation-based methods

The desire to discover trends in a data-driven, exploratory manner gave rise to morphometric techniques by John Ashburner and colleagues based on non-rigid registration. These are Voxel-Based Morphometry (VBM) [6], Deformation-Based Morphometry (DBM) [8], and Tensor-Based Morphometry [7]. These methods quantify shape variations in images when compared to a common reference in a *deformation field*. Subsequent analyses is often performed either on the deformation field, or the nonrigidly registered images. There are several open-source software packages available to perform deformation-based analysis. There include Statistical Parametric Mapping (SPM) [52] and FMRIB Software Library (FSL) [125].

First, deformation fields are best suited to model local volume expansion/contraction and are unable to capture variations caused by altered biophysical properties of tissue that appear as *texture* differences. Figure 1.5 illustrates several neurologic diseases for the main variations are in texture changes. Furthermore, the results of pixel-wise comparison often suffer from systematic registration bias, producing results that are a function of pixel misalignment as well as biological differences [24].

The authors state that while VBM can help index regionally into certain voxels where some metric of brain morphology may be affected by the factors under study, VBM cannot help answer the nature of these changes [53]. Another limitation of pixel-wise comparison is that it is mass-univariate; projecting a multidimensional morphologic space onto individual voxels can capture effects that are relatively localized but not those that are spatially distributed and involve multiple structures [40]. Furthermore, assessing group differences based on statistical parametric maps can-

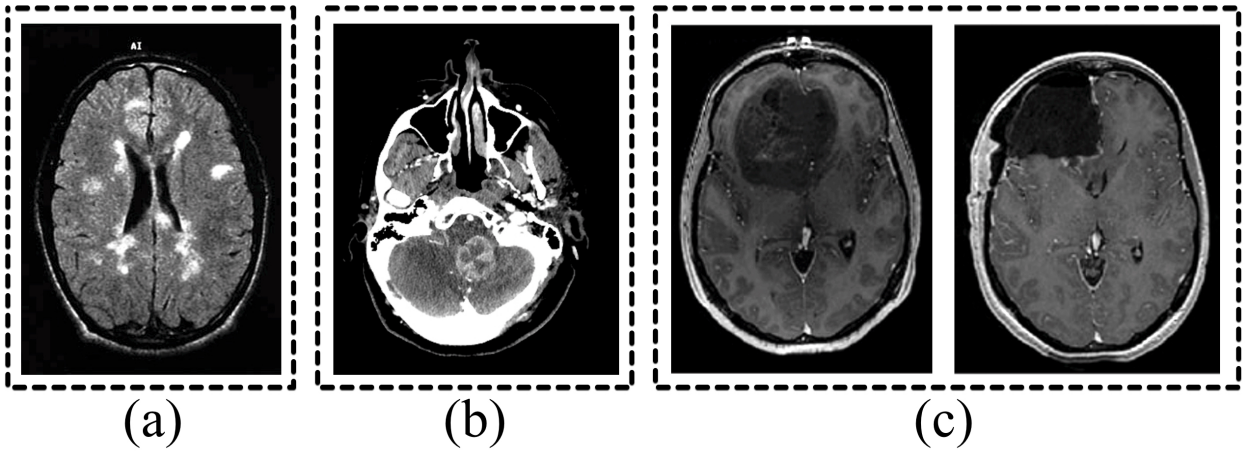


Figure 1.5: Neurologic conditions where shape analysis is not adequate to capture pathology affecting biophysical properties of tissue manifesting as texture variation. (a) Multiple sclerosis (source: [74]), (b) metastatic breast cancer tumor (source: Dept. of Radiology, University of Pittsburgh Medical Center), (c) GBM (left) and debulking procedure (right) (source: [31])

not necessarily distinguish intra-group variations from those at the interface between two groups [40]. Therefore, the answers offered are based on a narrow set of hypotheses.

Additionally, as Figure 1.4 illustrates, deformation fields are not unique to the source image; one image may map to many deformation fields. These methods are not generative and although it may be possible to localize differences, direct visualization is not possible.

Another class of deformation-based techniques focus on modeling shape and are those based on differential geometry. For example, approaches in computational anatomy attempt to match the cortical shape and those of deep nuclei in the brain by minimizing the amount of incremental effort to match the structures along a time dimension [57]. Similar approaches by Joshi *et al* [76] propose a landmark-matching approach to match shapes of structures based on large diffeomorphic transformations. The advantages of such approaches is that they do not lose information about shape, are generative, and enable interpolation between shapes that helps researchers understand and visualize the geometric properties of shape distributions. A limitation of these techniques, however, is that they focus on the analysis of shape to the exclusion of other variations that may be present in the dataset, such as texture.

1.3.3 Other approaches for nonlinear modeling

Active appearance models [38] and morphable models [21] have also been used to model the appearance of the brain [69]. Morphable models compose the appearance of an unknown object, such as a brain MRI, using a linear combination of known objects. Active appearance models compute modes of variation when normalized for the mean shape or texture. Although these approaches are nonlinear and generative, they require landmarks to be specified and their point-by-point correspondences.

Finally, manifold-learning techniques have been applied to learn the geometric structure of variations in MRI databases [28]. However, these techniques are not generative in general. That is, the graph-based structure is constructed from given samples in the dataset, but cannot be used to generate new images given an arbitrary point in the feature space.

1.3.4 Linear optimal transport

In contrast to other approaches, the linear optimal transport (LOT) distance metric is promising in being able to meet the goals of an automated framework for discovering and visualizing structural markers from MRI that can clinically differentiate patients. By viewing images as normalized nonnegative probability measures, the similarity between images can be quantified in a *transport map* using the Wasserstein distance, which is also known as Earth-Mover's Distance (EMD) in the field of computer science. The EMD measures the amount of work, defined by mass times distance transported, required to re-distribute the pixel intensities in one image to match another exactly using the mathematics of optimal mass transport (OT). In this process, texture and shape modeling are performed simultaneously. As Figure 1.7 illustrates, optimal mass transport is able to match two images exactly, capturing both shape and texture differences up to an interpolation error. Furthermore, the LOT embedding of the transport map is also generative, meaning that visualization of images is possible and does not require specification of any landmarks.

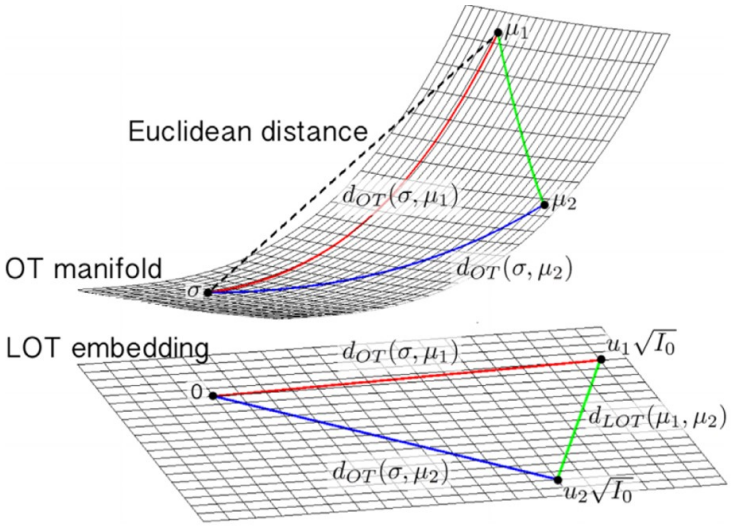


Figure 1.6: Source: [81]. Projecting distances on the manifold to a tangent space with respect to a template σ. The LOT-based modeling is generative, enabling images corresponding to arbitrary feature sets to be visualized.

The metric space defined by OT is a Riemannian manifold equipped with an inner product, where each point on the manifold can define a tangent space. The OT distance between two images on this manifold is the length of the shortest curve that connects them, also called the geodesic. The linearized version of this metric is defined by the projection of the OT distance onto the tangent space with respect to the common reference, as Figure 1.6 illustrates.

The key benefit of the linearized OT metric is that now *Euclidean distances defined in the LOT space correspond to modified versions of the OT metric*, which is referred to as a *generalized geodesic* [1, 81]. This is important because the implication is that complex, nonlinear morphology in the image space can be represented as simple, linear relationships in the LOT space. Indeed, recent work [12, 84, 135, 136] has demonstrated that examining 1D and 2D signals in the LOT space enhances their separability, aiding classification, discrimination, and visualization tasks. However, the mathematics of LOT have never been expanded to include 3D radiology images.

The distance between any two arbitrary images in the dataset can be computed by a linear functional of computed transport maps. This means that the number of transport maps that need to

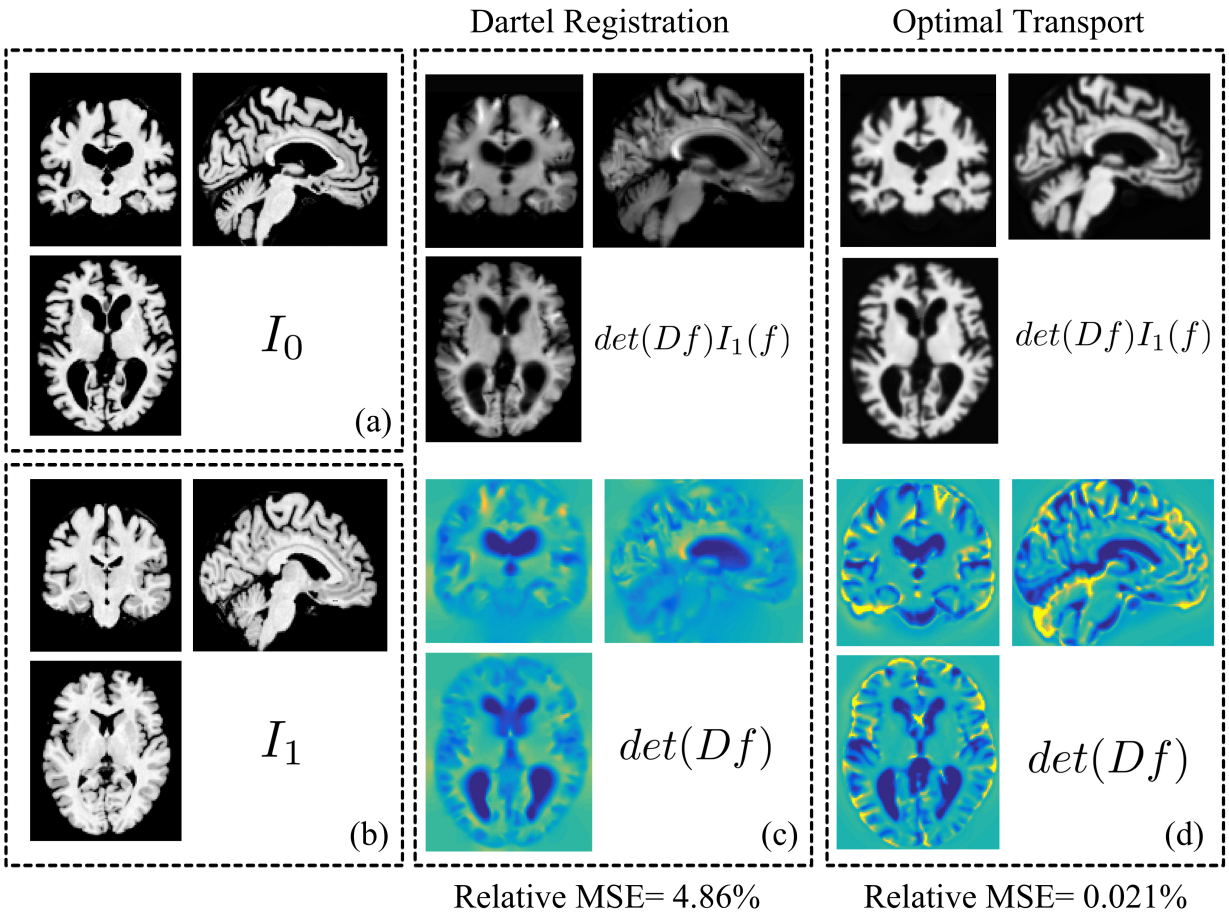


Figure 1.7: Compared to deformation fields computed using DARTEL [5], transport maps computed using optimal mass transport (OMT) captures both shape and texture differences between I_0 and I_1 and match images perfectly, up to an interpolation error. However, deformation fields lose texture variation information, thus resulting in high MSE when attempting to match source and target images.

be computed is N, where N is the number of images in the dataset, rather than all pairs of images that would require $\frac{N(N-1)}{2}$ computations, which saves number of computations needed [81].

This thesis focuses on how the mathematics of LOT can be adapted to create a framework to assist vision and detection tasks in radiology images.

1.4 Specific Contributions

This thesis builds upon previous work demonstrating that a linear optimal mass transportation framework [135, 136] can enhance linear separability of signals [106, 134] and ability to visualize discriminating changes [12, 81]. This work develops a Transport-Based Morphometry (TBM) approach for the first time that includes radiology images. The work in this thesis specifically focuses on analysis of magnetic resonance images. These images are large, often three-dimensional and easily comprise multiple million voxels. Current numerical OT methods suffer from difficulties including high computational complexity, non-convergence, and large numerical errors that result in poor, numerically non-invertible mappings with large medical images. The first part of this work focuses on developing an OT approach suitable for performing TBM with large medical images. In subsequent parts, we apply the developed approach toward several open clinical and research problems to enable state of the art results in these areas. In the process, we demonstrate how the TBM approach can be combined with simple machine learning techniques to enable classification, regression, and visualization in practical problems. The specific contributions in this thesis are:

- **Contribution 1: Develop a Multiscale OT approach for 3D radiology images.**

  The basis for this aim is prior work in cell microscopy imaging, which demonstrates that transforming images from an Eulerian (intensity-based) representation of an image to a Lagrangian (location-based) representation based on optimal transport facilitates problems in signal detection, classification, modeling and visualization of disease-related changes. Using concepts from optimal transportation theory, we extend the framework for the first time to radiology data, enabling analysis of tissue spatial distribution. In the process, we create a novel, more robust formulation of our Transport-Based Morphometry (TBM) technique and present a solution using variational optimization. The new formulation enables the TBM framework to be extended to 3D radiology images and avoids pitfalls and inadequacies of existing formulations.

- **Contribution 2: Classification and Discrimination tasks.**

  We apply the new TBM approach to a variety of discrimination and classification tasks in both clinical and research domains. One application area is in preclinical detection of osteoarthritis (OA) based on knee MRI. Results show that our TBM technique can predict future onset of OA 3 years in the future with 86% accuracy based on T2-weighted image appearance of cartilage. In studying the effects of 16p11.2 deletions/duplications on brain structure, we demonstrate that TBM enables 96% accuracy in predicting CNVs of 16p11.2 based on brain imaging studies alone, and illuminates the structural changes that enable high classification accuracy.

- **Contribution 3: Regression and Correlation tasks.**

  We extend the TBM framework as a discovery tool for statistical relationships between a continuous clinical variable and imaging data. We demonstrate that TBM is able to assess structural correlates of cognitive difficulty post-concussion. We also utilize the TBM approach to investigate the effects of aerobic fitness on brain tissue distribution in older adults. We discover that aerobic fitness influences areas of the brain that overlap with areas affected in normal aging.

1.5 Outline

This thesis comprises six core chapters. This chapter describes the problem statement and context for this work through practical, motivating examples. Chapter 2 lays the mathematical foundation related to optimal mass transport, including a brief exposition reviewing commonly-used solvers and key theoretical results. Chapters 3-5 are based on original research work.

Chapter 3 describes a new formulation and solution to the optimal mass transport problem that enables the Transport-Based Morphometry approach in this thesis for radiology images.

Chapter 4 describes a framework for utilizing TBM for classification and discrimination tasks.

TBM is applied to enable state of the art results in early detection of osteoarthritis and elucidating the relationship between gene and brain structure in 16p11.2 duplication/deletion carriers.

Chapter 5 describes a framework for regression and correlation analysis using TBM, and demonstrates how TBM enables state of the art results in prognosticating cognitive deficits after concussion and understanding the effects of aerobic fitness on older adult brains.

Finally, Chapter 6 concludes this work with a brief summary and directions for future research.

Chapter 2

Optimal Mass Transport and Use in Medical Image Analysis

2.1 Overview and Goals

The optimal mass transport (OT) problem is not new, but has found important applications in several fields. The original problem was posed by Gaspard Monge in 1781 [131]. He sought the most parsimonious way, in the sense of mass transported, to rearrange a pile of dirt into a castle of the same mass (of course with different distributions). Much of modern OT theory, however, was developed in 1942 by Leonid Kantorovich for the problem of resource allocation. He shared the 1975 Nobel Prize in Economics with Tjalling Koopmans.

Several formulations and numerical solutions have been proposed. While the discrete version of the problem can be solved by linear programming, the continuous version of the problem requires optimization to solve in dimensions higher than one. We review some of this work in this chapter.

Today, OT has found applications in many fields ranging from meteorology to image analysis. In this chapter, we focus on the applications specifically related to image analysis in content-based image retrieval, image registration, pattern recognition, and super-resolution imaging.

2.2 Review of Prior Work Related to Optimal Mass Transport

In this section, we briefly review the theory related to the optimal mass transport problem as well as describe existing numerical solvers.

Let Ω be a measurable space. Let μ and σ be probability measures defined on Ω, with corresponding positive probability densities I_1 and I_0, respectively. A mass preserving transform f that pushes σ to μ, or $f_{\#}\sigma = \mu$, satisfies the following,

$$\int_A d\sigma(x) = \int_{f(A)} d\mu(x), \quad \forall A \subset \Omega. \tag{2.1}$$

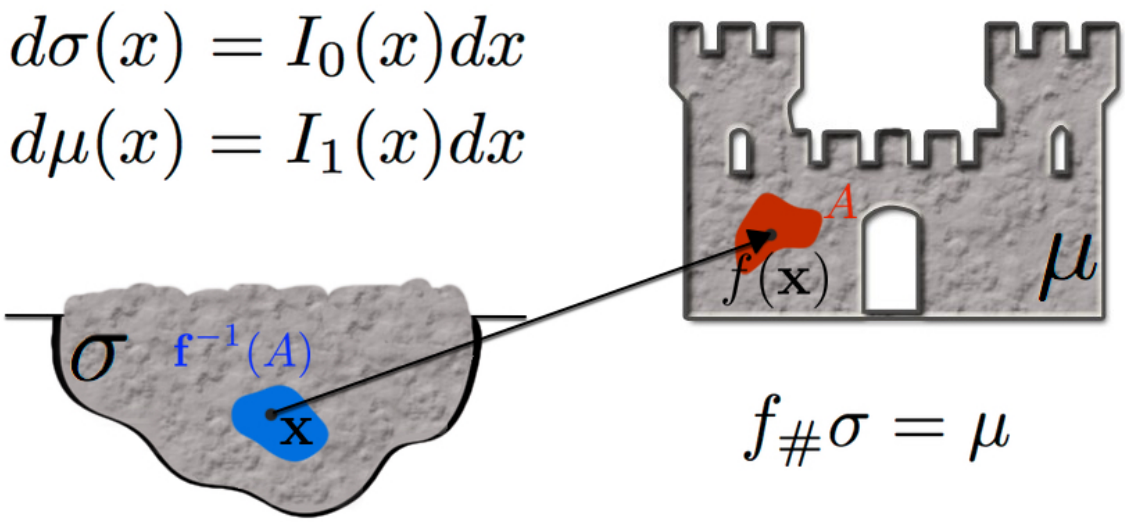


Figure 2.1: What should the optimal coupling be between units of mass in the pile of dirt and units of mass in the castle such that the transportation cost is minimized? Image credit: Soheil Kolouri

Figure 2.1 illustrates μ and σ, as well as the map $f_{\#}\sigma = \mu$. Such a mass preserving (MP) mapping f is in general not unique; in fact, infinitely many MP mappings may exist that satisfy Equation 2.1. However, we are interested in finding the MP mapping that is optimal in the sense of

mass transport, which we will define further in Equation 2.2. Let MP be the set of all such mass preserving mappings, $MP := \{f : \Omega \to \Omega | f_{\#}\sigma = \mu\}$. The optimal MP mapping in the mass transport sense can be written according to Monge's formulation, which minimizes the following cost function,

$$\min_{f \in MP} \int_{\Omega} c(x, f(x)) d\sigma(x) \tag{2.2}$$

Here, $c : \Omega \times \Omega \to \mathbb{R}^+$ is the cost functional. The functional c measures mass transportation cost and is often chosen to be the L^p-norm for which Equation (2.2) becomes the L^p-Wasserstein distance. The L^2-Wasserstein distance, $c(x, y) = |x - y|^2$, in particular has attracted rich attention in the image analysis, computer vision, and machine learning communities. For $c(x, y) = |x - y|^2$, Brénier [26] showed that there exists a unique optimal transportation map $f \in MP$ for which,

$$\int_{\Omega} |x - f(x)|^2 d\sigma(x) \leq \int_{\Omega} |x - g(x)|^2 d\sigma(x), \quad \forall g \in MP \tag{2.3}$$

when (i) $\Omega = \mathbb{R}^n$ and the probability measures have finite second-order moments (i.e. their densities vanish in the limit),

$$\int_{\Omega} |x|^2 d\mu(x) < \infty \text{ and } \int_{\Omega} |x|^2 d\sigma(x) < \infty,$$

and (ii) when σ is absolutely continuous with respect to Lebesgue measure.

Moreover, Brénier showed through polar factorization [26] that the transport map f must be the gradient of a convex function $\phi : \Omega \to \mathbb{R}$, $f = \nabla\phi$. The preceding property implies that when Ω is a convex and connected subset of $\mathbb{R}^n$, the optimal transport map is *curl free*. The

curl-free characteristic of the optimal transport map has been used in a variety of solvers utilizing optimization. Angenent *et al*. [2] and Haker *et al*. [62] solve for an initial MP map f_0 (not unique or optimal) through the Knothe-Rosenblatt rearrangement [22, 131] and then progressively update the initial map using composition (with another MP map s that satisfies $s_{\#}\sigma = \sigma$) so that it becomes curl free to signify optimality.

There are, as pointed out by Haber *et al*. [60] and Rehman *et al*. [128], two main shortcomings to the preceding numerical approaches. First, a robust method is needed to obtain an initial MP mapping, and the obtained initial map is often far from the optimal transport map. Second, and much more importantly, such methods update the transformation in a space which is tangential to the linearized MP constraint. Hence, for any finite step update used in the optimization, $f_0(s^k)$, the mapping deviates or *drifts* from the set of mass preserving mappings MP. While the level of *drift* may or may not be acceptable in practice for a 2D solution, the drifting is amplified for 3D images as demonstrated in Section 3.7. The *drift* phenomenon necessitates solution by alternative methods for 3D images.

Other convergent methods have also been proposed based on a fluid dynamics formulation of the problem [16, 88] or based on the solution of the Monge-Ampère equation [17]. Benamou *et al* [16], for instance, reformulated Monge's problem (Equation (2.2)) by considering all possible smooth enough, time dependent, density and velocity fields subject to the continuity equation. This formulation comes at the cost of an additional virtual time dimension, which is computationally expensive. For 2D images, we must solve a 3D problem and for 3D images, we must solve a 4D problem. Computational demands pose a challenge. Saumier *et al* [114] proposes a numerical solution that requires a PDE to be solved using a discrete linear system of equations. In medical images, and especially in 3D, where the number of voxels can easily number in the millions, solving a linear system of equations using Gauss-Siedel or Jacobi solvers requires hundreds to thousands of iterations over each of millions of voxels. The number of matrix systems to be solved may quickly exceed 10^{10}.

Another family of solvers [33, 39, 119] are based on Kantorovich's formulation of the problem. In short, Kantorovich's formulation searches for the optimal transport plan π defined on $\Omega \times \Omega$ with marginals μ and σ that minimizes the following,

$$\min_{\pi \in \Pi(\mu,\sigma)} \int_{\Omega \times \Omega} c(x,y) d\pi(x,y) \tag{2.4}$$

Here $\Pi(\mu, \sigma)$ is the set of all transport plans with marginals μ and σ. Optimization of Equation (2.4) in the discrete setting leads to a linear programming problem, which is computationally expensive. For large 3D images, the computational cost may even become prohibitive. Cuturi *et al* [39] used an entropy regularized version of the Kantorovich problem, which leads to an efficient solver based on the Sinkhorn algorithm. A shortcoming of such methods in medical imaging applications is that the concept of a *deformation map* is missing from the obtained transport plan. The reason is that a transport plan enables mass to be split and redistributed to push one measure into another. Moreover, the entropy regularized version of the problem encourages mass splitting. The methods based on the Kantorovich formulation are not readily applicable for medical image analysis. Chartrand *et al*. [33], on the other hand, retain the concept of a deformation map by solving the dual problem to the Kantorovich formulation. Chartrand *et al* obtain the optimal transport map through a gradient descent solution. The obtained transport map as pointed out in [33] and reconfirmed in this thesis comes with the tradeoff of undesired *artifacts*, especially when the images are not smooth. Thus, additional work is needed to overcome challenges related to quality of MP match with the Chartrand *et al* [33] approach.

Finally, there is work that describes an OT solution for shape matching that can match 3D shapes [121]. However, this work focuses on shape analysis, rather than volumetric density modeling of a 3D image. Second, this kind of method requires specification of landmarks and rigorous correspondence of these landmarks to be established between images. The need to specify additional parameters impedes a fully automated approach.

OT solvers in the existing literature primarily address smaller images, including those in 2D, and do not scale well to large 3D medical image datasets.

2.3 Applications of OT in Image Analysis

Recently, OT has generated excitement in the imaging community for its utility in image analysis. OT has been used in content-based image retrieval [113]. In medical imaging, OT is shown to be advantageous in interpolating time-series data (such as cell growth or migration) [55, 82], performing pattern recognition on cell microscopy images (classification, regression) [12, 70, 84, 126, 136], elastically registering images for surgical planning [60, 62, 128], and even segmenting brain structures [59]. There is high potential for OT-based solutions to play an important role in many new advances involving biomedical imaging.

2.3.1 Image Retrieval

Given an image, the retrieval problem consists of finding other images that are similar to the image in a database, illustrated by Figure 2.2.

A standard way to retrieve images is to measure a set of features from the image, such as texture features, color statistics etc. These features then lead to histograms of image signatures. However, the histograms do not always employ the same bins; in fact some histograms utilize dynamic binning. Comparing these histograms can be a challenge, especially when the binning does not directly correspond.

Several papers [87, 107, 113] have used the Earth Mover's Distance (EMD) metric to compare histograms. Compared to other metrics such as L_1 distance, L_2 distance, χ^2 statistic, and Jeffrey's divergence, [113] show that the EMD achieves the highest precision-recall performance.

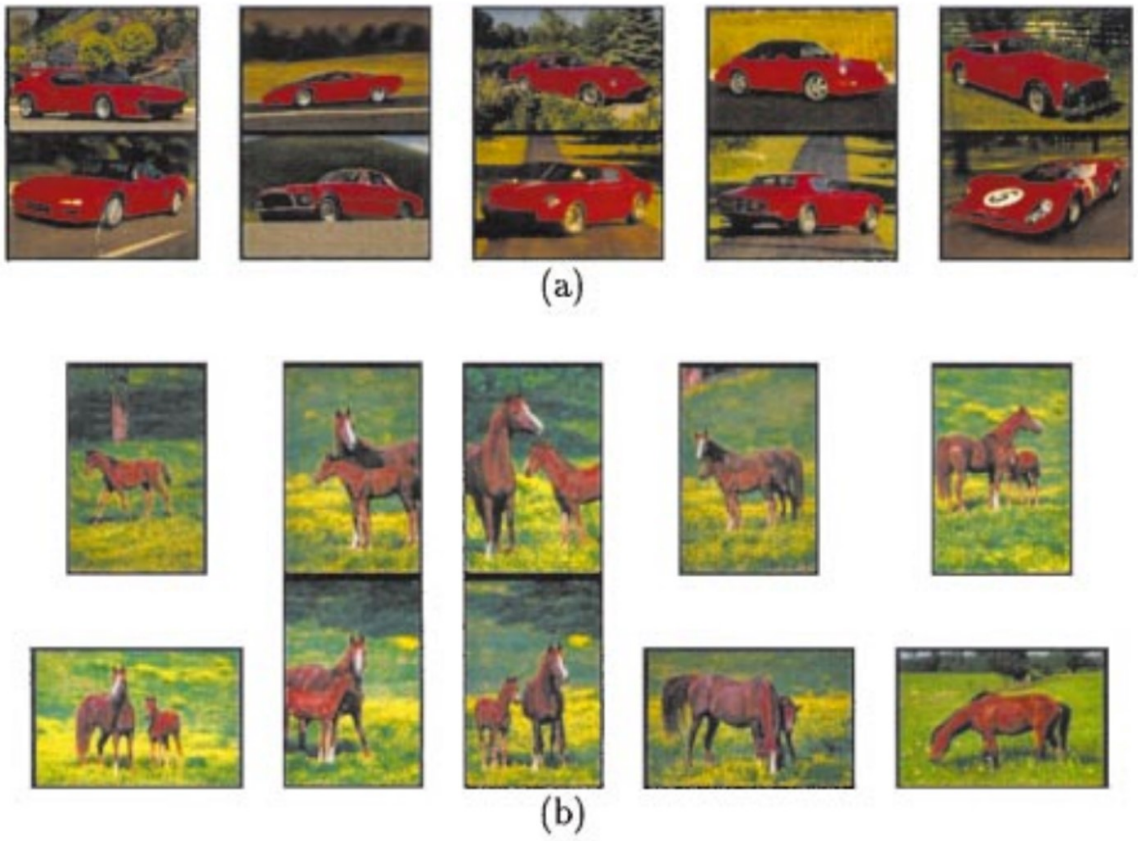


Figure 2.2: (a) red cards and (b) horses showing pictures that are alike in a database. Source: [113]

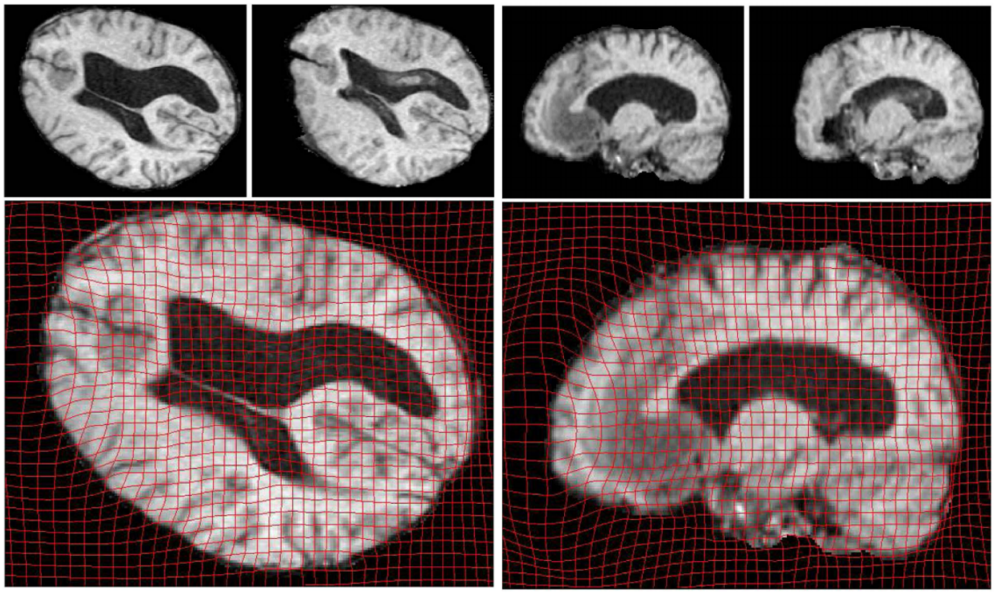
2.3.2 Registration

Image registration is the problem of computing a geometric transformation between the coordinate systems of two different images in order to achieve maximal alignment between the images.

Although the registration problem may include many different transformations, such as rigid-body, affine, elastic, etc., as well as similarity metrics in order to measure degree of alignment, such as mutual information, cross-correlation, intensity-based, etc., intensity-based elastic registration is often of interest in medical imaging applications. Lung CT images on expiration and inspiration contain constant amount of tissue, but different tissue densities as the lung expands and contracts during breathing. Cardiac MRI that shows the beating heart also shows constant tissue, but changing tissue density during systole and diastole. Proton-density imaging MRI and fMRI also show changing densities [61, 62].

Optimal mass transport is an attractive method for computing the transformation to elastically register two images. First, it is naturally suited to compare two tissue densities, especially when the tissue mass is fixed such as in lung CT and cardiac MRI. Second, the computed registration by OT is landmark-free and parameter-free, as it is an intensity-based method. Third, the transformation is symmetric, meaning that the order of the images does not influence the results. Finally, the transformation using OT is the solution of a convex problem with a unique minimizer.

Optimal transport has been used for elastic registration in [60, 62, 128]. Some papers have even utilized the OT-based elastic registration method to enable segmentation of brain structures [59, 60]. Figure 2.3 shows how elastic registration can quantify the appearance of the brain before and after a tumor is removed. Figure 2.4 shows how elastic registration using OT can be used to parcellate cortical structures.



(a) (b)

Figure 2.3: Source: (a) Pre-operative (left) and post-operative (right) axial slices and deformation shown (below), (b) Pre-operative (left) and post-operative (right) sagittal slices and deformation shown (below). Source: [128]

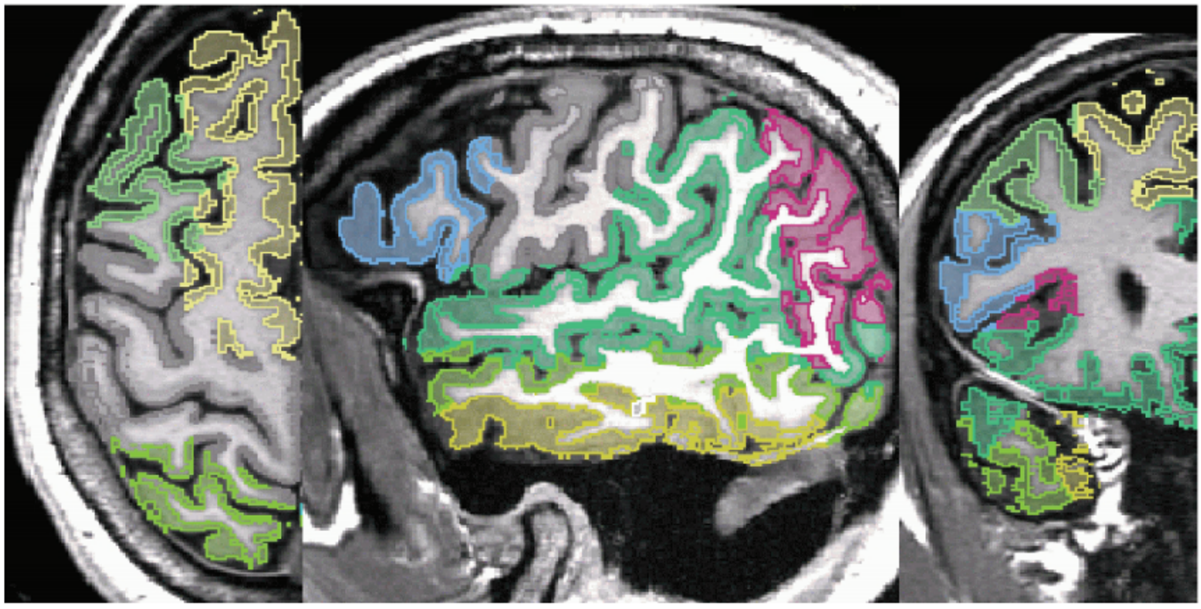


Figure 2.4: Use of OT-based registration to elastically register images for segmenting cortical structures. Source [60]

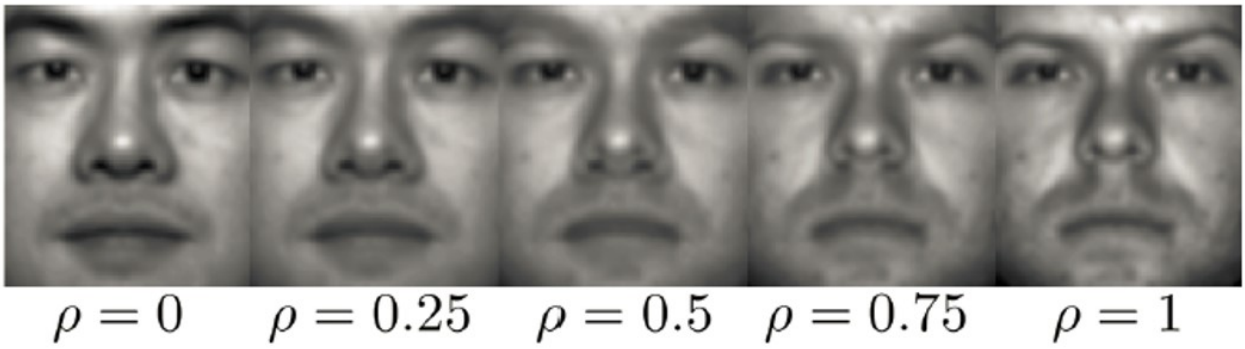


Figure 2.5: Source: [83]. Morphing one face to another, with distance along the OT geodesic indicated by ρ

2.3.3 Image Reconstruction

Color synthesis [49] and texture synthesis applications, as well as super-resolution imaging have been addressed using transport-based metrics [83].

The problem of super-resolution involves constructing a high-resolution image from a low-resolution image. The geodesic on the OT manifold provides a morphing from a template image to a sample face image. In the training phase, a basis is learned in the transport space for the displacement maps needed to compose a morphing from any face image to the template. Figure 2.5 is an example of morphing one face to another.

In the testing phase, the morphing is sought that gives the most likely high-resolution face image in terms of maximum a posteriori probability. In [83], it is demonstrated that modeling variations in images in the transport space improves image reconstruction quality compared to modeling variations in the image space.

2.3.4 Pattern Recognition

The specific application of OT that is the focus of this thesis is pattern recognition. The mathematics of LOT described in Section1.3.4 have been used to transform images from a pixel-based (Eulerian, in PDE parlance) to location-based (Lagrangian, in PDE parlance) representation to

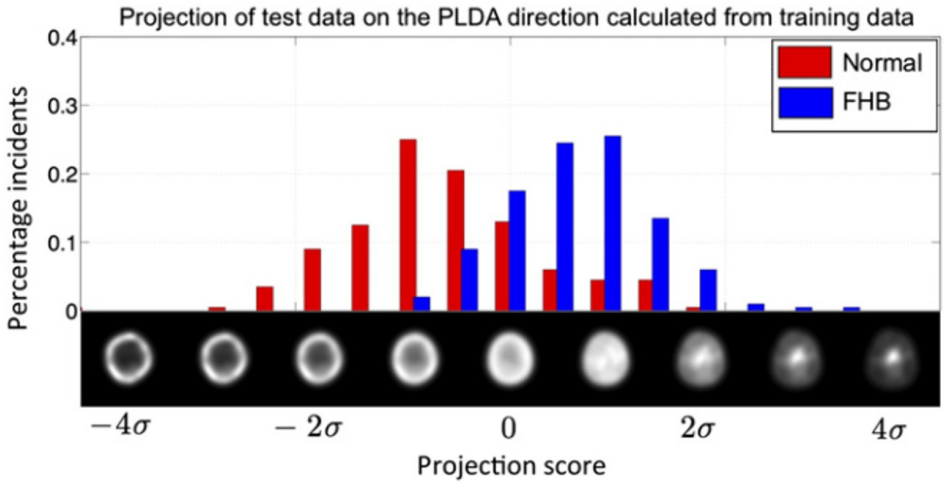


Figure 2.6: Source: [12]. Projection of data onto the direction in the LOT space that separates FHB cells from normals. We see that the LOT-based metric is able to model the nonlinear changes in cell appearance that increase the likelihood that it is malignant.

facilitate information extraction in the transform domain. In fact, separate 1D and 2D signal transformations for pattern recognition have been defined based on OT mathematics [80, 111]. Prior work has found that 2D microscopy images are generally more or equally separable in the LOT space, which will be henceforth abbreviated as the ”transport space” or ”transport domain” [12, 84, 111, 135, 136]. Furthermore, statistical learning in this space is more easily able to discover trends in the data and facilitates discrimination and regression tasks [12, 84, 111, 135, 136]. Because the LOT embedding is generative, any discriminant or regression direction computed in the transport space can be inverted to visualize shifts in the morphologic profile.

Using these principles, cell morphology associated with Fetal-type Hepatoblastoma (FHB) have been investigated in the transform space (see Section 1.3.4) in [12].

Although the LOT framework is a promising technique for automated discovery and visualization of structural markers, prior work on LOT is based on 1D and 2D signals. As discussed earlier in this chapter, the numerical solution of OT for large 3D images faces several challenges.

2.4 Summary

Although the OT problem is not new, it has many emergent applications in image analysis, especially in pattern recognition and super-resolution imaging. The LOT framework for quantifying and visualizing variations in sets of images is promising, although the prior work is based on 1D and 2D signals. The mathematics of LOT have never been extended to include large medical images, including 3D radiology images. Numerical OT for large images faces several challenges, especially for digital image datasets that can easily comprise millions of voxels in 3D. Some of these challenges include high computational complexity, especially in approaches that require solution of a discrete system of equations [60, 114] or introduce an additional time variable [16, 17]. For approaches based on gradient flow [2, 62] and other continuous-domain constraints [33], authors describe non-convergence on the solution path for digital images [62, 128]. Finally, approaches for shape matching and comparison [121] require many landmarks to be specified and matched.

Our goal is to address the aforementioned limitations of current OT solvers to enable broader application of Transport-Based Morphometry (TBM) to large medical imaging data, including radiology data.

Chapter 3

Multiscale Variational Optimal Transport

3.1 Introduction

Transport-Based Morphometry [12, 136], an image transformation framework based on the mathematics of LOT, can potentially address shortcomings of current techniques that are reviewed in Chapter 1. However, as described in Chapter 2, solving optimal mass transport for large digital medical images with multi-million voxels faces multiple challenges, including high computational complexity, non-convergence, and large numerical errors that result in poor, numerically non-invertible mappings. In this chapter, we present a new solution to the OT problem using variational optimization that can be solved using a series of simple gradient descent update equations. The main contribution of this work is a new technique for analyzing tissue distribution in large medical images. A novel mathematical formulation of the OT problem resulting in superior numerical solution for the first time enables application of TBM to large medical images, such as magnetic resonance imaging.

Applying the proposed OT approach to brain MRIs of healthy older adults, we see that the technique proposed enables superior results compared to existing OT techniques. Statistical learning using the proposed technique was able to discover aging-related brain morphology changes that are well-corroborated and well-accepted in the clinical literature on aging. In contrast, prior OT

approaches failed to generate mass-preserving (MP) mappings due to aforementioned problems. The proposed solver especially enables a broader set of applications involving 3D medical images.

This chapter is structured as follows. First, in Section 3.2, we review the equations for Transport-Based Morphometry for synthesis and analysis. In Section 3.3, we present our approach, including our modified formulation and multiscale solver that enables superior numerical results compared to existing methods. Section 3.6 describes experimental methodology. In Section 3.7, we present our results showing robustness of the proposed OT approach compared to prior work; we also see that Transport-Based Morphometry is able to accurately assess dependent brain morphology changes with age. Finally, in Section 3.8, we present discussion of the results and implications for future work.

This work underscores that there are potentially numerous new applications of TBM to study diseases for which the genotype-structural-behavioral associations are still unknown.

3.2 Signal Transformation Equations

By treating MRIs as smooth density functions, the similarity in tissue *spatial distribution* between two images can be quantified based on the L^2-Wasserstein distance. Any MRI modality that generates scalar intensity maps (i.e. T1-weighted, T2-weighted, FLAIR, fractional anisotropy, etc.) is amenable to analysis by the TBM framework.

Consider a set of magnetic resonance images $I_1, ..., I_N : \Omega \to \mathbb{R}^+$, corresponding to experimental subjects $i = 1, ..., K$, where $\Omega = [0, 1]^3$, the images are first intensity normalized to produce densities such that

$$\int_\Omega I_i(x)dx = 1. \tag{3.1}$$

A common reference image I_0, is chosen and an optimal transport mapping is calculated from the reference image to each MRI, I_i, for a total of K maps computed. Let $f_i : \Omega \to \Omega$ be a mass preserving mapping from I_0 to I_i. Then, the analysis equation [84] that transforms an image to its

corresponding representation in transform domain can be written based on

$$\begin{aligned} f_i^*(x) &= \underset{f_i \in \mathrm{MP}}{\arg\min} \int_\Omega |x - f_i(x)|^2 I_0(x) dx, \\ \text{s.t.} \quad & \det(Df_i(x)) I_i(f_i(x)) = I_0(x) \qquad \text{for } \forall x \in \Omega \end{aligned} \tag{3.2}$$

Here, D denotes the Jacobian matrix and MP is the family of all mass preserving mappings from I_0 to I_1. The problem in Equation (3.2) is well-posed and existence of a unique optimal solution $\mathbf{f}_i^*$ to above optimization was shown by Brénier [26]. Thus, the transformation is bijective.

The transport maps $f_i^*(x)$ are vector fields that define the direction and quantity of mass transport needed to match $I_i(x)$ to $I_0(x)$ with error zero. Then it can be shown that $\hat{I}_i(x) = (f_i^*(x) - x)\sqrt{I_0(x)}$ provides a natural isometric linear embedding for image I_i with respect to the LOT [84, 136]. This linear embedding is generative, thus, any geometric analysis in the LOT domain can be directly visualized in the image domain [84, 136].

An image corresponding to a given linear embedding can be recovered and visualized according to the synthesis equation [84]

$$\begin{aligned} I(x) &= det(Df^{-1}(x)) I_0(f^{-1}(x)) \\ &\text{where } f^{-1}(x) \text{ is the inverse mapping of } f(x) \end{aligned} \tag{3.3}$$

We note that once images have been transformed, Euclidean distances in the transport domain correspond to linearized versions of the transport metric on the OT geodesic [12, 84, 136]. This suggests that complex, nonlinear changes modeled by OT can be described by simple linear models in the transport domain, which will be explored further in Sections 3.5 and 3.7.

3.3 Formulation

In this section, we present our solution to the OT problem. This section is organized as follows. First, in subsection 3.3.1, we present a relaxed formulation of the Monge problem. Next, in subsection 3.3.2, we present gradient descent update equations derived using a variational optimization approach. Lastly, in subsection 3.3.3, we present a solver based on accelerated gradient descent in a multiscale scheme that enables superior numerical solution to the OT problem.

3.3.1 Variational formulation of the problem

In order to find the optimal transport map, we reformulate the minimization in (2.2) by relaxing the MP constraint. Assuming that Ω is a convex and connected subset of $\mathbb{R}^n$, as is the case for most image analysis problems (i.e. $\Omega = [0,1]^n$), and assuming that the probability measures μ and σ are atomless and absolutely continuous, we can write the differential counterpart of Equation (2.1) as,

$$det(Df(x))I_1(f(x)) = I_0(x), \quad \forall f \in MP \tag{3.4}$$

where $I_1(f(.)) = I_1 \circ f(.)$, D is the Jacobian matrix, and $det(.)$ denotes the determinant operator. The minimization in (2.2) for $c(x,y) = |x-y|^2$ can first be relaxed into the following optimization problem,

$$\begin{aligned} \arg\min_f \quad & \frac{1}{2}\int_\Omega |x - f(x)|^2 I_0(x) dx \\ s.t. \quad & \|det(Df)I_1(f) - I_0\|^2 \leq \epsilon \end{aligned} \tag{3.5}$$

for some small $\epsilon > 0$. Next, we use the result from Brénier's theorem which states that the optimal transport map is a curl free mass preserving map. Therefore we propose to modify the optimization

problem in Equation (3.5) by regularizing the objective function with the curl of the mapping, f,

$$\begin{aligned} \arg\min_f \quad & \frac{1}{2}\int_\Omega |x - f(x)|^2 I_0(x)dx + \frac{\gamma}{2}\int_\Omega |\nabla \times f(x)|^2 dx \\ s.t. \quad & \|det(Df)I_1(f) - I_0\|^2 \leq \epsilon \end{aligned} \tag{3.6}$$

where γ is the regularization coefficient and $\nabla \times (.)$ is the curl operator. We note that modifying (3.5) to penalize the objective with the curl does not change the optimal solution, but solving (3.6) in practice helps guide the solution toward the curl free map. We can relax the optimization problem above further and write it as a regularized (or penalized) unconstrained optimization problem,

$$\begin{aligned} \arg\min_f \frac{1}{2}\int_\Omega |x - f(x)|^2 I_0(x)dx \quad & + \quad \frac{\gamma}{2}\int_\Omega |\nabla \times f(x)|^2 dx \\ & + \quad \frac{\lambda}{2}\int_\Omega (det(Df(x))I_1(f(x)) - I_0(x))^2 dx \end{aligned} \tag{3.7}$$

Hence the formulation above contains terms explicitly signifying properties of MP mapping $\|det(Df)I_1(f) - I_0\|^2$ and a curl-free mapping $\|\nabla \times f\|^2$.

The optimization problem in Equation (3.7) is not a convex problem. We use a multiscale variational optimization technique to solve it. The multiscale scheme helps guide the solution toward the global optimum. We will see in the results section that the multiscale scheme is able to achieve solutions comparable to those obtained using convex methods when they apply. Section 2.4 describes the multiscale variational solver we devise for the optimization in (3.7).

3.3.2 Euler-Lagrange equations

The objective function in (3.7) can be written as,

$$M(f) \quad = \int_{\Omega} \mathcal{L}(x, f(x), Df(x))dx. \tag{3.8}$$

The Euler-Lagrange equations for the transport field f then are of the form,

$$\frac{dM}{df^i} = \frac{\partial \mathcal{L}}{\partial f^i} - \sum_{k=1}^{n} \frac{d}{dx^k}(\frac{\partial \mathcal{L}}{\partial f^i_{x^k}}), \quad i = 1, ..., n \tag{3.9}$$

where the superscripts denote the coordinate index for the vectors, and the subscripts denote partial derivatives, $f^i_{x^k} = \frac{\partial f^i}{\partial x^k}$. Writing the Euler-Lagrange equations for the objective function in (3.7) leads to,

$$\begin{aligned} \frac{dM}{df} = (f - id)I_0 \quad &+ \quad \lambda(det(Df)\nabla I_1(f) - \nabla \cdot (adj(Df)I_1(f)))I_{error} \\ &+ \quad \gamma(\nabla \times \nabla \times f) \end{aligned} \tag{3.10}$$

where $id(x) = x$ is the identity function, $adj(.)$ denotes the adjugate operator, $\nabla \cdot (.)$ is the divergence operator, and $I_{error} = det(Df)I_1(f) - I_0$. The derivation for the equation above is presented in the Section 3.4. Equation (3.10) is a key result from our formulation. The complexity for computing each gradient descent update step here is $\mathcal{O}(NlogN)$, where N is the number of pixels or voxels in the image. The computational complexity is determined by the cost of computing gradients $\mathcal{O}(N)$ in (3.10), but is dominated by the cost of cubic interpolation in computing the $det(Df)I_1(f)$ term.

3.3.3 Multiscale accelerated gradient descent

Local optimality is guaranteed for non-convex problems, and we can guide the solution toward the globally optimal solution by a multiscale scheme as depicted in Figure 3.1 based on gaussian pyramids. Nesterov's accelerated gradient descent method [96] is used at each scale to find the corresponding optimal transport map from Equation (3.7). The optimal transport map is then interpolated and used as the initial point for the accelerated gradient descent method in the next scale (finer scale).

The accelerated gradient descent update for k'th iteration ($k > 1$) at each scale is as follows,

$$\begin{cases} g^k = f^{(k-1)} + \frac{k-2}{k+1}(f^{(k-1)} - f^{(k-2)}) \\ f^k = g^k - \alpha_k \frac{dM(g^{(k)})}{df} \end{cases} \tag{3.11}$$

where α_k is the gradient descent step size, and is automatically chosen at each gradient descent update such that the maximum displacement is fixed. The update at k = 1 is the usual gradient descent update.

Compared to existing solvers, our multiscale approach is based on a relaxed formulation of the Monge problem and can avoid traditional problems of drifting, computational complexity, and artifact, as demonstrated in the experimental results.

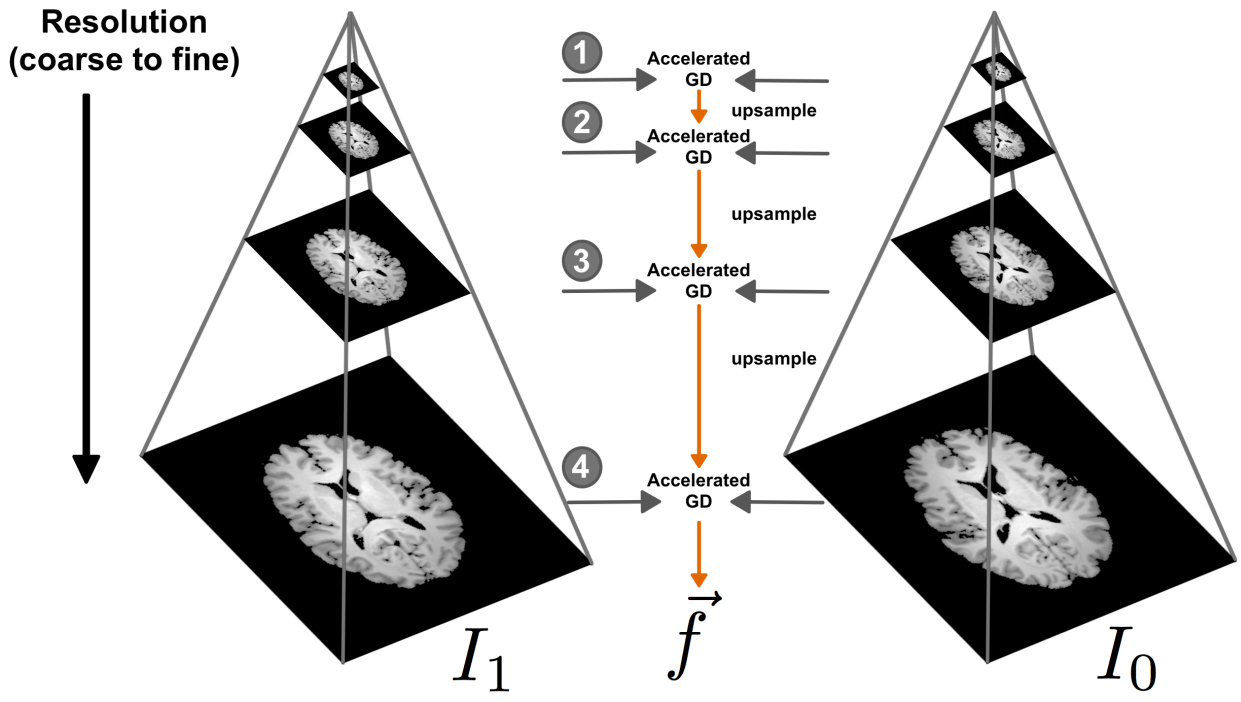


Figure 3.1: The schematic of the multiscale approach devised in this chapter. The solution to the accelerated gradient descent is first calculated at a coarse level and then refined as the optimization proceeds.

3.4 Derivation of Euler-Lagrange equations

Here we present the derivation of the Euler-Lagrange equation in (3.10). Starting from the objective function in Eq. (3.7) we have

$$\begin{aligned} M(f) &= \underbrace{\frac{1}{2}\int_\Omega |f(x)-x|^2 I_0(x)dx}_{M_1(f)} \\ &\quad + \underbrace{\frac{\lambda}{2}\int_\Omega (det(Df(x))I_1(f(x))-I_0(x))^2 dx}_{M_2(f)} \\ &\quad + \underbrace{\frac{\gamma}{2}\int_\Omega |\nabla \times f(x)|^2 dx}_{M_3(f)} \end{aligned} \tag{3.12}$$

where the first term, $M_1(f)$, is the transport cost, the second term, $M_2(f)$, enforces f to be mass preserving, and the third, $M_3(f)$, enforces f to be curl free.

The general Euler Lagrange equation for,

$$M(f) = \int_\Omega \mathcal{L}(x, f(x), Df(x))dx \tag{3.13}$$

is given as

$$\frac{dM}{df^i} = \frac{\partial \mathcal{L}}{\partial f^i} - \sum_{k=1}^{n} \frac{d}{dx^k}(\frac{\partial \mathcal{L}}{\partial f^i_{x^k}}), \quad i = 1, ..., n \tag{3.14}$$

Following the above equation for the first term, $M_1(f)$, we get,

$$\frac{dM_1}{df^i} = (u^i)I_0, \quad i = 1, ..., n \tag{3.15}$$

where $u = f - id$ is the displacement field, and $id(x) = x$ is the identity function.

For the second term we have,

$$\frac{\partial \mathcal{L}_2}{\partial f^i} = \lambda \det(Df)\frac{\partial I_1(f)}{\partial f^i} I_{error} \tag{3.16}$$

where $I_{error} = det(Df)I_1(f) - I_0$. Let C be the cofactor matrix of Df. Then $det(Df)$ can be written as the sum of the cofactors of any columns or rows of Df,

$$\begin{aligned} det(Df) &= \sum_{i=1}^{n} f^i_{x^j} C_{i,j}, \ \forall j \in \{1, ..., n\} \\ &= \sum_{j=1}^{n} f^i_{x^j} C_{i,j}, \ \forall i \in \{1, ..., n\} \end{aligned} \tag{3.17}$$

Using the cofactor matrix, C, we can write,

$$\frac{\partial \mathcal{L}_2}{\partial f^i_{x^k}} = \lambda C_{i,k} I_1(f) I_{error} \tag{3.18}$$

And from Equations (3.16) and (3.18) we have,

$$\frac{dM_2}{df^i} = \lambda((det(Df)\frac{\partial I_1(f)}{\partial f^i} - \sum_{k=1}^{n} \frac{d}{dx^k} C_{i,k} I_1(f)) I_{error}) \tag{3.19}$$

and writing the vector form of the above equation for all i and using $C^T = adj(Df)$ we can write,

$$\frac{dM_2}{df} = \lambda((det(Df)\nabla I_1(f) - \nabla \cdot (adj(Df) I_1(f))) I_{error}). \tag{3.20}$$

Finally for the third term, $M_3(f)$, we have

$$\frac{\partial \mathcal{L}_3}{\partial f^i} = 0 \tag{3.21}$$

Furthermore, assuming that $n = 2, 3$ and using the Levi-Civita symbol ϵ we can write the norm squared of the curl of f as follows,

$$|\nabla \times f|^2 = \sum_{p=1}^{n} (\sum_{l=1}^{n} \sum_{m=1}^{n} \epsilon^{plm} f^m_{x^l})^2 \tag{3.22}$$

which leads to,

$$\frac{\partial \mathcal{L}_3}{\partial f^i_{x^k}} = \gamma \sum_{p=1}^{n} \epsilon^{pki} (\sum_{l=1}^{n} \sum_{m=1}^{n} \epsilon^{plm} f^m_{x^l}). \tag{3.23}$$

Therefore we have,

$$\begin{aligned}
\frac{dM_3}{df^i} &= -\gamma \sum_{k=1}^{n} \frac{d}{dx^k} (\sum_{p=1}^{n} \epsilon^{pki} (\sum_{l=1}^{n} \sum_{m=1}^{n} \epsilon^{plm} f_{x^l}^m)) \\
&= -\gamma \sum_{k=1}^{n} \sum_{p=1}^{n} \epsilon^{pki} \frac{d}{dx^k} (\sum_{l=1}^{n} \sum_{m=1}^{n} \epsilon^{plm} f_{x^l}^m) \\
&= \gamma \sum_{k=1}^{n} \sum_{p=1}^{n} \epsilon^{ikp} \frac{d}{dx^k} (\sum_{l=1}^{n} \sum_{m=1}^{n} \epsilon^{plm} f_{x^l}^m) \\
&= \gamma (\nabla \times \nabla \times f)^i
\end{aligned} \tag{3.24}$$

Finally, combining Equations (3.20) and (3.24) will lead to,

$$\begin{aligned}
\frac{dM}{df} = (f - id) I_0 + \lambda (det(Df) \nabla I_1(f) \\
-\nabla \cdot (adj(Df) I_1(f))) I_{error} \\
+\gamma \nabla \times \nabla \times f
\end{aligned} \tag{3.25}$$

3.5 Modeling shape and appearance of the brain

Previous work has established that treating signals as densities increases their separability in the transport domain [12, 80, 84, 111, 135, 136]. Recall that Euclidean distances in transport domain correspond to linearized versions of the transport metric on the OT manifold. Here, we describe a framework for linear regression, discrimination, and blind signal separation in the transport space.

The data matrix $X \in \mathbb{R}^{d \times K}$ stores the vectorized transport maps x_m corresponding to each subject $m \in 1, ..., K$ where d is the number of elements in the vectorized transport map and K is the number of subjects.

3.5.1 Regression and correlation analysis with a clinical variable

The influence of an independent clinical variable $v \in \mathbb{R}^{K\times 1}$ on brain tissue distribution can be investigated by computing the direction in the transport domain w_{corr} such that the linear correlation with age is maximized according to (3.26) [12].

$$w_{corr} = arg\max_{w} \frac{w^T X v}{\sqrt{w^T w}} = \frac{Xv}{\sqrt{v^T X^T X v}} \tag{3.26}$$

Here, the direction $w = \bar{x} + \nu w_{corr}$ is a vector field that represents the direction and magnitude by which tissue is re-distributed due to v and ν represents the increment or decrement to sample along the maximally correlated direction.

The images corresponding to the computed direction w can be visualized through inverse TBM transformation by Equation (3.3) and illustrate the morphology that is associated with outcome v.

3.5.2 Discriminant analysis to differentiate groups of subjects

Another class of problems facilitated by the TBM technique is that of discriminating two groups based on MRI appearance, such as the one posed in Figure 3.2. For these problems, penalized linear discriminant analysis (PLDA) [134] performed in the transport domain can find the direction in transport space that maximally separates C classes. The PLDA direction is given by (4.3)

$$w_{PLDA} = arg\max_{||w||=1} \frac{w^T S_T w}{w^T (S_W + \alpha I) w} \tag{3.27}$$

where $S_T = \frac{1}{M}\sum_m (x_m - \bar{x})(x_m - \bar{x})^T$. Here, $\bar{x} = \frac{1}{M}\sum_{m=1}^{M} x_m$.

The within-class scatter matrix is $S_W = \sum_C \sum_{n\in C} (x_n - \bar{x_c})(x_n - \bar{x_c})^T$. The parameter α controls the tradeoff between the traditional linear discriminant analysis (LDA) direction and one that lies in the principal component analysis (PCA) subspace. The parameter α can be chosen by plotting the stability of the subspace as a function of α.

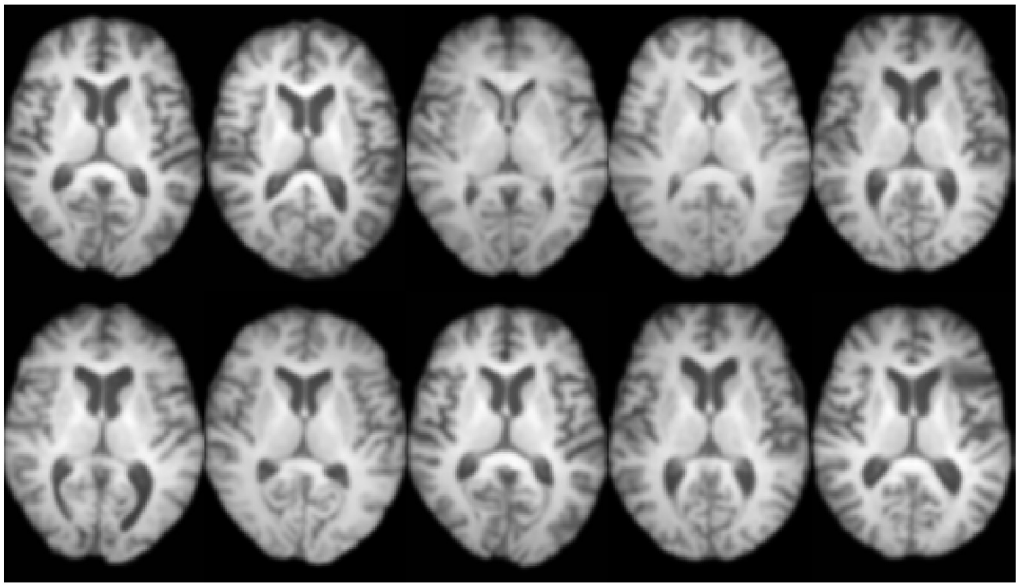


Figure 3.2: MR images belonging to 10 older adults in their $6th$ or $7th$ decades of life. The images correspond to subjects who are either 2σ above or below the mean aerobic fitness as assessed by vO_2 L/min. Is there a common morphologic feature that separates these groups? If so, what is it?

Sampling along and inverting the direction w_{PLDA} yields images showing the typical morphology of a class and how it changes as one progresses from one class to another.

3.5.3 Visualizing principal phenotypic variations in the brain

Given the covariance matrix S_T defined in Section 3.5.2, the principal components are given by the eigenvectors of S_T and can be used to explain the main modes of variability in the dataset [12].

The factorization in Equation (3.28) gives both the principal components and eigenvalues, where the diagonal components of Σ represent the variance for each principal component.

$$S_T = U\Sigma U^T \tag{3.28}$$

Each principal component can be inverted and visualized to yield the principal phenotypic variations that comprise the images in the dataset.

3.6 Computational Experiments

Here we describe image acquisition, pre-processing, optimal transport minimization, and statistical learning steps. The code was prototyped in MATLAB (MathWorks, Natick, MA) using built-in libraries.

3.6.1 Datasets

Comparing MP registration methods

Healthy, adult brain images were obtained from the IXI dataset, Biomedical Image Analysis Group at the Imperial College in London [72].

10 images were selected at random from Guy's Hospital in UK. Subjects were male and ranged from 41 to 86 years of age at the time of imaging (mean age 57.8 years, standard deviation 15.7 years). The images were T1-weighted images, obtained using a Philips Medical Systems Intera 1.5 T scanner, with the following imaging parameters for T1. Repetition time = 9.813, echo time = 4.603, number of phase encoding steps = 192, echo train length = 0, reconstruction diameter = 240, flip angle = 8. The images are 128 x 128 x 128, with 1 mm^3 resolution.

Modeling healthy aging in the brain

Images of 135 healthy subjects were obtained, ranging in age from 58 to 81 years (mean age 66.6 years, standard deviation 5.9 years). Both male and female subjects are included. T1-weighted brain images were collected using a 3D Magnetization Prepared Rapid Gradient Echo Imaging (MPRAGE) protocol with 144 contiguous slices. Images were acquired on a 3 T Siemens Allegra scanner with repetition time = 1,800 ms, echo time = 3.87 ms, field of view (FOV) 256 mm, and acquisition matrix 192×192 mm , flip angle = 8 [47].

These images provide an expanded dataset of older subjects on which age-related brain mor-

phology can be investigated.

3.6.2 Multiscale variational optimal transport

Image preprocessing

Images were skull-stripped and affinely registered to a common template using Statistical Parametric Mapping (SPM) software version 12 [52]. Images were normalized so that the sum of intensities was equal in both images (equal *mass*). By normalizing to a large positive number, 10^6, numerical precision errors resulting from computations with small numbers were avoided. We also add a small constant 0.1 to the normalized images and renormalize so that they are strictly positive [33] to satisfy OT conditions.

Optimization parameters

The step size for accelerated gradient descent is chosen such that the maximum displacement per update is 0.01 of a pixel, the same for our method and comparison methods. At every step of gradient descent, there is a check to maintain that the mapping is diffeomorphic and the step size is reduced as necessary in order to ensure a diffeomorphism. The parameters were obtained experimentally: $\lambda = 100$, $\gamma = 6.5 \times 10^4$ when the MSE reaches 25% of the initial MSE to steer the solution towards a curl-free MP mapping, and number of scales = 3. The multiscale approach was also implemented for the two comparison methods in this chapter, although the results did not significantly change when the scheme was used.

The termination criteria for all methods implemented is when MSE of the morphed source image relative to the template image reaches 0.55%. When the drifting phenomena in the Haker method [63] produces MSE 100x the initial MSE, we forcibly terminate the code. We report the mean L_2 norm of the curl per voxel and MSE relative to the template.

We use a numerical discretization scheme in which values are placed at pixel or voxel centers.

A consistent second-order finite difference approximation was used for all differential operators, utilizing the DGradient toolbox for MATLAB [90].

3.6.3 Experiment 1: Comparing MP registration methods

We compare the approach of Haker *et al.*[63] and Chartrand *et al.*[33] to the approach described in this chapter. The computational complexity of the gradient descent update step for all three methods implementd is $\mathcal{O}(NlogN)$. All methods implemented utilized the same preprocessed images.

All unique pairs of images were registered to each other for the 10 images, resulting in 45 total registration problems. The statistics reported in this section are based on the registrations performed in turn with our method, that of Chartrand *et al.*, and that of Haker *et al.*

3.6.4 Experiment 2: Modeling the effects of aging on brain tissue distribution

Regression analysis was performed to assess the relationship between brain tissue distribution and age using the approach in Section 3.5.1. The common reference image I_0 was computed by the Euclidean average of all the subjects. Statistical significance of the computed direction is assessed using permutation testing with T = 1000 tests.

The reults of regression analysis in the transport space were compared to those obtained using deformation-based analysis. Images were skull-stripped and the DARTEL [5] toolbox in SPM12 was used to compute deformation fields. DARTEL is commonly used to perform standard VBM analysis. Images were smoothed ($\sigma = 1$) and normalized similar to the OT procedure following non-rigid registration by DARTEL. The most correlated direction was computed on the raw pixel intensities of DARTEL-registered images using Equation (3.26).

3.6.5 Experiment 3: Assessing the effects of aerobic fitness on brain health

Aerobic fitness is measured by vO_2 L/min in this study and is assessed based on a graded maximal exercise test. The individuals were grouped into low fitness and high fitness groups based on those who had a vO_2 L/min greater than one standard deviation above the mean (high fitness: n = 22) and lower than one standard deviation below the mean (low fitness: n = 16). The PLDA approach in transport space is used in order to learn discriminating differences between high fitness and low fitness groups, which is described in Section 3.5.2.

3.6.6 Experiment 4: Visualizing principal phenotypic variations

Unsupervised learning using PCA was performed. The images corresponding to the top three eigenvectors in transport space were visualized to understand the principal phenotypic variations using the approach described in Section 3.5.3.

3.7 Results

We compare our approach to current methods in the literature based on both the Monge formulation (Haker *et al.*[63]) and Kantorovich formulation (Chartrand *et al.*[33]).

3.7.1 Comparing MP registration methods

We report results for both 2D and 3D MP registration using optimal transport. The 2D image dataset was derived from the 3D dataset by extracting the same axial slice from the middle of every 3D brain image. The solver and accelerated gradient descent update equations are the same whether working with 2D or 3D images.

2D optimal mass transport

Table 3.2 displays the mapping statistics for the set of 2D images. The mass transported is lowest for the Chartrand *et al* method and compares favorably to the mass transport achieved using our method. The MSE for Chartrand *et al* indicates that this method also produces the poorest MP mapping of the three methods. The MSE achieved is 3-4 times that achieved by the other two methods.

Our method achieves the lowest MSE in addition to mass transport distance. As reviewed in Section 2.2, the optimal MP mapping is the MP mapping that achieves minimum mass transport.

The curl, a measure of optimality of the MP mapping, is zero by design for the method of Chartrand *et al* as expected (see Section 2.2). The mean curl is the highest for our method compared to the other two methods, but is still small in an absolute sense (10^{-4}).

Our method produces the best results in terms of mass-preservation and mass transport. In Figure 3.3, we see the optimal transport fields and their corresponding morphings for several 2D brain images. Visually, the transport fields and quality of morphings are similar for all three methods.

The proposed method was prototyped in MATLAB using built-in functions. The average runtime for 256×256 brain images with 3 scales was 19.36 ± 7.91 seconds.

3D optimal mass transport

While in 2D all three methods seem to produce visually similar OT mappings, we see in 3D that several phenomena become evident. Examining Figure 3.5, we compare the plots of curl and relative MSE over gradient descent iterations for several brain images. We see that the magnitude of the curl (on the order of 10^6) is large for the Haker *et al* method. The curl for the Chartrand *et al* method remains at zero by design. Our method produces curls for all images that tend toward zero with iterations of gradient descent.

We see that relative MSE with the Haker *et al* approach increases significantly until we forcibly

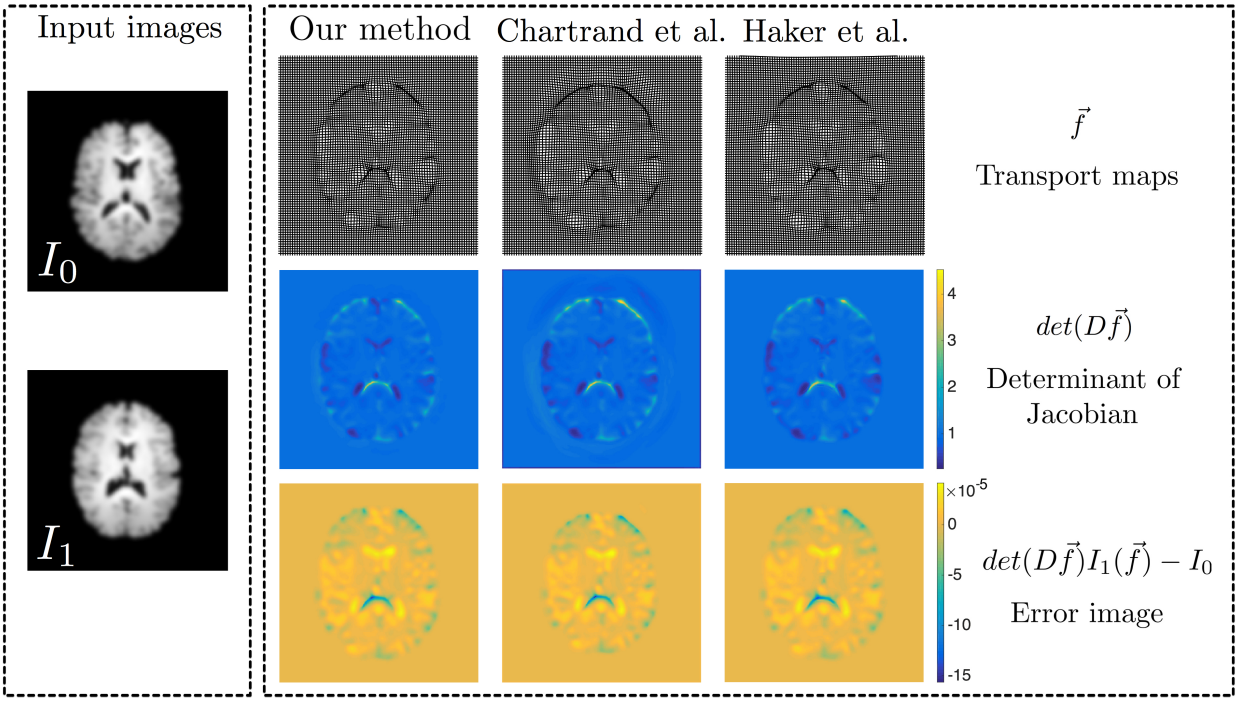


Figure 3.3: The source image I_1, target image I_0, and their calculated optimal transport map $\vec{f}$, corresponding determinant of Jacobian matrices, and the error image for the Haker method, the Chartrand method and our method. All are comparable for 2D OT

terminate the code when the MSE reaches 100x its initial value. Hence, starting after around 100 iterations of gradient descent, the phenomenon of *drift* with Haker *et al.*'s approach becomes evident. The relative MSE of Chartrand *et al.*'s approach decreases, but remains large in magnitude (5-10%) at termination. The large MSE results in visual artifacts in the quality of the MP match, which we can see in Figure 3.4. In contrast, for our method, all images are able to achieve the 0.55% termination criterion.

Table 3.2 corroborates the plots in Figure 3.5. Our method produces the lowest relative MSE (best MP mapping), and all brain images are able to achieve the termination criterion of 0.55%. Furthermore, our curl at termination is 8 orders of magnitude lower than that obtained using the Haker method.

In terms of mass transported, the Chartrand *et al* method produces the lowest transport distance, although the MSE of the MP mapping is about 5-10x higher than that achieved using our method. We can also see artifacts visually in the mappings produced by the Chartrand *et al* method compared to our method (Figure 3.4).

In Figure 3.4, we compare axial, sagittal and coronal slices mapped using our method and that of Chartrand *et al* (The method of Haker *et al* failed to produce a viable solution, which is why it is not shown.) We see that mappings produced by our method result in visually similar images to the target image I_0, whereas those produced by the Chartrand *et al* method contain several artifacts.

Overall, our method outperforms both comparison methods for 3D images. Our method achieves the lowest MP mapping, while at the same time achieving small curl and mass transported.

The median runtime was under 20 minutes per brain in MATLAB using built-in libraries on a general purpose computer. There is significant opportunity for improvement with an implementation in native C.

Thus, we see that our approach is able to overcome traditional limitations of drift, artifact, and computational complexity. Our approach is able to solve numerical OT for MRI data, unlike most traditional OT solvers. Numerical OT is an active area of research and the emerging OT

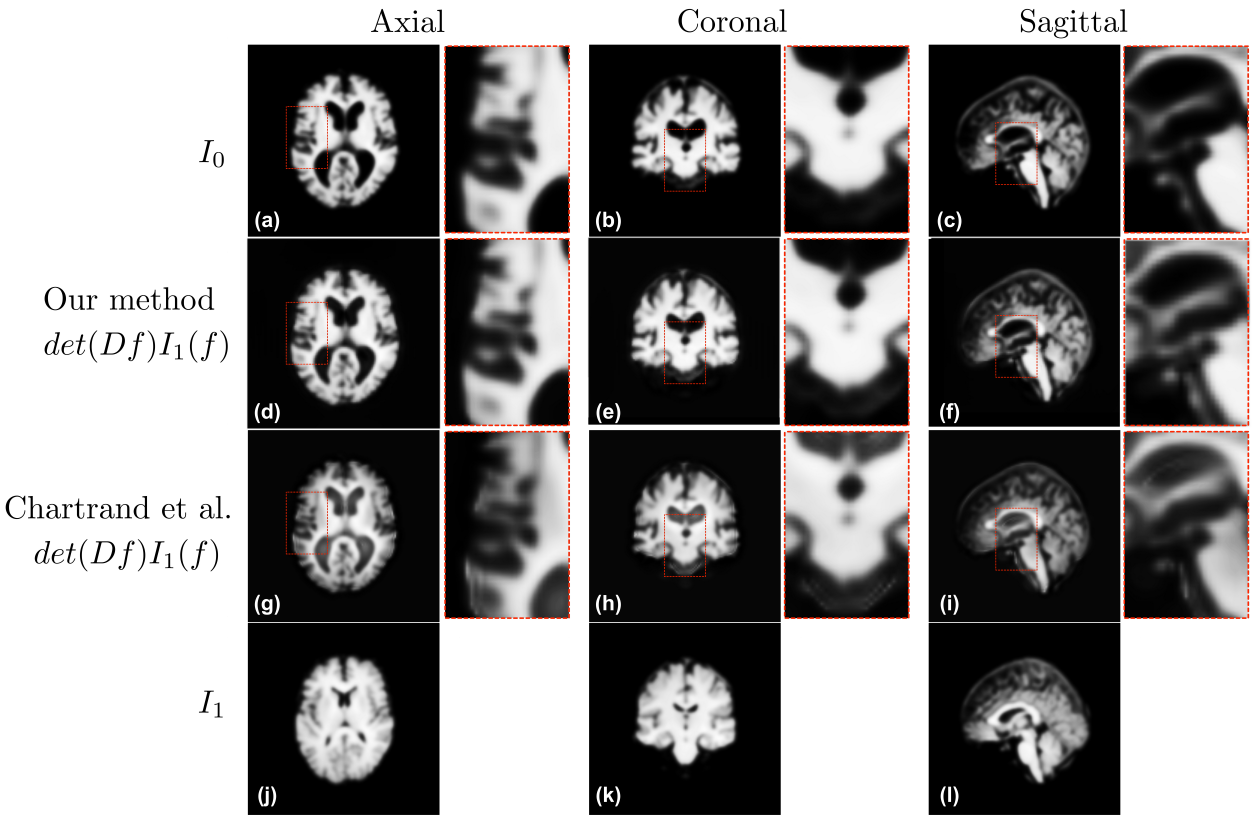


Figure 3.4: The target image (a-c), the morphed image in axial, coronal and sagittal cuts using our method (d-f) and the method presented by Chartrand *et al.*[33] (g-i), and the source image (j-l)

literature may highlight other potentially feasible approaches in addition to ours that can enable Transport-Based Morphometry to be extended to MRI data.

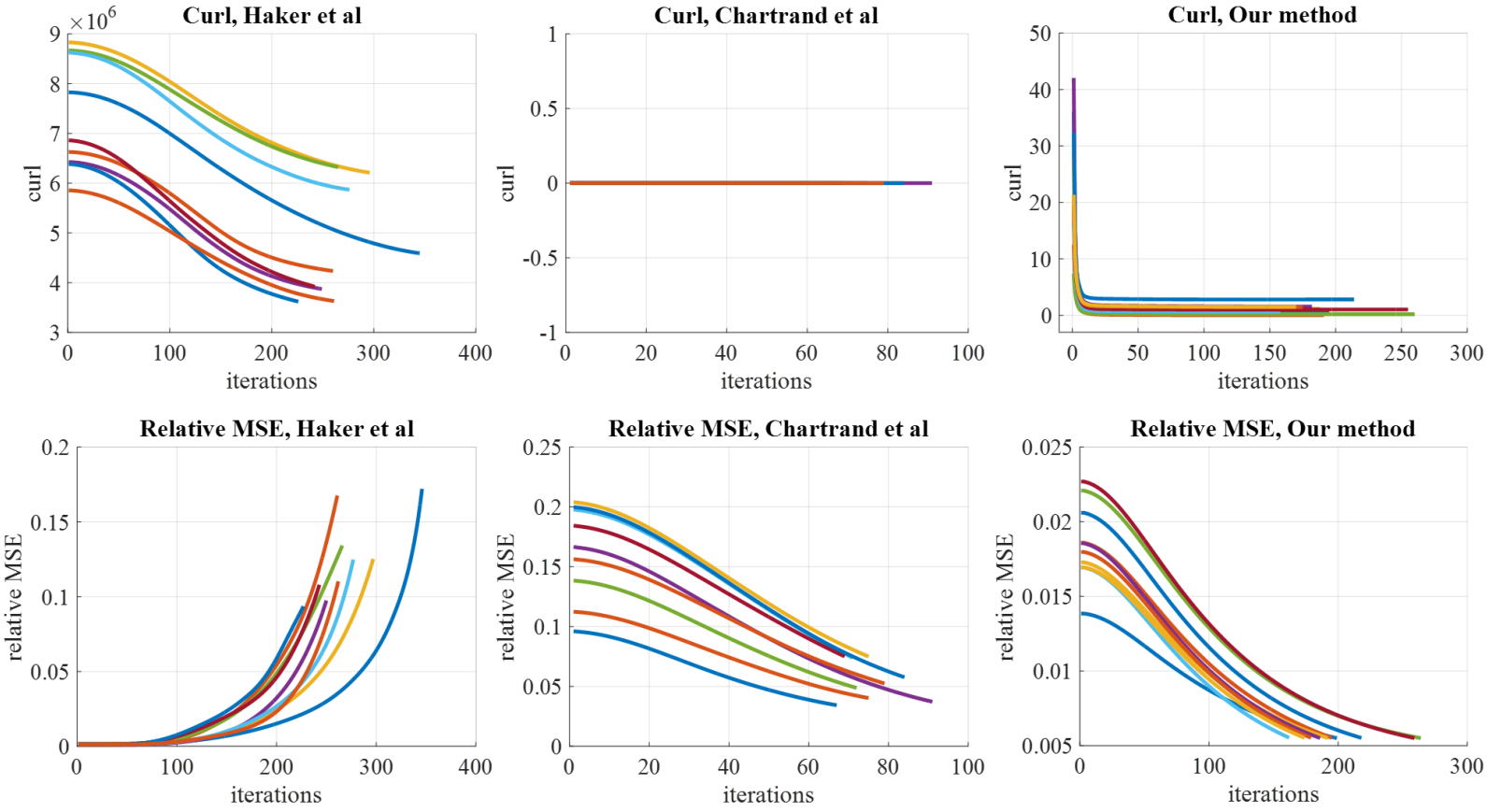


Figure 3.5: We see the plots for MSE, curl and mass transported for all three methods. The plots for our method are shown only for the last scale of the GP, using an initial point already close to the final point.

Table 3.1: Comparing methods of solving OT in 2D

2D	**OT Mapping Statistics**		
Method	*Relative MSE*	*Mean curl*	*Mass transported*
Our method	$0.23 \pm 0.056\%$	$(8.7 \pm 5.3) \times 10^{-4}$	1.63 ± 0.57
Chartrand *et al.*	$1.8 \pm 2.9\%$	0	1.56 ± 0.46
Haker *et al.*	$0.45 \pm 0.59\%$	$(7.0 \pm 0.16) \times 10^{-6}$	2.37 ± 0.79

Table 3.2: Comparing methods of solving OT in 3D

3D	**OT Mapping Statistics**		
Method	*Relative MSE*	*Mean curl*	*Mass transported*
Our method	$0.55 \pm 0.0011\%$	0.37 ± 0.61	1.3 ± 0.50
Chartrand *et al.*	$3.0 \pm 1.8\%$	0	0.07 ± 0.04
Haker *et al.*	$9.9 \pm 2.5\%$	$(4.5 \pm 1.6) \times 10^{6}$	12.5 ± 4.8

3.7.2 Modeling the effects of aging on brain tissue distribution

Figure 3.7 shows the fraction of variance captured by principal components of the image domain (raw voxel values after affine registration) compared to DARTEL registration and optimal mass transport solved using the described approach. Fewer components are needed to represent more of the variance in the transport space than either for image domain or DARTEL pixel-wise comparison. Therefore, the information about variability in the dataset appears to be better captured by examining tissue distribution using OT rather than comparing raw pixel values individually, before or after nonrigid registration.

The direction maximally correlated with age computed in the transport space is found statistically significant (Pearson's r = 0.4605, $p < 0.001$). Figure 3.8a shows the data when it is projected onto the maximally correlated direction, with each datapoint representing a subject's image.

The most correlated direction shown in Figure 3.8a can be inverted to visualize the dynamic morphology underlying age-related differences, which is shown by Figure 3.8b (images are colorized to aid visual interpretation). We see that the changes captured by the TBM regression framework are well-corroborated by known changes in the clinical literature. The changes we see

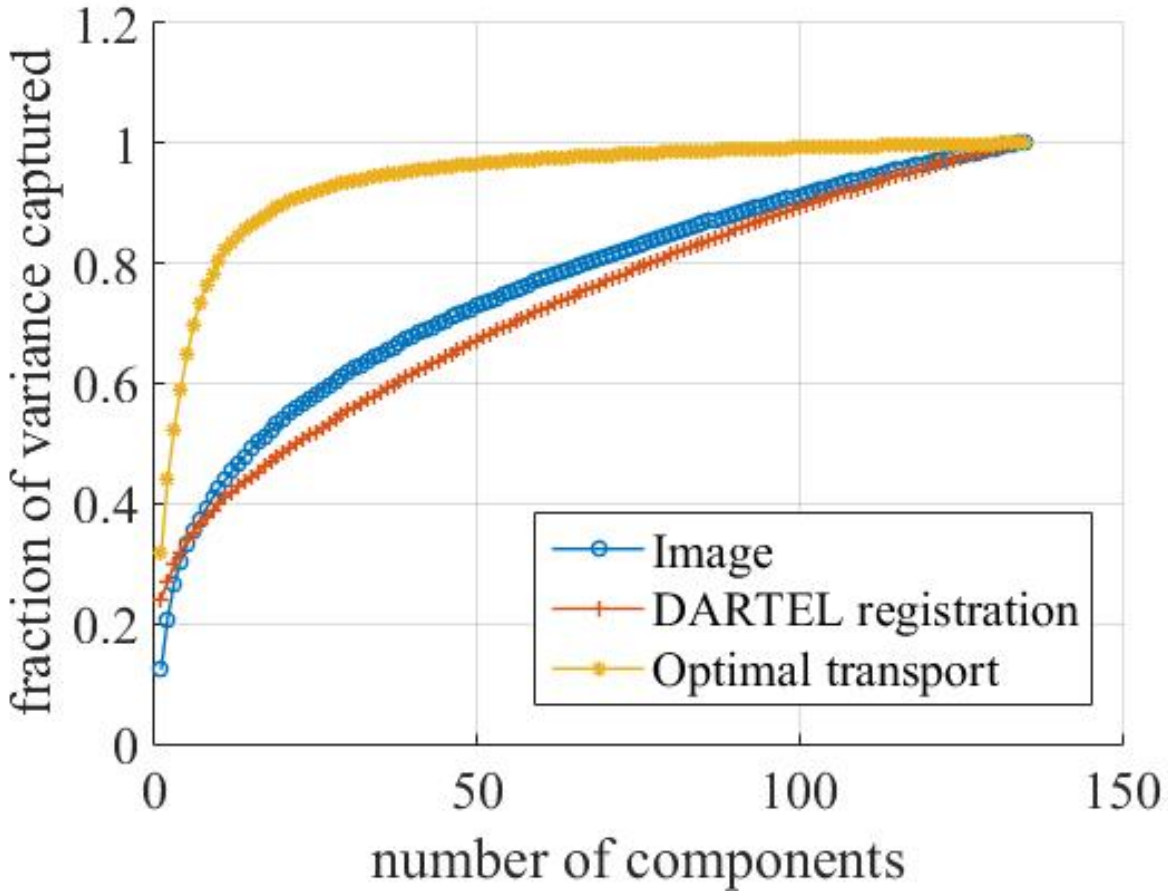
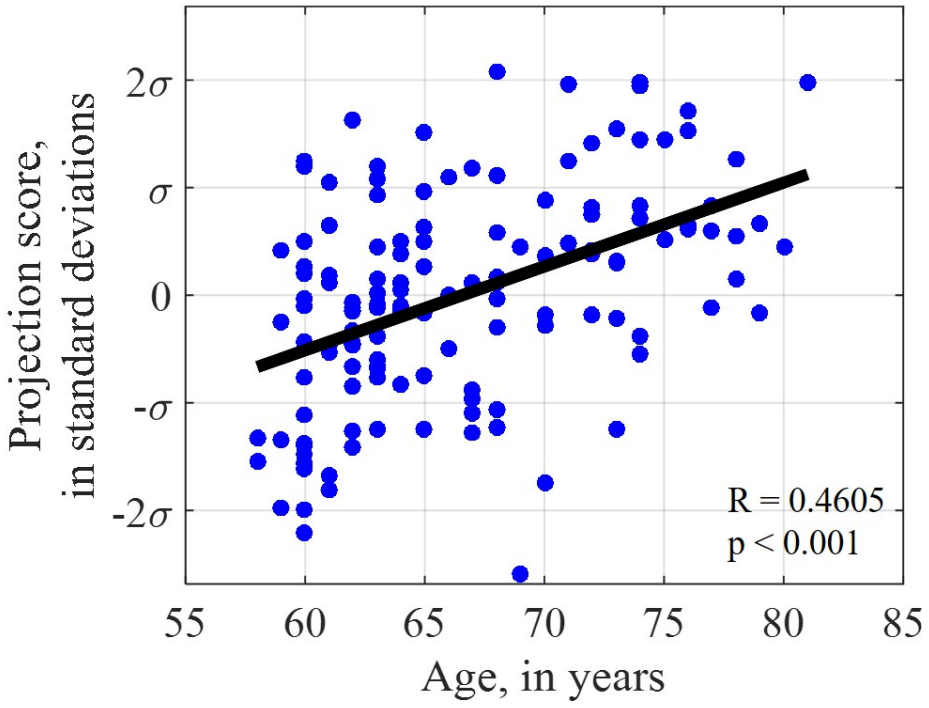
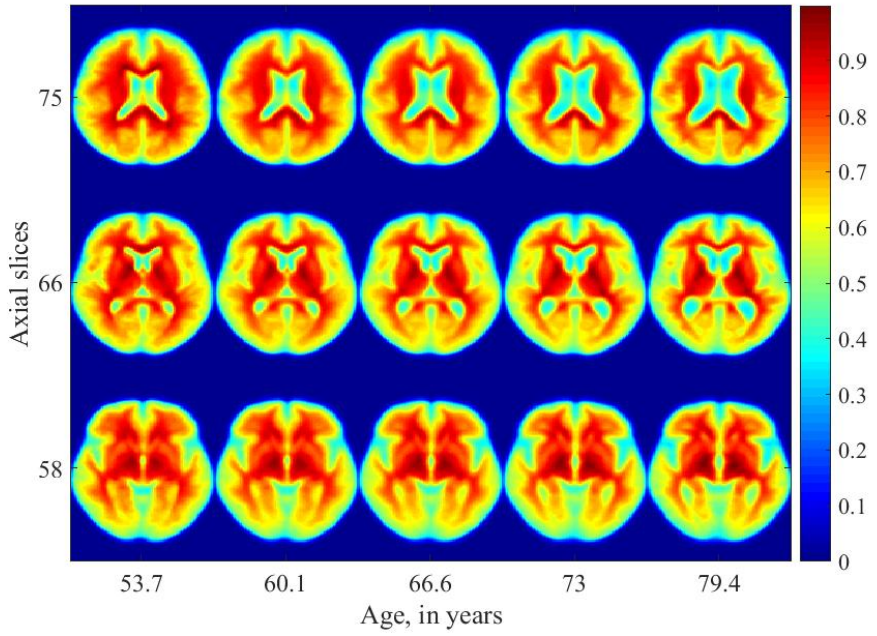


Figure 3.7: Compared to the models utilizing pixel-wise comparison (Eulerian and DARTEL registration), the model based on OT is able to capture more of the variability in the dataset with fewer principal components.



(a)



(b)

Figure 3.8: (a) Projection of data onto the direction that maximizes linear correlation with age, (b) Visualization of changes in tissue distribution that are statistically dependent on age. The vertical axis shows various axial slices from a 3D dataset from rostral (towards head) to caudal (towards toe). The horizontal axis shows the effect of increasing age from left to right on that axial slice. We see enlarging ventricles, and global atrophy of both gray matter and white matter with increasing age.

here are enlarging ventricles and global gray matter/white matter thinning, validating that analysis by TBM can discover and visualize biologically meaningful relationships in a fully automated manner.

In contrast, regression analysis performed on the raw pixels registered by DARTEL was unable to find a significant relationship between brain morphology and age (Pearson's r = 0.6196, p = 0.1700). The DARTEL toolbox is commonly used in the standard VBM analysis pipeline. Figure 3.9 shows the images generated by attempting to fit a regression model on individual pixel values. The changes related to aging are not well visualized. Focusing on the differences identified, Figure 3.6 shows the heat map corresponding to the change captured by VBM, showing which isolated pixels are used in the model, and the magnitude of their linear increase/decrease with age. Figure 3.6 shows the change in voxel intensity values with an approximately 5 year increase in age from the mean. Overall, it appears that aging-related changes are better captured by a tissue distribution analysis than by comparing individual pixels.

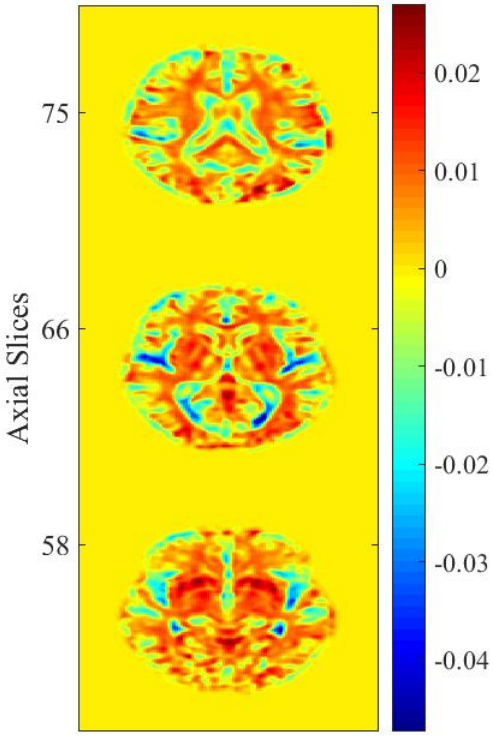


Figure 3.6: Heat maps showing the voxels whose intensity levels are linearly correlated as a function of changes in age.

3.7.3 Assessing the effects of aerobic fitness on brain health

Using the PLDA approach for discriminant analysis can separate high fitness from low fitness individuals.

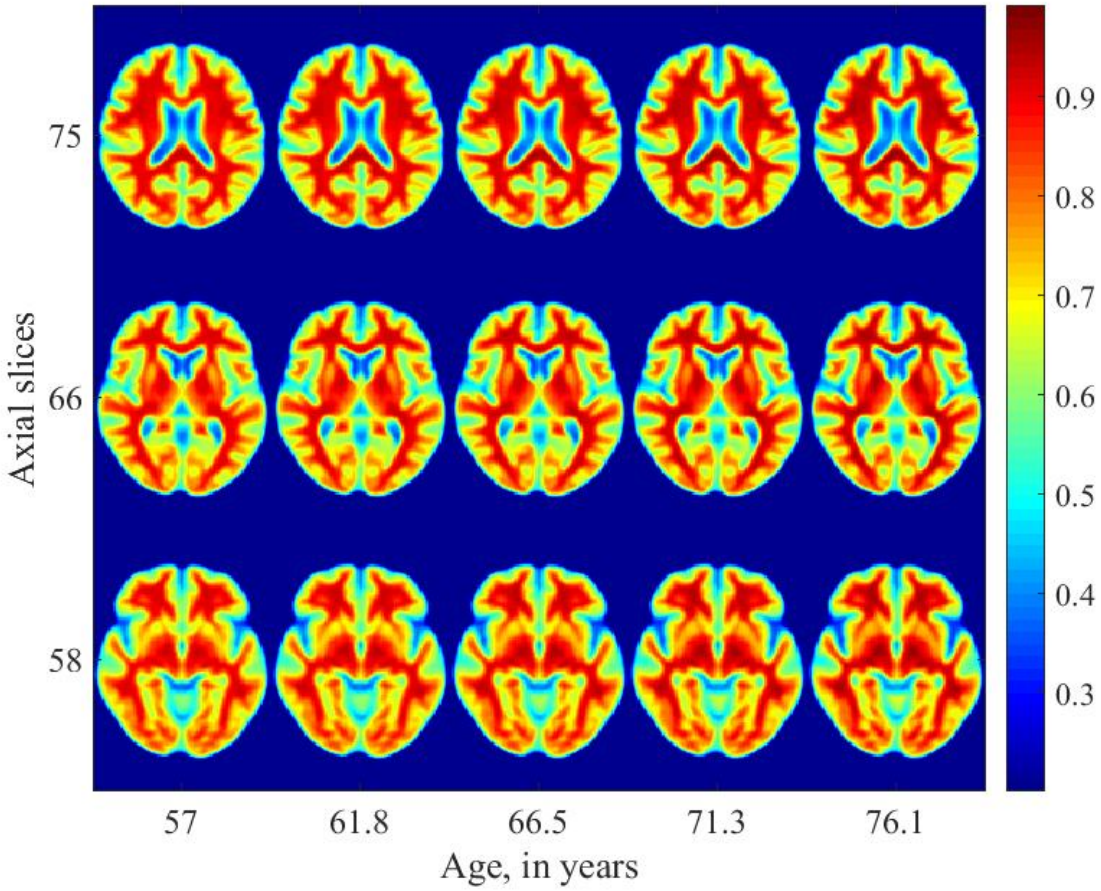
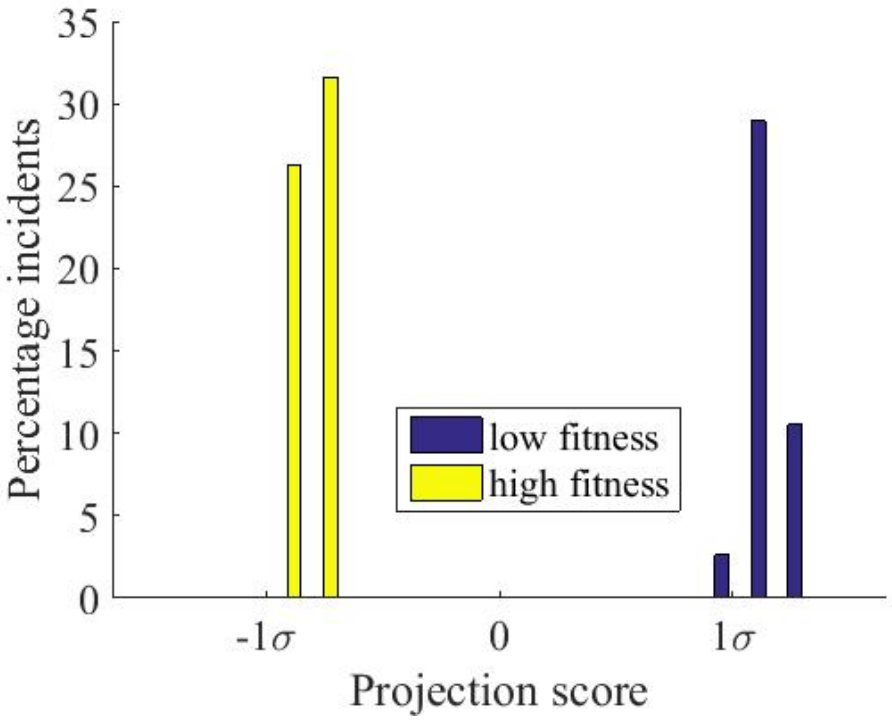
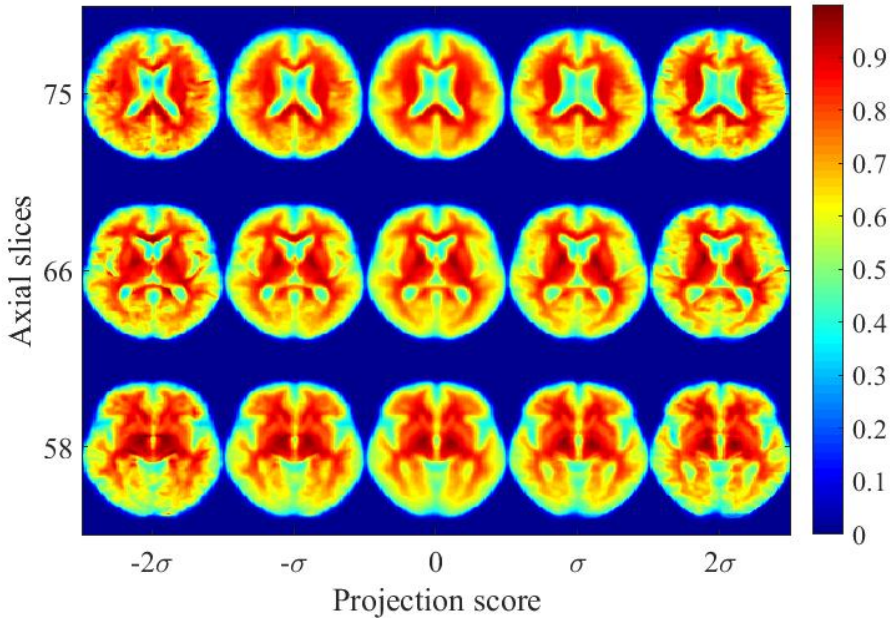


Figure 3.9: Visualization of aging-related changes captured by pixel-wise comparison after DARTEL registration, a common practice in Voxel-Based Morphometry.



(a)



(b)

Figure 3.10: (a) High fitness and low fitness individuals can be perfectly separated based on their images alone when images are projected onto the most discriminant direction computed by PLDA, (b) images illustrating the differences between high fitness and low fitness individuals.

Most importantly, inverting the computed discriminant direction using TBM inverse transformation shows that along different projection scores, brains corresponding to low fitness individuals appear to demonstrate changes in tissue distribution that are visually similar to those due to advancing age in Figure 3.8b. This represents a new finding for more detailed investigation in larger studies. We note that the measure of aerobic fitness used, vO_2 L/min is an absolute measure that is not corrected for body weight. Future studies will also need to investigate aerobic fitness when corrected for weight. Therefore, the role of fitness in brain health later in life can be assessed using the TBM linear discriminant framework to reveal new information about discriminating differences. Future studies using the fully automated approach are needed to build on these preliminary findings to yield important biological insights.

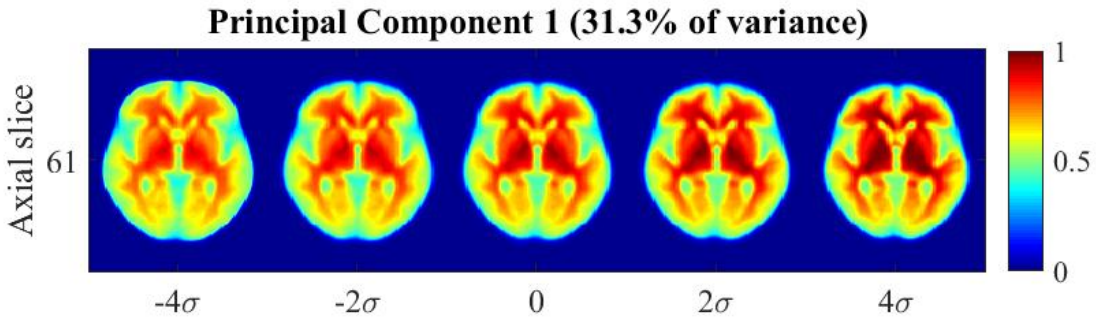
3.7.4 Visualizing principal phenotypic variations

We can visualize the top three PCA directions to gain a sense of the principal modes of variation in the dataset, which capture variability in brain size, level of tissue atrophy, and prominence of midbrain structures shown by Figure 4.8. Therefore, TBM can capture spatially diffuse patterns of change that lend to biological interpretation, unlike methods that rely on pixel-wise comparison.

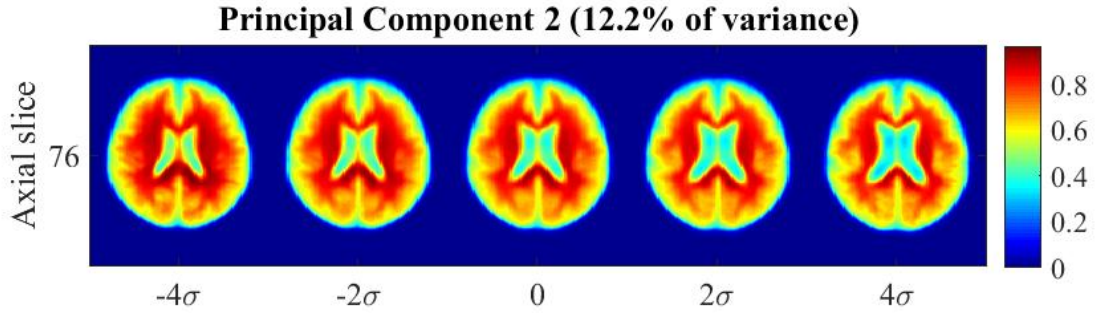
The TBM transformation framework, coupled with tools for statistical learning, can discover trends in MRI data with regard to regression with a clinical variable, discrimination of classes, and unsupervised learning. Analyzing tissue distribution enables modeling and discovery of phenomena previously not possible with existing techniques.

3.8 Conclusion

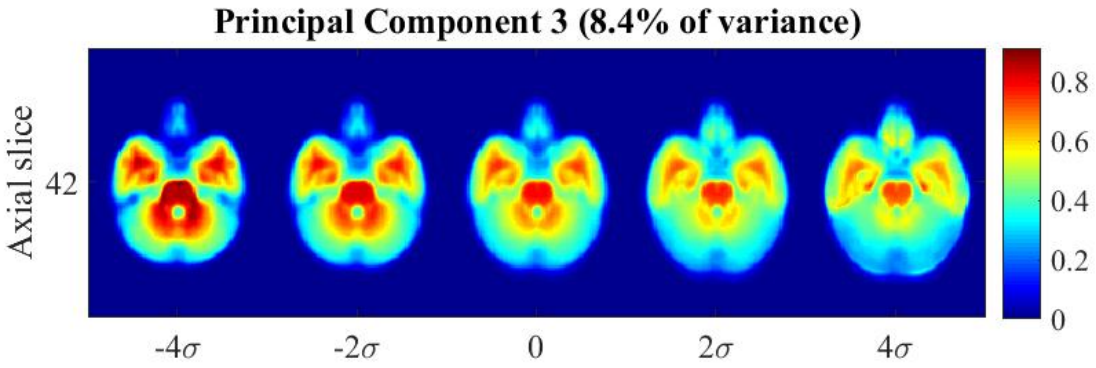
We describe a fully automated MRI analysis technique that facilitates discovery of trends from patient images. Unlike traditional approaches, our approach examines the images in a transform domain where the variations are more directly described by measuring differences in tissue spa-



(a)



(b)



(c)

Figure 3.11: (a) Captures variability in brain size, (b) Captures variability in brain tissue atrophy and size of ventricles, (c) Captures variability in prominence of midbrain and brainstem structures

tial distribution using OT. The concept of transforming images to a transform domain for pattern recognition is based on prior evidence that many 1D and 2D signals become more separable when they are transformed using OT [12, 84, 111, 135, 136]. Here, we extend the mathematics of LOT-based transformation to include 3D radiology data for the first time. This is the first application of TBM to any radiology problem, enabling automated discovery and visualization of structural trends. The results confirm our hypothesis that tasks of regression, discrimination, and signal separation are facilitated in the transform domain. Furthermore, Euclidean distances in the transport domain correspond to linearized versions of the transport metric on the OT geodesic, and we indeed see that linear trends in the transport domain can capture complex, nonlinear changes in tissue distribution. Our approach is able to correctly discover morphologic changes due to aging that are well-corroborated by the aging literature, unlike traditional pixel-wise comparison techniques. Additionally, our technique is able to identify that there are morphological differences between high fitness and low fitness groups and principal phenotypic variations that are biologically interpretable, demonstrating its use for discrimination and unsupervised learning. The TBM transformation is bijective and fully invertible, closing an important gap in computer-aided pattern recognition and modeling. For the first time, Transport-Based Morphometry can yield visualization of the dynamic changes in morphology that underlie various clinical outcomes through inverse transformation. Thus, TBM is a powerful tool to enhance scientific understanding of the mechanisms underlying clinical observations or identify image biomarkers for further investigation in future clinical trials. Our code is made freely available for academic purposes in [104].

Traditional methods for assessing structural correlates in the image domain, such as those that utilize numerical descriptors or pixel-wise comparison are able to test only a subset of the information available. Figure 1.7 shows that compared to deformation fields, transport maps yielded by OT can capture variations in both shape and texture between two images. At best, traditional techniques may produce statistical maps identifying the location of changes in the form of heat maps like Figure 3.6. However, current approaches cannot provide insight into the dynamic morphol-

ogy changes underlying the statistical maps. By analyzing differences in tissue spatial distribution, TBM can capture phenomena that affect the brain affecting the voxels nonlinearly such as aging. In contrast, pixel-wise comparison approaches implicitly assume that the changes affecting the brain can be described via a linear composition of isolated voxel intensities. Many pathologic changes in biophysical properties of tissue may not be well-modeled by volume/expansion contraction alone that deformation-based analysis assesses and Figure 3.7 shows that the components of OT better capture the natural structure in the data than pixel-wise comparison before or after DARTEL registration. Also, the tissue spatial distribution is more robust to field inhomogeneities/random noise compared to raw pixel-wise comparison. Another drawback of deformation-based modeling (Voxel-Based Morphometry, Deformation-Based Morphometry, Tensor-Based Morphometry, etc.) is that deformation fields are not unique to the pair of images [24]. Analyzing the deformation field through DBM or its Jacobian through Tensor-Based Morphometry may produce results that vary widely according to the particular field that is generated by the algorithm. Methods based on pixel-wise alignment such as VBM are sensitive to misregistration of voxels [24]. Thus, results obtained using deformation-based analysis are influenced by limitations of the method, which confound biological insights. In contrast, transport maps yielded by OT are unique and bijective. Therefore, statistics computed on transport maps can be mapped unambiguously to source images.

There are several limitations of this work. First, our approach for optimal transport minimization is non-convex. Although the approach does not guarantee that the global minima will be achieved theoretically, the experimental results demonstrate that the multiscale scheme guides the minimization to the global minima and the results are comparable to those using convex algorithms in 2D. We pose it as a future problem to couple the TBM framework presented in this chapter with solvers that can overcome limitations with large 3D images and at the same time are convex. Another limitation is that analyzing the spatial distribution of voxels requires a normalization of images. Thus, the TBM transform does not directly consider whether there are statistically significant differences in the sum of voxel intensities. However, the sum of voxel intensities can

be included as a feature when statistical analyses are performed in the feature domain.

We note that our formulation and TBM solver are fully general and applicable to many image modalities. Given the key role of OT-based modeling in numerous applications in medical imaging, including developments in Transport-Based Morphometry [12, 70, 84, 126, 136], and super-resolution imaging [83], an efficient numerical solution of OT suited for 2D as well as 3D digital imaging data opens the door to numerous research and clinical advances.

Chapter 4

TBM for Classification and Discrimination

4.1 Overview and Goals

In this chapter, we demonstrate how TBM can be applied to facilitate classification and discrimination tasks integral to several important clinical and research problems:

Osteoarthritis: Today, almost 1 out of 7 adults aged 25 years and older suffer from osteoarthritis, including more than a third of adults aged 65 and older [51]. Currently, clinicians and patients must wait for symptoms to develop, which include pain and irreversible changes on x-ray before a diagnosis can be made. However, the best chance at modifying disease trajectory lies in its early detection. Can osteoarthritis be accurately detected before symptoms develop using TBM to analyze the appearance of knee cartilage on MRI?

Brain morphology due to 16p11.2 duplications/deletions: Many patients with neurodevelopmental and neuropsychiatric disorders such as autism spectrum disorder, cerebral palsy, bipolar disorder and schizophrenia have copy number variants (CNVs) at the chromosomal locus 16p11.2 [93, 138]. Can the relationship between gene and effects on brain structure as seen on MRI be better elucidated with the help of TBM?

We demonstrate how transforming images from an Eulerian to Lagrangian representation using Transport-Based Morphometry can facilitate statistical learning using simple linear classifiers

and allows visualization of the discriminant changes. TBM enables these problems to be addressed, enabling state of the art results and closes important gaps in understanding osteoarthritis and neuroscience of the 16p11.2 genetic locus. Section 4.2 will focus on preclinical detection of osteoarthritis using TBM, while the work in Section 4.3 elucidates how TBM can be used to better understand the link between gene and brain structure in carriers of 16p11.2 CNVs.

4.2 Preclinical Detection of Osteoarthritis

4.2.1 Introduction

Articular cartilage in the knee undergoes early biochemical changes that are associated with osteoarthritis (OA). The normal biochemical content of articular cartilage consists of water, proteoglycans, and type II collagen. While there is a well-defined spatial organization of molecular content in healthy cartilage, there is evidence that the normal structure is disrupted in patients who go on to develop OA. MRI is a tool that can assess these changes in biochemical content and organization [9, 30, 41, 68, 78, 123, 129, 142]. However, as Figure 4.1 shows, human visual inspection is not sensitive enough to detect the subtle biochemical changes on MRI. Computer-aided pattern detection is needed to decipher and illuminate what the complex changes are. Early detection of OA would enable intervention when it has the largest potential to improve patient outcomes.

However, traditional pattern recognition techniques suffer from several limitations that prevent them from achieving adequate answers to questions addressed in Figure 4.1 [9, 41, 78, 123, 129]. Some studies have found that certain numerical descriptors, such as texture statistics, appear to be perturbed in subjects who progress to OA [41], but a fundamental gap in knowledge lies in how these numerical metrics correspond to appearance of cartilage on imaging. For instance, Tamura texture features and histogram statistics do not have direct biological interpretation. The potential for clinical translation as well as understanding the OA pathogenesis are hampered by the inability

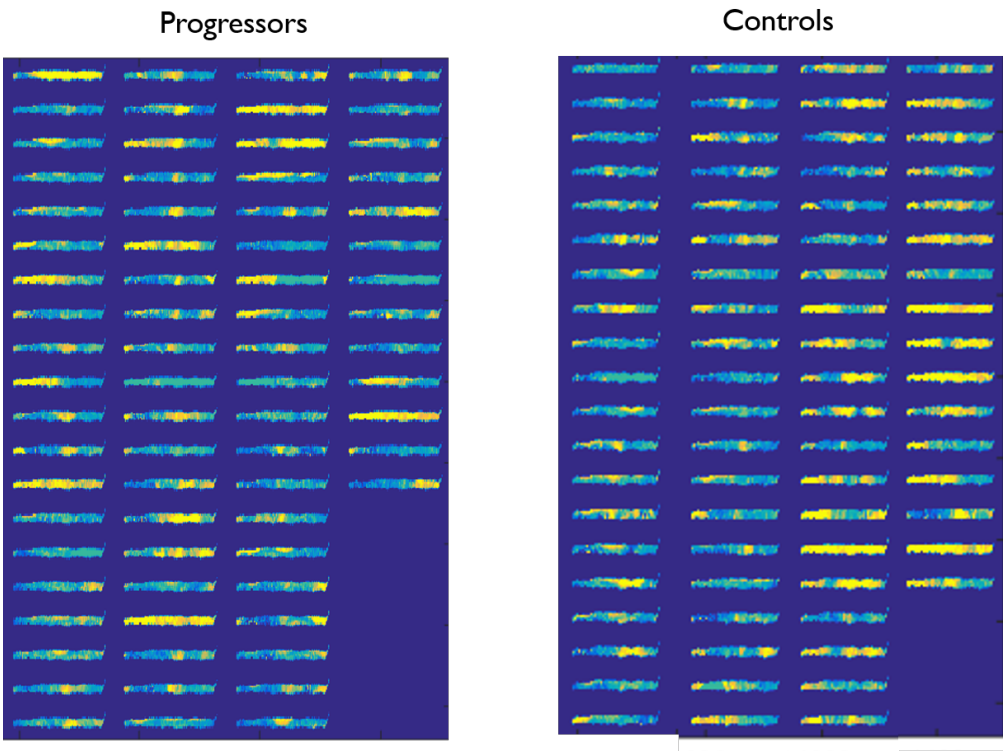


Figure 4.1: Lateral condyle images corresponding to individual subjects. These are T2-weighted MRI of two groups of asymptomatic subjects. In three years, only one of these groups will go on to develop clinical OA. Relevant questions include: Can these groups be differentiated based on image appearance alone? If so, what features are most discriminating? How can we model or visualize the abnormal variation leading to discrimination between the two groups?

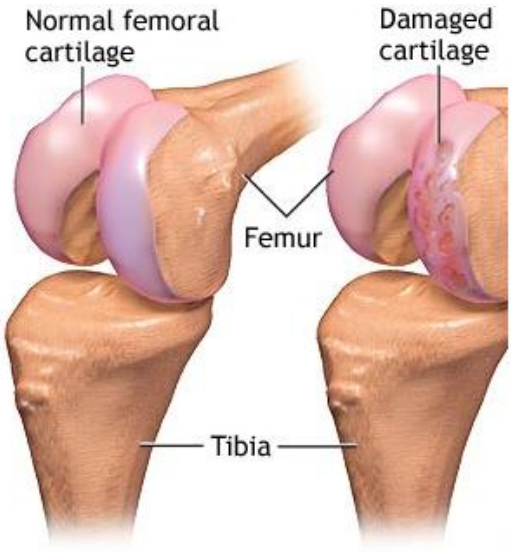


Figure 4.2: The femoral cartilage consists of two C-shaped segments, the medial and lateral condyle cartilages. Source: [32]

to model or visualize changes.

We apply TBM [12, 85, 136] to characterize variations in cartilage texture as seen on T2 imaging. The subjects used in this study are selected from the Osteoarthritis Initiative (OAI) database. These subjects initially show no clinical or radiographic evidence of OA at baseline, but on three-year follow-up, a subset of these patients go on to develop clinical OA as determined by scores on the Western Ontario and McMaster Universities Arthritis (WOMAC) questionnaire. Using TBM, we are able to train a classifier that yields accuracy in prediction of up to 86.2%, the highest of any result reported in the literature to date. For the first time, the changes that enable preclinical, pre-symptomatic OA can be visualized and we see that signature patterns of injury to medial and lateral condyle emerge that are sensitive for future progression to OA.

4.2.2 Methods

Patient Cohort

The data were obtained from the OAI database, an NIH-sponsored database available for public access at [103]. The flow diagram in Figure 4.3 summarizes how subjects were selected for the study. A total of 70 subjects (31 from the unexposed control subcohort and 39 from the incidence subcohort) were included in the study. Non-progression subjects were chosen from the unexposed control subcohort, defined as those subjects for whom WOMAC [15] score < 10 and KL score ≤ 2, both at baseline and at three-year follow-up, and no risk factors for OA progression. The symptomatic OA progression cohort was selected from the incidence subcohort based on initial baseline criteria of WOMAC score ≤ 10 and baseline KL score ≤ 2, but on three-year follow-up these subjects were required to have a change in WOMAC score > 10 from baseline indicating progression to symptomatic OA. Computational analysis included only the subjects for whom both the lateral and medial condyle segmented cartilage images were available (subsection 4.2.2). Patients were excluded from the OAI study cohort based on the following criteria: those who had a history of rheumatoid arthritis, bilateral total knee join replacement, or positive pregnancy test. Institutional review board approval was obtained at all participating institutions in the OAI, and informed consent was obtained by all participants in the study [97].

MRI Acquisition

The 3D sagittal double-echo steady-state (DESS) and T2 mapping images were obtained from the OAI imaging database by request (http://oai.epi-ucsf.org) [108]. The knee radiographs obtained at baseline were standard bilateral standing posterior-anterior fixed flexion knee radiographs. The knee radiographs were graded according to the KL scoring system [79].

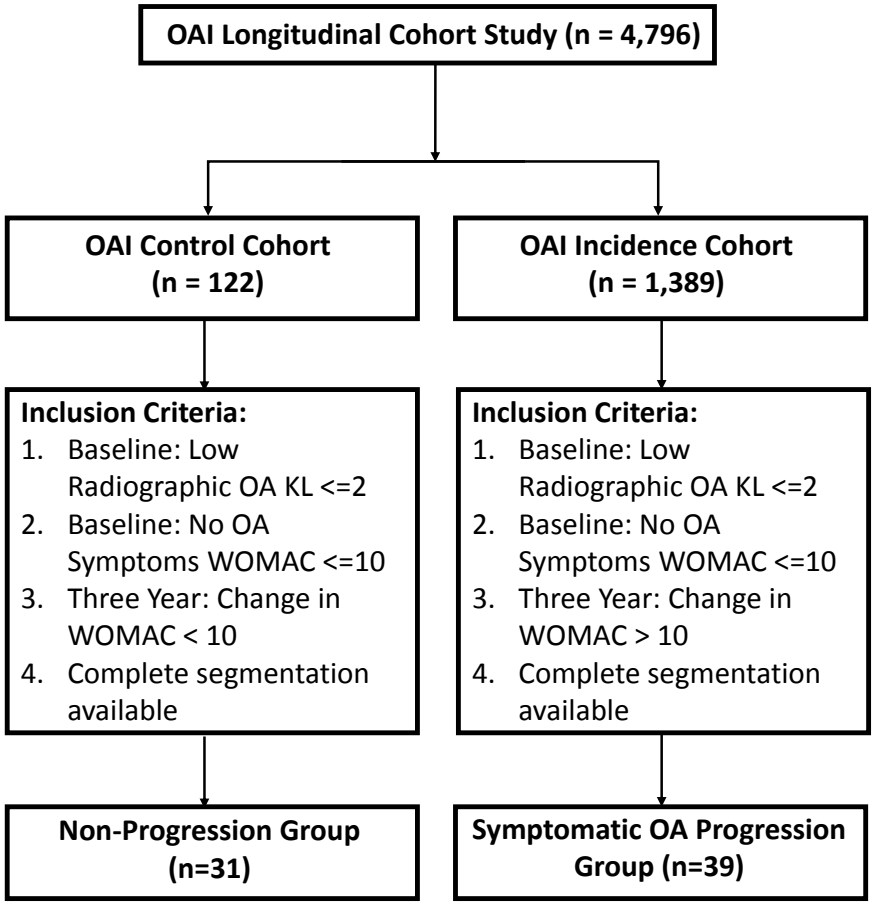


Figure 4.3: Subjects were selected from the OAI cohort based on no radiographic or symptomatic evidence of OA at baseline. On three-year follow-up, the progression cohort was defined as those subjects who developed OA symptoms as assessed by change in WOMAC score.

Registration, Segmentation and Preprocessing

All images were automatically segmented using nonrigid registration to a template image. The template image was segmented by hand. The algorithm was developed by our collaborators Dr. Richard Spencer and Beth Ashinsky at the National Institutes of Aging. Further details will be provided once the publication is available in the public domain.

After segmentation, the sagittal curvature of the femoral condyles (C-shaped) were mapped to a linear plane (flattened) to facilitate visual interpretation. Landmark-based thin plate spline (TPS) registration [23] was used to compute a nonrigid mapping. The mapping was done slicewise and the landmarks were chosen by an automated procedure that extracted the centerline as well as major and minor curvatures automatically from each slice of the condyle. For each 3D slice, points were uniformly sampled on the centerline and mapped to a set of points of the same spacing/width apart on a horizontal line of the same length. The thickness along each condyle slice was preserved by mapping a pair of points straddling each centerline point, one each on the minor and major curvatures, to a pair of points placed on either side of the horizontal line landmark points such that the vertical distance was preserved. The mapping functions to flatten each slice of each condyle was computed based on the first image in the dataset, and applied to all the remaining images. Thus, the effect of flattening is standardized across all images in the dataset. Figure 4.4 illustrates the flattening process that preserves the dimensions of the condyle.

Given sets of corresponding landmark points $x_1, ..., x_N$ and $y_1, ..., y_N$, the parameters $p = A, t, c_1, , c_N$ are sought such that

$$y_i = f(x_i) = Ax + t + \sum_{k=1}^{N} c_k \phi(||x_i - x_k||) \tag{4.1}$$

where $\phi(x)$ are radially symmetric basis functions. In 2D, $\phi(r) = r^2 log(r)$. The mapping function in Equation 4.1 was computed separately for each sagittal slice based on the first image in the dataset, and the same function was applied to all subsequent images.

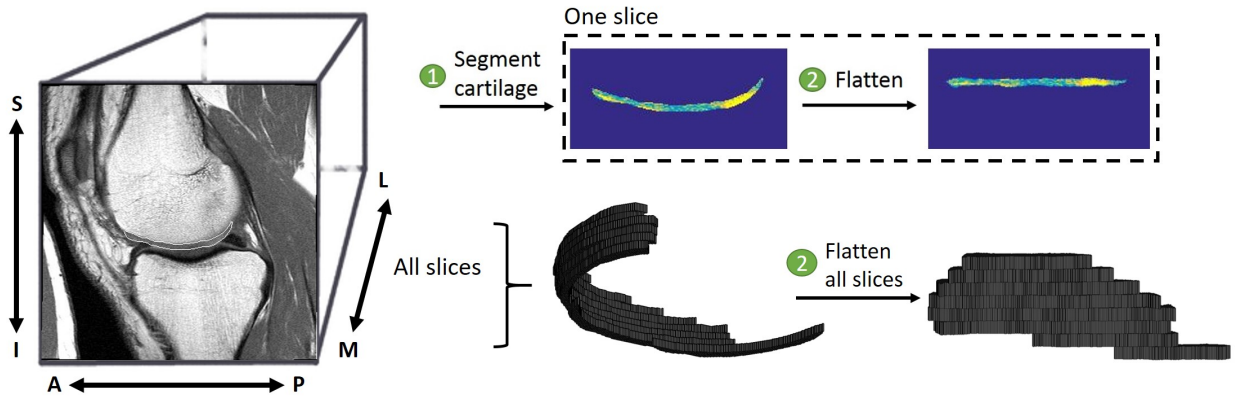


Figure 4.4: The top row shows the flattening process of a single sagittal slice of the curved condyle image. The sagittal curvature of the condyle image was mapped to a 3D plane with the same dimensions (length, width, and thickness), as the bottom row shows through volume rendering.

Quantifying texture variation using Transport-Based Morphometry

Given a set of preprocessed knee cartilage images, $I_1(x), ..., I_N(x)$ corresponding to experimental subjects $1, ..., N$ (where x is a coordinate in 3D space), we can compute the unique TBM transformation for the image $I_1(x)$ according to the analysis equation (3.2).

A transport map $f_i(x)$ is computed for each subject i that quantifies the T2 spatial variation in each image $I_i(x)$ compared to the common reference image $I_0(x)$. The direction that best predicts OA progression in the transport domain can be inverted to visualize signature variations distinguishing normal from OA progression by the synthesis equation (3.3). All TBM codes were implemented in MATLAB (MathWorks, Natick, MA).

Classification

Variations that characterize or separate the progressor and non-progressor cohorts are sought using the linear discriminant analysis (LDA) technique. The LDA technique finds a discriminant boundary between the control and progressor cohorts in the transport space according to Equation 4.2

$$w_{LDA} = arg \max_{||w||=1} \frac{w^T S_T w}{w^T S_W w} \tag{4.2}$$

Note that the direction w_{LDA} can equivalently be computed using the formula $w_{LDA} = arg \max_{||w||=1} \frac{w^T S_B w}{s^T S_W w}$, where the "between-class scatter matrix" is $S_B = S_T - S_W$. Given the *m*th TBM transformed image f_m, let us call refer to the vectorized version of the transport map f_m as x_m. The covariance matrix S_T can be written as

$$S_T = \frac{1}{M} \sum_m (x_m - \bar{x})(x_m - \bar{x})^T \text{such that } \bar{x} = \frac{1}{M} \sum_{m=1}^{M} x_m$$

Here, $S_W = \sum_C \sum_{n \in C} (x_n - \bar{x_c})(x_n - \bar{x_c})$, and represents the "within-class scatter matrix." Here, where $C = 2$ and represents the number of classes. The discriminant direction w_{LDA} was computed and validated using complete leave-two-out cross validation, where all pairs of control and progressor images were held out sequentially to determine accuracy, sensitivity, and specificity of the computed discriminant direction.

Before classification using LDA, the principal components analysis (PCA) technique was used to eliminate dimensions containing little or no contribution to the overall variance in the dataset. Only the top 95% of the variance was retained for the medial and lateral condyle images. All LDA and PCA codes were implemented in MATLAB.

Statistics

Group differences were assessed by multivariate analysis of variance (MANOVA) using the SPSS software. Receiver operator characteristic (ROC) curves were computed using the projection score when images are projected onto the computed discriminant boundary.

Table 4.1: Subject demographics

Parameter	Progressors (n = 39)	Control (n = 31)	p
Age	60.2 ± 9.0	56.5 ± 9.2	0.09
BMI	29.0 ± 4.3	25.1 ± 4.4	3.7×10^{-4}

Age and body-mass index of subjects in each group at baseline

4.2.3 Results

Age and BMI

There were no statistically significant differences in age between progressor and control cohorts, but the differences in BMI were statistically significant as Table 4.1 summarizes. In Table 4.2, we see that age and BMI alone predict OA progression with accuracy above chance, but are not sufficient to enable discrimination with high sensitivity and specificity.

Multivariate Group Differences

There were significant differences in the signal variation between progressor and non-progressor cohorts. Multivariate analysis of variance (MANOVA) was performed for the medial condyle images and lateral condyle images separately on the transport maps. For medial condyle images, $F(29, 40) = 3.530$, $p = 1.3 \times 10^{-4}$, Wilks $\Lambda = 0.281$. For lateral condyle images, $F(30, 39) = 1.845$, $p = 0.036$, Wilks $\Lambda = 0.413$. When both condyles were considered jointly, $F(40, 29) = 5.983$, $p = 2 \times 10^{-6}$, Wilks $\Lambda = 0.108$.

Classification

The T2 spatial variations were assessed for their ability to discriminate progressor and control images using TBM. The test accuracy is reported in Table 4.2, computed using an LDA classifier and complete leave-two-out cross validation as described in Section 4.2.2. The highest predictive accuracy can be achieved when both condyles are considered jointly rather than considering solely

Table 4.2: Classification results

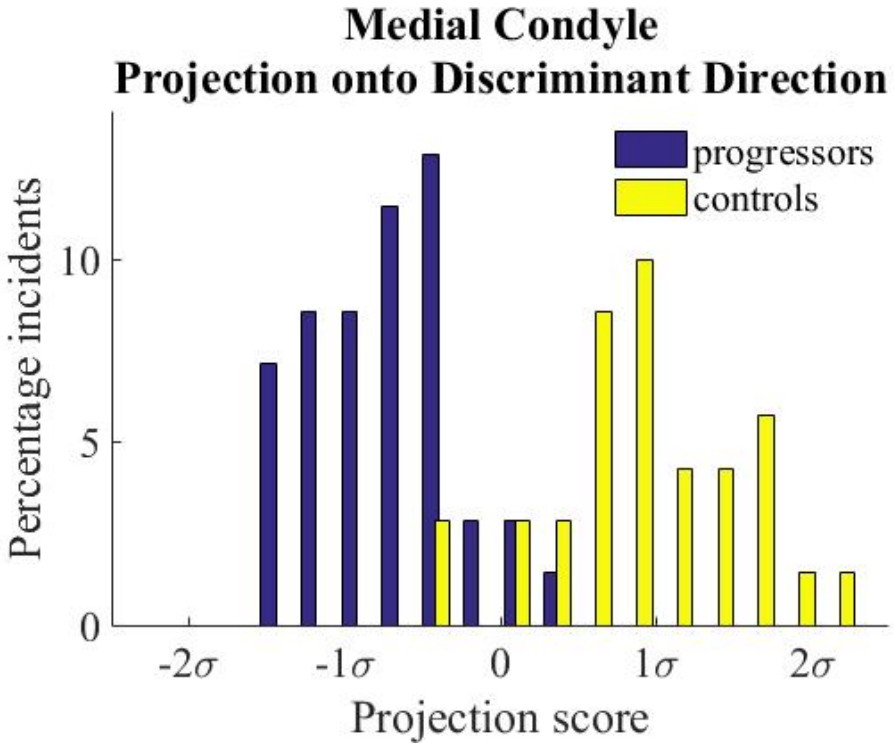
Region	Test Accuracy	Sensitivity	Specificity	Cohen's Kappa
Age and BMI only	72.3%	73.6%	71.0%	0.45
Medial condyle	73.0%	77.1%	69.0%	0.46
Lateral condyle	65.6%	66.5%	64.6%	0.31
Both condyles combined	86.2%	86.6%	85.9%	0.72

medial or lateral condyles individually.

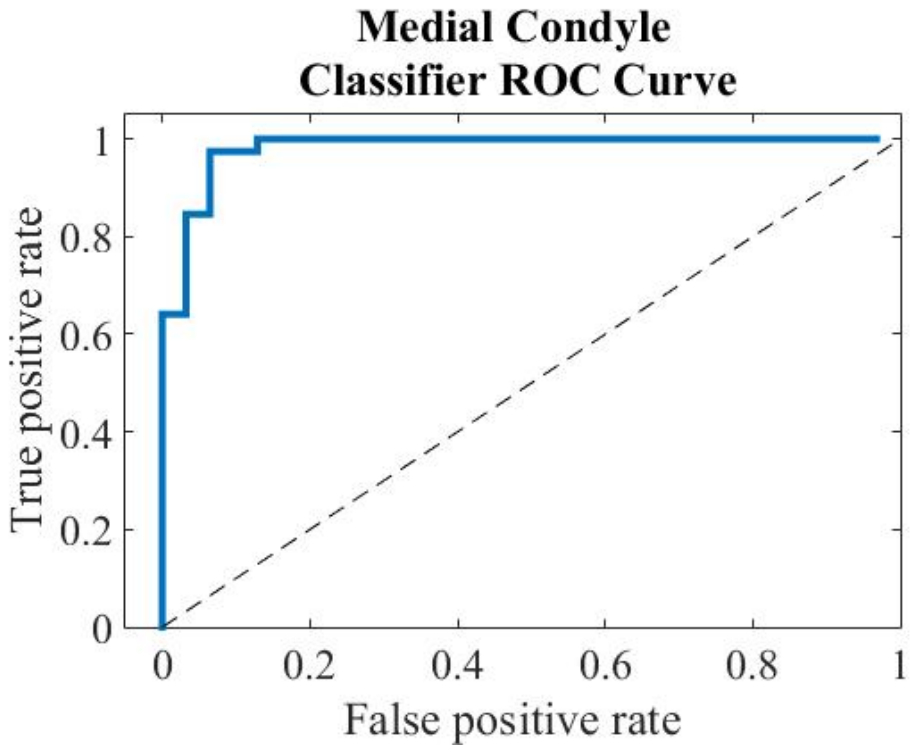
Figures 4.5 and 4.6 summarize the LDA classifiers for medial condyles and lateral condyles respectively. In Figure 4.5a and 4.6a, the histograms show the distribution of the entire subject data when it is projected onto the computed discriminant direction. The receiver-operator characteristic curves in Figures 4.5b and 4.6b illustrate the quality of the binary classifiers as the discrimination threshold is varied. The shape of the ROC curve indicates that a robust classifier can be obtained.

Visualizing Discriminating Variations

The separability of the cohorts when both condyles are considered together is shown by Figure 4.7. Images generated along the direction of the classification boundary are shown in Figures 4.7b and 4.7c. Lower slice numbers for both lateral and medial condyles correspond to the lateral aspect of the knee, whereas higher slice numbers correspond to medial location. Focal damage in the center of the medial condyle, as indicated by brightness in intensity, appears to be a sensitive change associated with future OA. In lateral condyles, especially in the more medial aspect (slices 5-8), the pattern of brightness in intensity appears to become more diffuse for progressor images compared to control images.

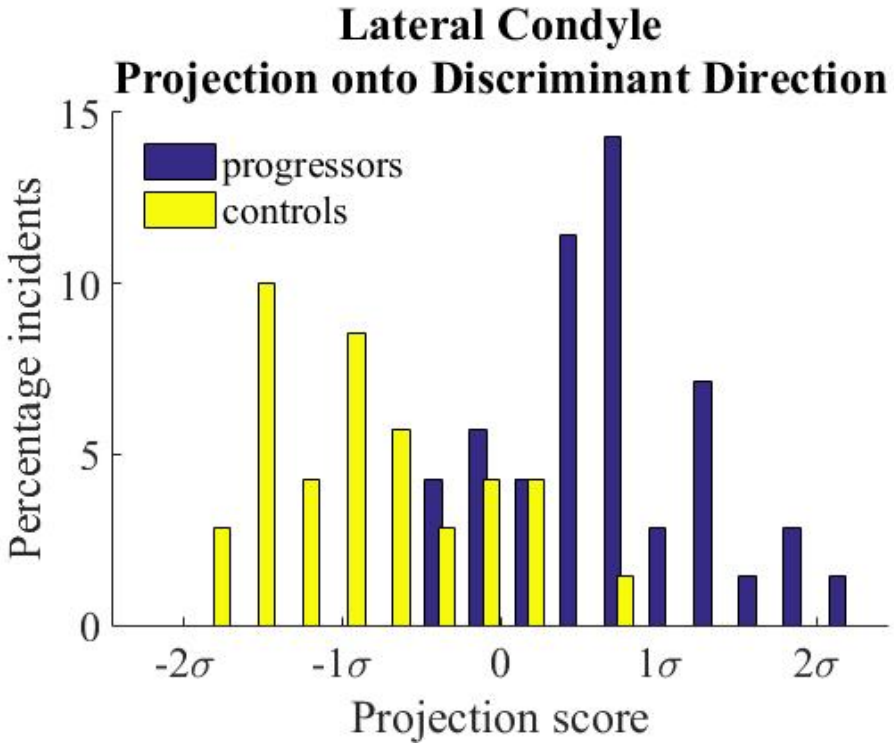


(a)

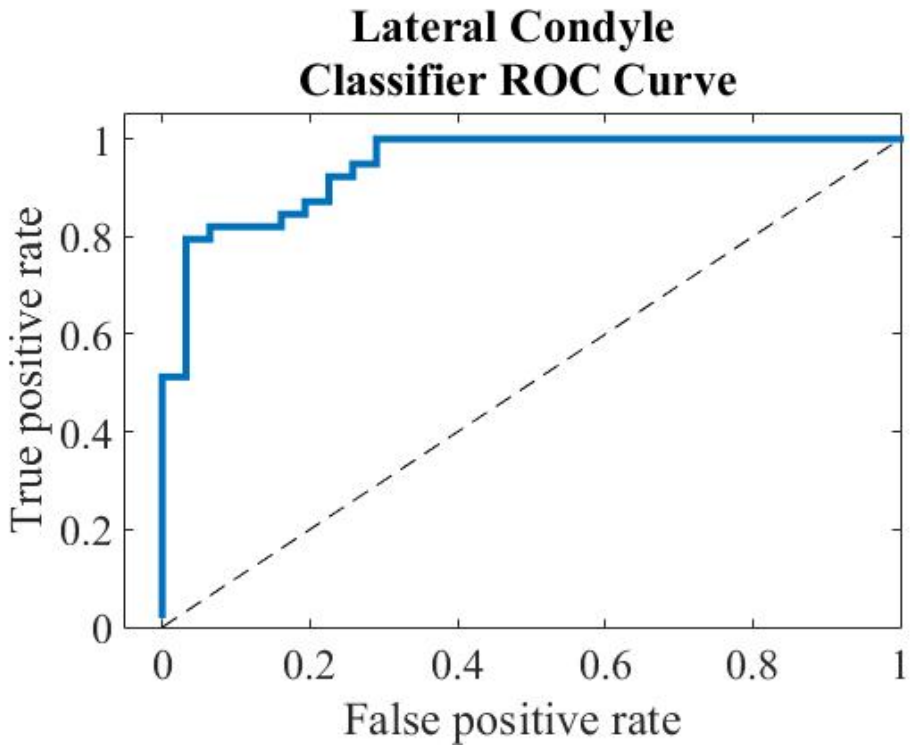


(b)

Figure 4.5: Medial condyles. (a) histogram illustrating the projection of the dataset onto the LDA classifier boundary (b) ROC curve illustrating performance of the classifier for multiple projection scores



(a)



(b)

Figure 4.6: Lateral condyles. (a) histogram illustrating the projection of the dataset onto the LDA classifier boundary (b) ROC curve illustrating performance of the classifier for multiple projection scores

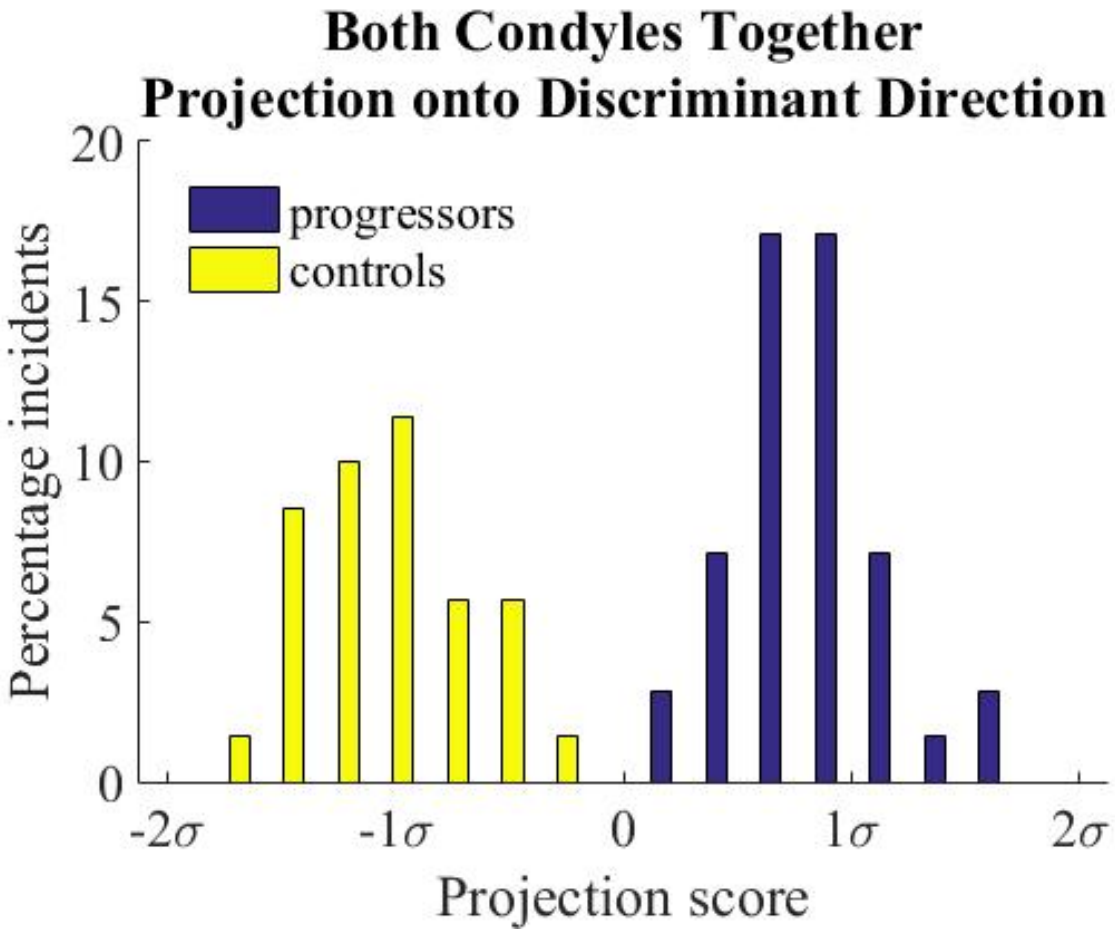


Figure 4.7: Both condyles combined. (a) histogram illustrating the projection of the dataset onto the LDA classifier boundary

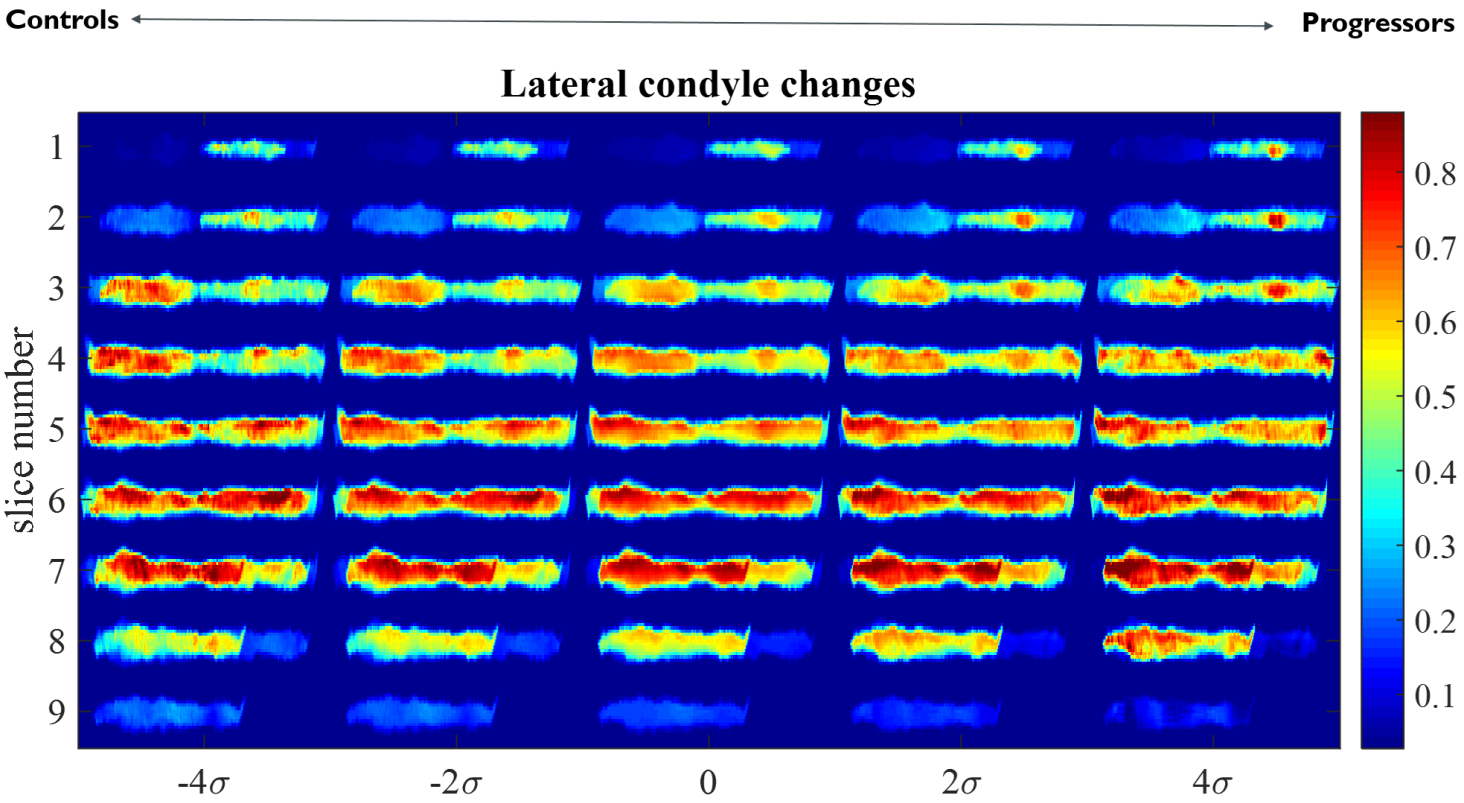


Figure 4.7: (b) lateral condyle images showing how cartilage texture changes going from control to progressor images

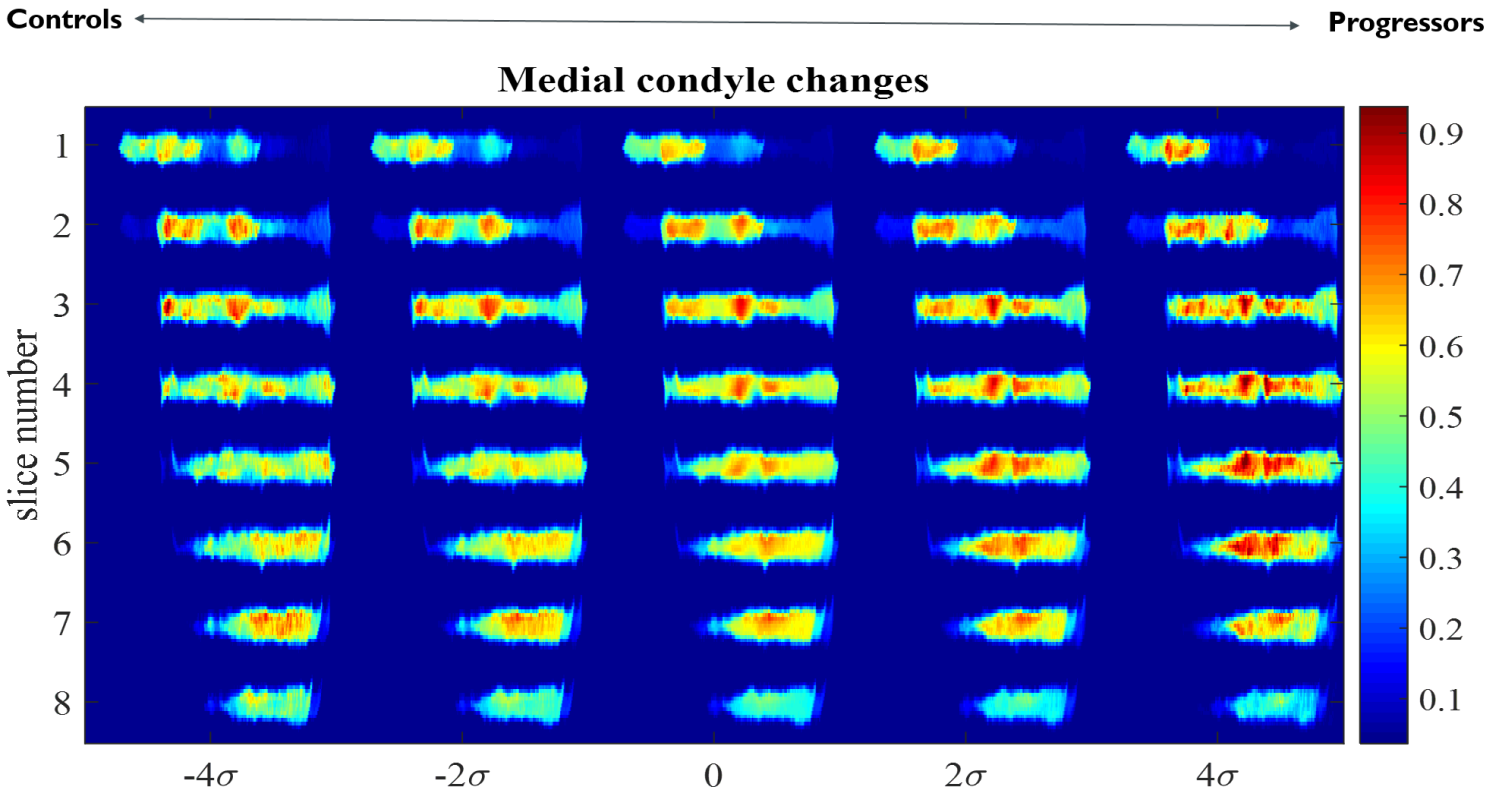


Figure 4.7: (c) images showing how medial condyle cartilage texture changes while moving from the control to the progressor subjects

We can confirm that the transport domain better captures nonlinear changes in T2 spatial distribution compared to the image domain by examining the scree plots in Figure 4.8. Fewer principal components are needed to represent the same amount of variability in the dataset in the transport domain than in the image domain, indicating that the transport domain components better capture the natural structure in the dataset compared to the image domain components.

4.2.4 Discussion and Summary

We apply the Transport-Based Morphometry (TBM) [12, 85, 136] technique to analyze T2 spatial distribution in articular cartilage of the knee. The goal was to predict incidence of osteoarthritis three years in the future based on the appearance of cartilage on T2-weighted imaging. We find that TBM can enable sensitive prediction of future OA based on analyzing T2 spatial distribution in cartilage. Furthermore, TBM for the first time allows variations that enable sensitive prediction to be modeled and visualized, revealing signature patterns of damage on T2-weighted imaging that can be used as clinical biomarkers for future OA.

The T2 map in normal cartilage shows spatial variation in signal intensity that describes the distribution of water molecules, proteoglycans, and orientation of collagen fibers [65, 141, 142]. Changes in articular cartilage texture on T2 imaging implies a biochemical change in composition of the cartilage. It has been reported that in early OA, the water content in the cartilage increases as well as anisotropy of collagen fiber orientation [142]. We discover by TBM modeling that a focal increase in brightness at the medial condyle center is sensitive for future OA. The pattern of T2 hyperintensity appears to become more diffuse in the lateral condyle, especially in the posterior segments of its medial aspect. The changes in the distribution of signal hyperintensity may suggest a possible mechanism of damage that includes fissuring and replacement of damaged tensile tissue with water. A possible mechanistic explanation may include a fissuring process in which fibrous tissue is damaged and water (bright intensity) diffuses into the holes.

Risk factors for OA include advanced age, BMI, gender, genetic predisposition, joint injury

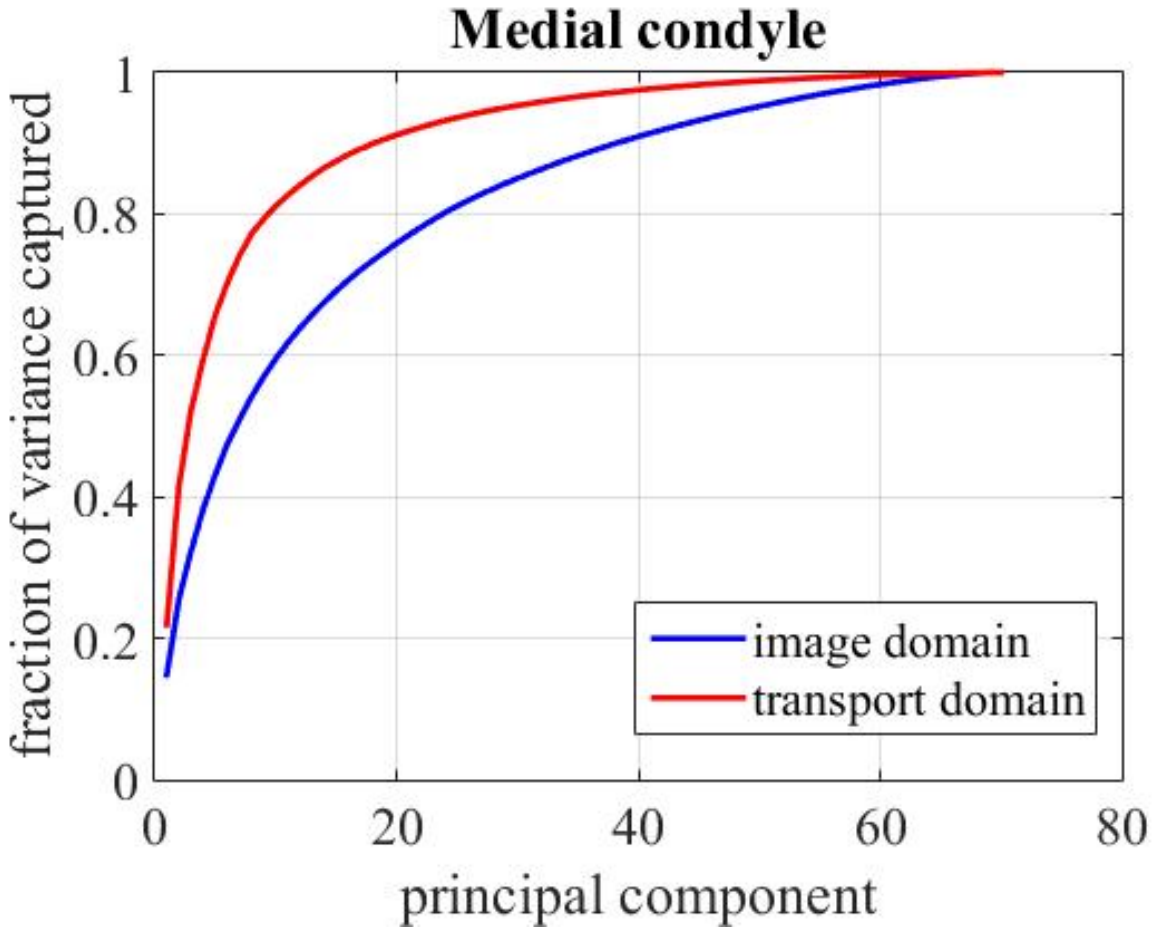


Figure 4.8: Scree plots show that fewer components are needed to represent images in the TBM space than in the image domain. Thus, more of the variations are captured by considering spatial distribution of intensities rather than voxel-by-voxel comparison. The plots for (a) medial condyle

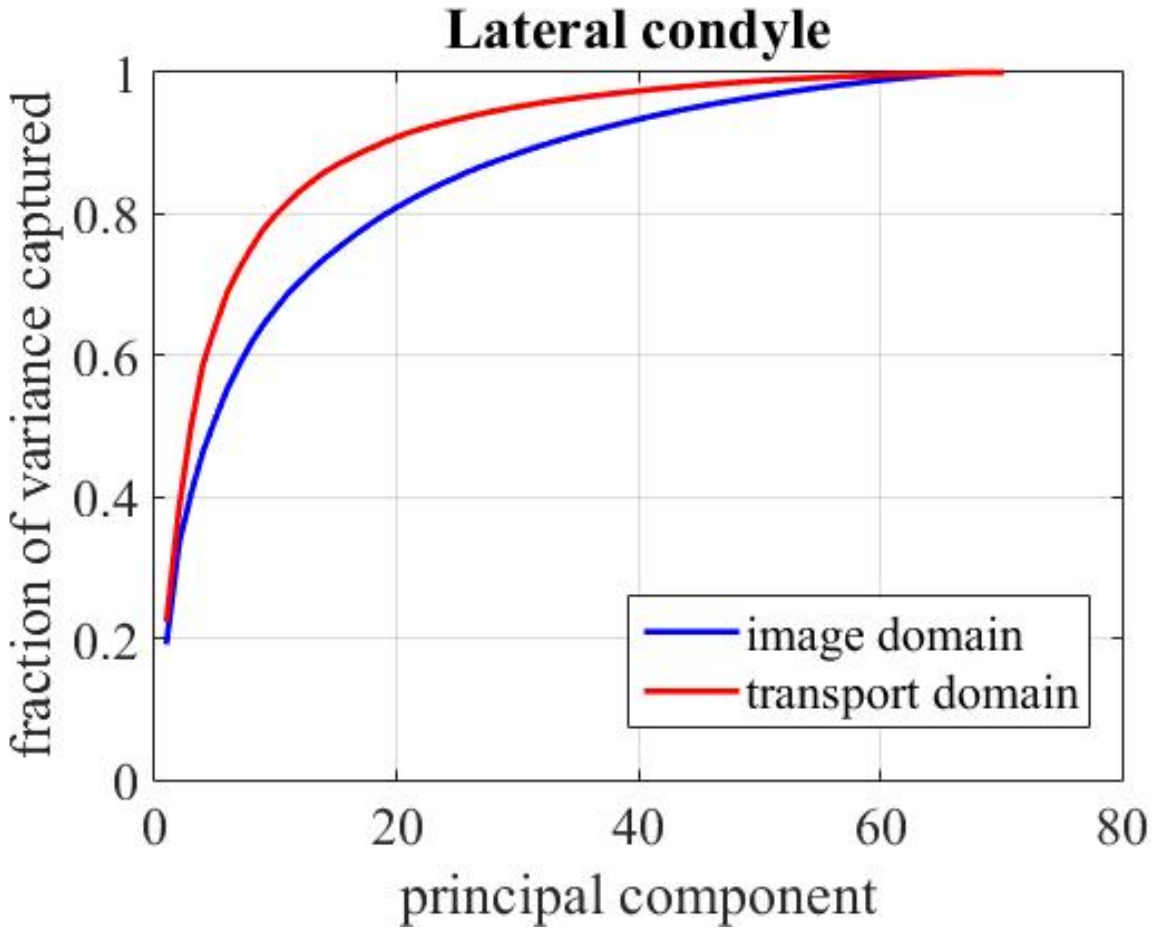


Figure 4.8: (b) lateral condyle are shown separately.

[51], etc, which all contribute to the ultimate appearance of cartilage on T2 weighted imaging. However, these are only two of many risk factors that are responsible for damage to cartilage. We see that the prediction accuracy can be greatly improved when the T2 images of cartilage are used for prediction rather than individual risk factors that contribute to its appearance (Table 4.2).

Ours is not the first study to investigate association of T2 spatial variation in cartilage to OA development, but our prediction accuracy of 86.2% is the highest reported to date. Both Urish *et al* [129] and Ashinksy *et al* [9] investigate whether early detection of OA is possible three years in advance of symptoms (note that KLU is an author on all three studies). The best prediction accuracy reported by Urish *et al* is $76.2 \pm 0.7\%$ and Ashinsky *et al* report 36-70% accuracy when the images are used in binary classification of progressor and non-progressors using T2 maps. In addition to enhanced prediction ability, our method additionally confers the ability to model and visualize the images enabling sensitive prediction for the first time, revealing distinct patterns of injury that are useful for clinical identification and also for learning more about the pathogenesis of OA at the structural level.

The Transport-Based Morphometry approach has only recently become available for automated discovery and visualization of structural markers from radiology images by the enabling work of Kundu *et al* [85]. The TBM approach bears some similarity to registration-based image analysis techniques such as Voxel-Based Morphometry (VBM) [6] and Deformation-Based Morphometry (DBM) [8], but there are several critical differences that make TBM an ideal tool study T2 spatial variation. First, VBM and DBM can capture differences only in shape, not texture. Given that this study focuses on T2 texture variation, VBM and DBM would not be candidate methods for this analysis. In contrast, TBM can analyze differences in shape as well as texture. Second, the transport maps produced by TBM enable generative modeling and direct visualization of the subclinical damage.

There are several limitations of this study. Sources of noise in this study are due to scanner resolution, and segmentation/registration procedure. The number of subjects used in this study is

relatively small. Larger scale studies utilizing the TBM technique and clinical studies are needed to explore and build further on these results.

In the future, the TBM approach can be used to model shape of cartilage as well as texture simultaneously. In this study, we were interested in studying how T2 spatial variation can be used to predict OA. However, the TBM technique is entirely general to any imaging modality. In the future, approaches based on molecular imaging and high-resolution imaging of cartilage can allow a more detailed investigation of how biochemical structure and molecular composition change as a function of future OA. There is strong potential for TBM to be coupled with molecular imaging techniques in the future to enable very detailed analysis of cartilage structure and changes over time. Thus, the potential for TBM to help solve many mysteries about OA is enormous.

4.3 Understanding brain morphology associated with the 16p11.2 genetic locus

4.3.1 Introduction

There is evidence that genetics play a role in disorders such as autism, but assessing the relationship between genotype and phenotype is challenging because many complex genetic variations result in the same phenotype. There have been a limited number of studies investigating the effects of common genes that predispose to neuropsychiatric and neurodevelopmental disorders using genetically-stratified cohorts. The 16p11.2 chromosomal locus (BP4-BP5) is one of these genes and copy number variants (CNVs) at this locus are associated with several disorders such as autism, schizophrenia, epilepsy and others [20, 64, 73, 112, 117]. In fact, carrying a 16p11.2 deletion is associated with 38.7 fold increase (95% CI: $13.4 - 111.8$) in developing autism or developmental delay, and a duplication is associated with 20.7 (95% CI: $6.9 - 61.7$) fold increase [93]. Prior studies have investigated whether group-level differences in brain structure between

duplication, deletion, and control cohorts exist [105, 109], but these studies do not provide insight into which differences enable reliable discrimination of these cohorts [53]. No study to date has investigated whether 16p11.2 copy number variation can be reliably predicted in individuals with a range of developmental impairments. The latter is important because establishing a causal link between genotype and phenotype would significantly advance understanding of gene-brain-behavior relationships.

Unfortunately, predictive studies are generally not performed in analyzing brain structure [53], as the creators of VBM summarize that "in [predictive] models, the parameters have no physical meaning unless the generative model [used to explain the data] is invertible" [53]. Fortunately, we have demonstrated that Transport-Based Morphometry can bridge the gap in physical interpretation. In this study, we utilize the TBM image transformation framework to enable both classification and visualization of sensitive differences that predict 16p11.2 CNV based on the appearance of the brain on MRI alone. The regional gray matter and white matter morphological changes that enable differentiation shed light on the underlying neurobiology of 16p11.2 CNVs.

4.3.2 Methods

Participants

Subjects were recruited as part of the Simons VIP Project [37]. Participants were referred by clinical genetic centers or testing laboratories, web-based networks, and self-referral of families who learned about Simons VIP (http://SimonsVIPconnect.org) and underwent an initial screening and review of medical records by Geisinger and Emory University. Inclusion criteria were those who had recurrent breakpoints at BP4-BP5 of 16p11.2 without other pathologic CNVs or known genetic diagnoses/syndromes. Exclusion criteria were those who had a history of environmental insults with potential influence on neurocognitive status (such as fetal alcohol syndrome), severe birth asphyxia, severe prematurity, or lack of fluency in English.

Table 4.3: Subject demographics

Cohort (N = 235)	**Age (years)**
Deletion carriers (N = 51)	15.1 ± 11.4
Duplication carriers (N = 53)	24.0 ± 16.0
Controls (N = 131)	24.5 ± 14.5

There are significant differences between the ages of deletions vs. controls and deletions vs. duplications. M:F ratio $= 1.35$

Behavioral testing included the Autism Diagnostic Observation Schedule (ADOS), Autism Diagnostic Interview (ADI) and broad screening measures of social impairment such as the Social Responsiveness Scale (SRS). The core sites for phenotyping were University of Washington Medical Center, Baylor University Medical Center and Boston Children's Hospital.

Neuroimaging was performed at University of California (UC) and Children's Hospital of Philadelphia (CHOP).

The control cohort consisted of subjects recruited locally near the core imaging sites (UC-San Francisco, UC-Berkeley, CHOP) from the general population and were matched for age, sex, handedness, and nonverbal IQ. Exclusion criteria included major DSM-IV diagnoses based on clinical psychologist review or immediate family history of ASD, other developmental disorders, dysmorphic features, or genetic abnormalities. Subjects were also excluded for a Symptom Checklist 90 (SCL-90) score > 62, an axis I psychiatric diagnosis on the Diagnostic Interview for Children, a Broader Autism Phenotype Questionnaire aloof subscale score $> 4/3.5$ (male/female), a pragmatics subscale score $> 3.25/3.5$, or a rigidity subscale $> 4/3.7$. Control subjects also had a chromosome microarray, neurologic exam, a clinical psychologist interview, and a photograph evaluation for dysmorphology. Subject demographics are summarized in Table 4.3.

Deletion and duplication carriers had a range of neurodevelopmental diagnoses, summarized in Table 4.4.

Table 4.4: Diagnoses across duplication and deletion carriers

Diagnoses	**Number with diagnosis (%)**
ADHD	18 (19.8%)
Anxiety, OCD, Phobia	17 (18.7%)
Articulation disorder	21 (39.9%)
Behavioral disorder	9 (9.9%)
Clinical ASD	14 (15.4%)
Coordination disorder	24 (26.4%)
Tic disorder or Tourette syndrome	6 (6.6%)
Enuresis	9 (9.9%)
Language disorder	16 (17.6%)
Learning disorder	6 (6.6%)
Mood disorder	10 (11%)
Intellectual disability	20 (22%)
Stereotypic movement disorder	2 (2.2%)
Mean number of diagnoses per subject	2.0 ± 1.6

The 16p11.2 duplication and deletion carriers had a range of diagnoses. Abbreviations: ADHD = attention deficit hyperactivity disorder, OCD = obsessive-compulsive disorder, ASD = autism spectrum disorder. Note: This table excludes the 13 subjects for which diagnoses were unavailable.

MRI acquisition

Structural imaging was performed at UC and CHOP sites. The images were collected on matched 3T TIM Trio MRI scanners using vendor-supplied 32 channel phased-array radio frequency head coils. Structural imaging included a high resolution multiecho T1-weighted magnetization-prepared gradient-echo image (ME-MPRAGE). Scan parameters were as follows: $TR = 2350$ ms, $TI = 1200$ ms, $TE = 1.64$ ms, $FA = 7°$, voxel size $1 \times 1 \times 1$ mm, and $FOV = 256$ mm.

Images with severe wrapping, poor head coverage, ringing/striping/blurring, ghosting, inhomogeneities or susceptibility artifacts were not included in the analysis (13 subjects were discarded for quality control).

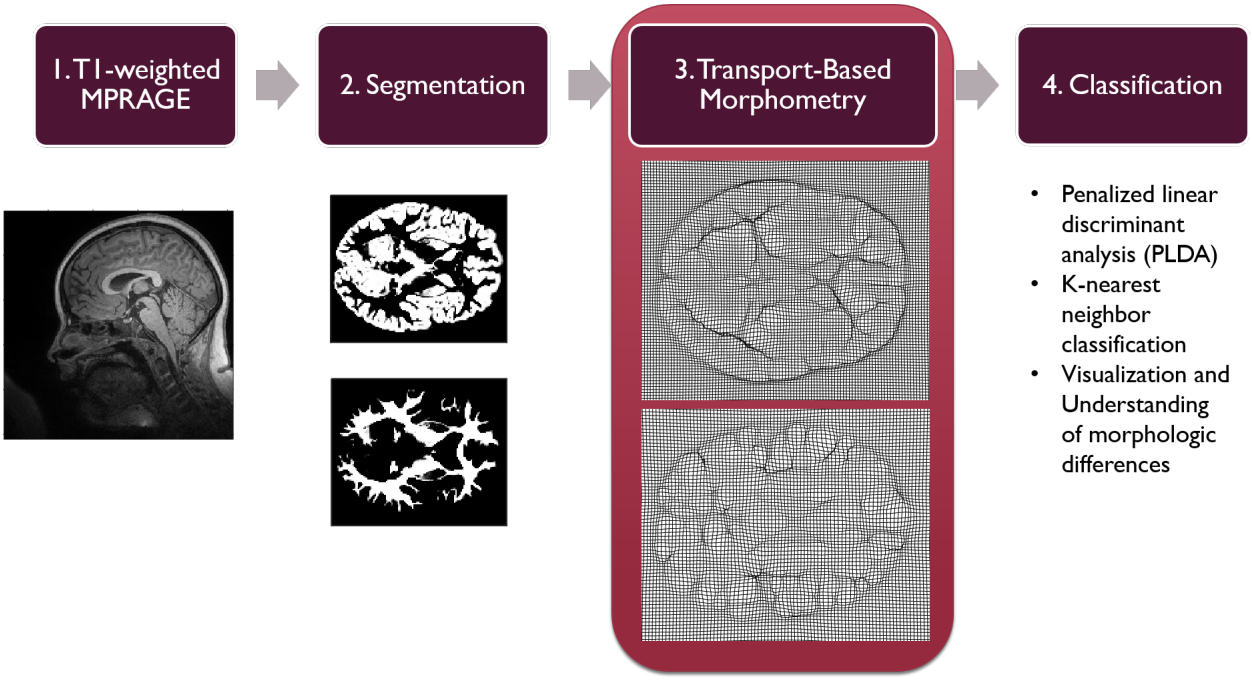


Figure 4.9: After segmentation, transport maps are computed for both white matter and gray matter channels as compared to a common reference template using optimal mass transport. Subsequent statistical analysis is performed on the transport maps that quantify the variation of each image from the common reference.

Capturing structural variations using Transport-Based Morphometry

Statistical Parametric Mapping (SPM12) [52] was used to affinely coregister and segment white matter and gray matter tissues.

Common reference images for gray matter and white matter images were computed by taking the Euclidean average of all of the gray matter and white matter images, respectively.

Subsequently, TBM was applied to generate transport maps to characterize the individual spatial tissue distribution compared to common template images for gray or white matter channels separately. Equations for analysis (Equation 3.2) and synthesis (Equation 3.3) are described in Chapter 3.

The transport maps were vectorized and concatenated into a data matrix, which served as the

feature set for subsequent statistical learning.

Classification

The standard principal components analysis (PCA) technique was applied to reduce the dimensionality of the data matrix before classification. The data was projected onto the components representing the top 96% of the variance.

After dimensionality reduction, the penalized linear discriminant analysis (PLDA) method was used for classifying individual samples. A $C-1$ dimensional subspace in the transport space is sought using PLDA method such that the projections of the $C=3$ classes (duplication, deletion, control) are maximally separated. The PLDA directions are given by the solution to the optimization problem in Equation 4.3

$$w_{PLDA} = arg \max_{||w||=1} \frac{w_T S_T w}{w^T (S_W + \alpha I) w} \tag{4.3}$$

The matrices S_T and S_W are defined according to the same convention in Equation (4.2). The penalty weight α represents the tradeoff between the traditional LDA direction and the topmost PCA directions. The parameter α is chosen by plotting the stability of the subspace as a function of α and selecting the value for which the change between subsequent α values is $< 3\%$.

Test accuracy is evaluated using leave-one-subject out cross-validation. First, the discriminant subspace is computed using the training samples. Based on training sample estimates, the quantities $\hat{\mu}_j$ and $\hat{\pi}_j$ that indicate the class mean (i.e. class centroids) and prior probability of each class j respectively. These quantities are defined as follows.

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{y_i = j} x_i \tag{4.4}$$

$$\hat{\pi}_j = \frac{n_j}{n} \tag{4.5}$$

Table 4.5: Classification accuracy using PLDA scheme

Tissue	Test Accuracy	Sensitivity	Specificity	Cohen's Kappa
Age and Gender only	32.4%	31.5%	91.6%	0.13
White matter	96%	94%	99%	0.93
Gray matter	92.8%	91.8%	98.3%	0.88

Classification accuracy using age and gender alone does not achieve sufficiently high accuracy in differentiating duplication vs. deletion vs. controls compared to white matter and gray matter tissues. (Note: the sensitivity and specificity reported are calculated based on identifying deletion carriers from all the rest.)

In the above, x_i indicates transformed images, and y_i indicates their class labels for $i \in 1, ..., n$, where n is the total number of samples.

To classify a test sample, the class centroids are projected onto the discriminant subspace to give the transformed centroids $\tilde{\mu}_j$ and the sample is assigned to the class j according to the nearest centroid, based on minimizing $||\tilde{\mu}_j - \tilde{x}||_2^2 - \log \hat{\pi}_j$.

In order to compare to the classification accuracy using age and gender only, we use the same PLDA classification scheme setting $\alpha = 0$.

Visualization and Interpretation

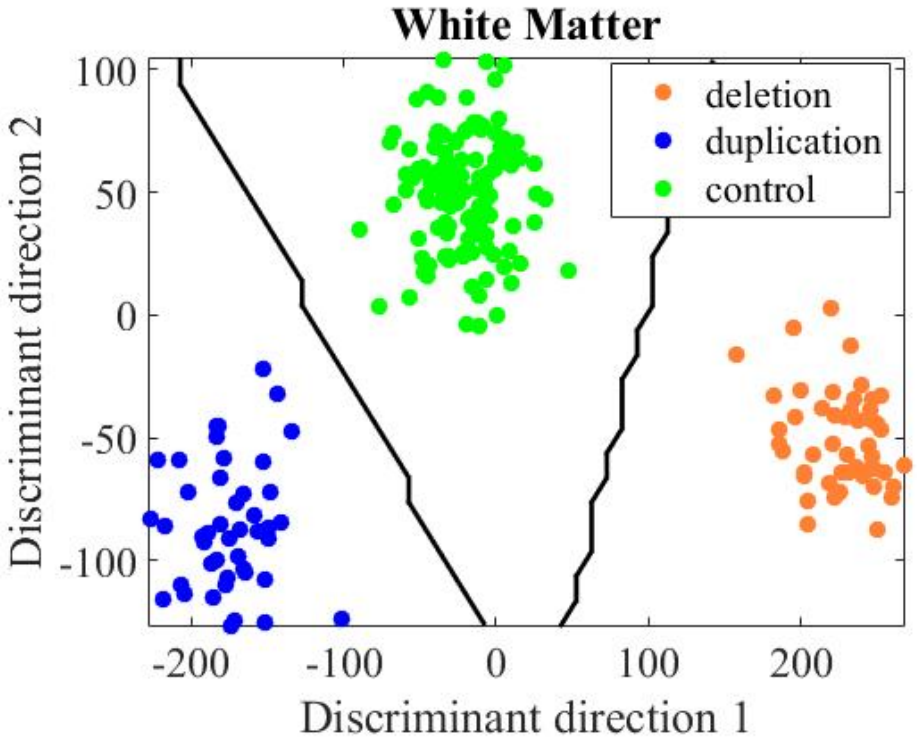
The discriminant subspace was sampled and inverted to generate the corresponding images through inverse TBM transformation using Equation (3.3). A neuroradiologist with $18+$ years of experience interpreted the final images.

4.3.3 Results

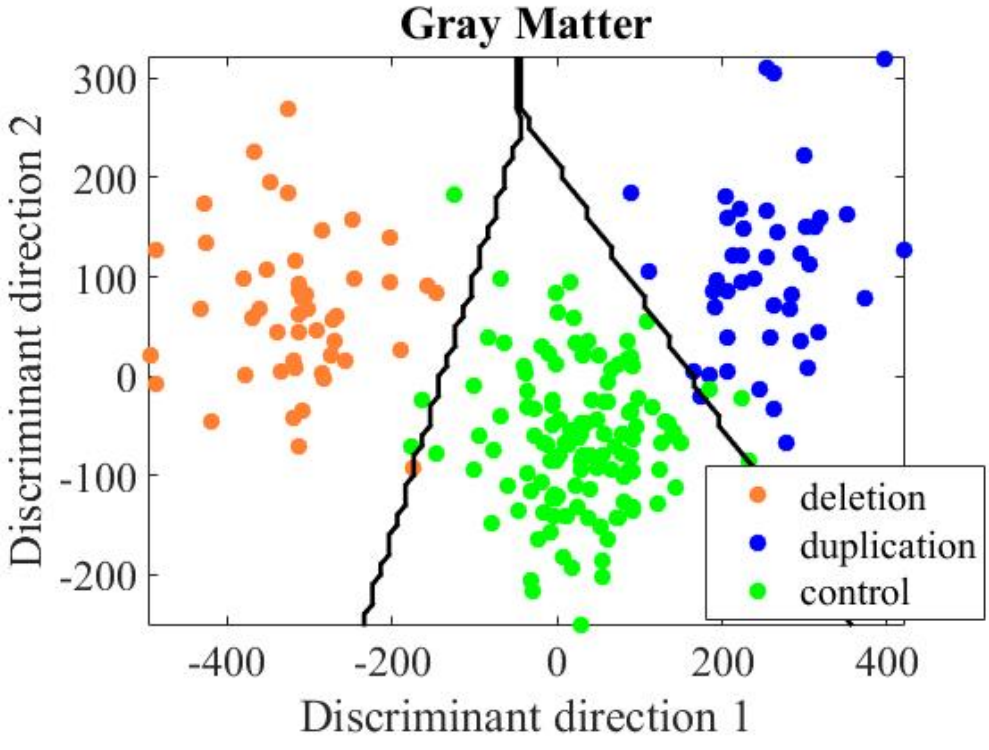
Classification

As Table 4.5 indicates, using white matter alone enables 96% test accuracy in discriminating the cohorts. Gray matter alone enables 92.8% test accuracy. When the data is projected onto the maximally discriminant subspace yielded by PLDA, distinct clusters emerge, as shown in Figure 4.10.

Figure 4.11 shows a single sample axial slice of white matter images generated by sampling the discriminant subspace in shown in the scatter plot in Figure 4.10a. These images can be interpreted to assess characteristic changes in image morphology, described in the next subsection.



(a)



(b)

Figure 4.10: Projection of all the data onto the maximally discriminant subspace yielded by PLDA shows that duplication, deletion, and control carriers form distinct clusters. The classifier boundaries are drawn in black for (a) white matter, as well as (b) gray matter based on distance to the nearest centroid

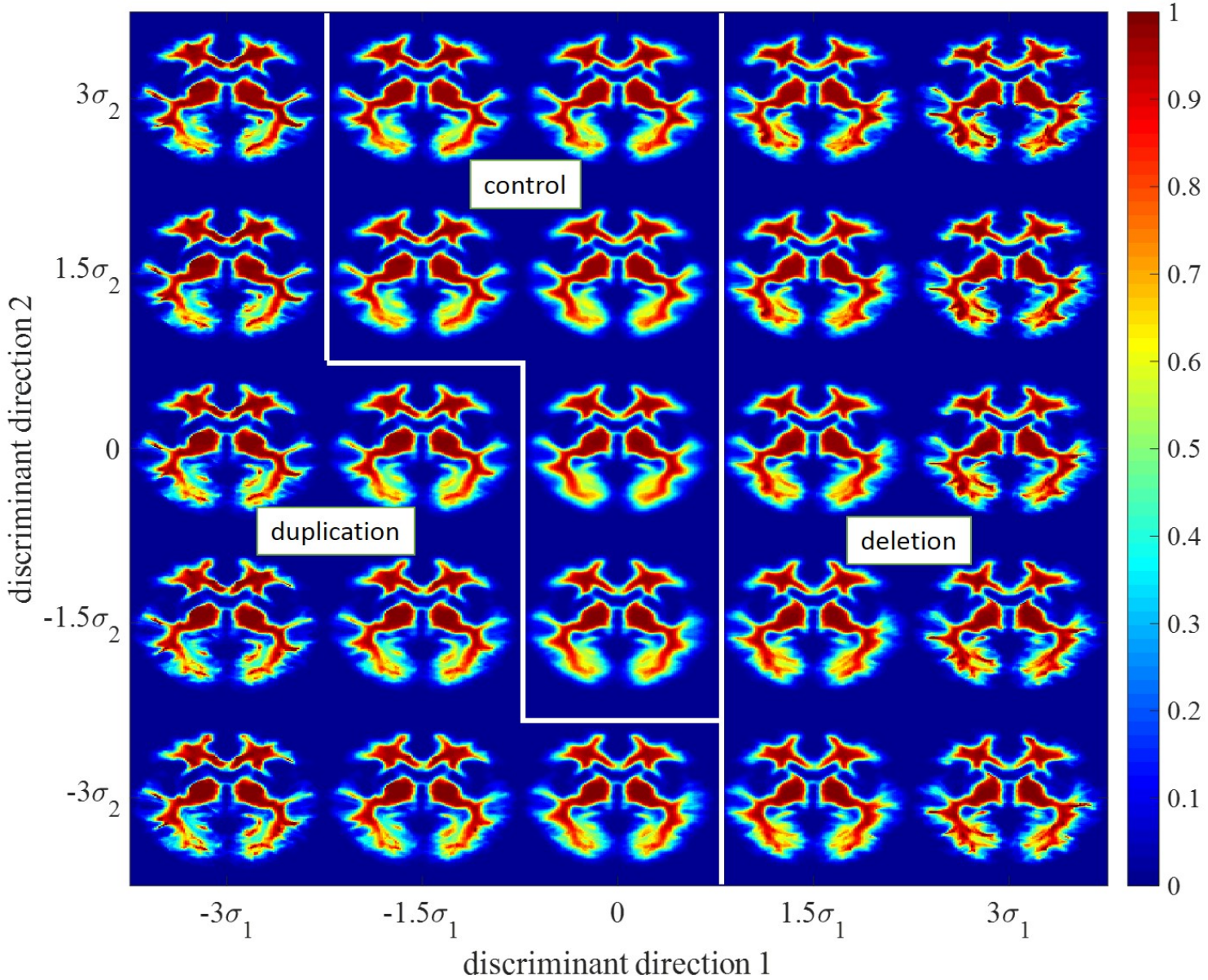


Figure 4.11: A single axial slice of computer-generated white matter images corresponding to the sampled discriminant space shown in Figure 4.10a. These images can then be interpreted to understand the differences that enable high classification accuracy.

Note that the Discriminant direction 1 and Discriminant direction 2 indicated in Figure 4.10 refer to different sets of directions for gray matter and white matter.

Visualizing Discriminating Differences

In Figures 4.12 and 4.13, we show the axial slices that best summarize the white matter differences discovered among the groups. In slice 85, we see a characteristic increase in white matter density in the superior cerebral (frontoparietal) white matter (deletion carriers $>$ controls $>$ duplication carriers). Slices 73 and 63 show changes in occipital white matter density, including the splenium of the corpus callosium (deletion carriers $>$ controls $>$ duplication carriers). Slice 53 shows differences in the inferior temporal (duplication carriers $>$ controls $>$ deletion carriers) and inferior frontal (deletion carriers $>$ controls $>$ duplication carriers) regions, while the brainstem is unchanged. In slice 47, the superior vermis of the cerebellum is affected (deletion carriers $>$ controls $>$ duplication carriers). Finally, in slice 39, we see that the cerebellar hemispheric white matter is affected such that (deletion carriers $>$ controls $>$ duplication carriers).

As for the gray matter differences shown, Figure 4.14 shows the representative axial slices of gray matter that summarize the discriminating differences between the subjects. Slice 55 shows increased gray matter density in the orbital frontal regions (deletion carriers $>$ controls $>$ duplication carriers). Furthermore, there are changes in gray matter density in the dentate nucleus of the cerebellum (duplication carriers $>$ controls $>$ deletion carriers). Slice 81 shows changes in gray matter density in the insular regions (deletion carriers $>$ controls $>$ duplication carriers) as well as in the temporal parietal junction (deletion carriers $<$ controls $<$ duplication carriers). Furthermore, the posterior cingulate shows changes as well (deletion carriers $>$ controls $>$ duplication carriers). Finally, slice 90 shows that gray matter density in the inferior perirolandic areas are affected as well, (duplication carriers $<$ controls $<$ deletion carriers), with left showing more changes than the right.

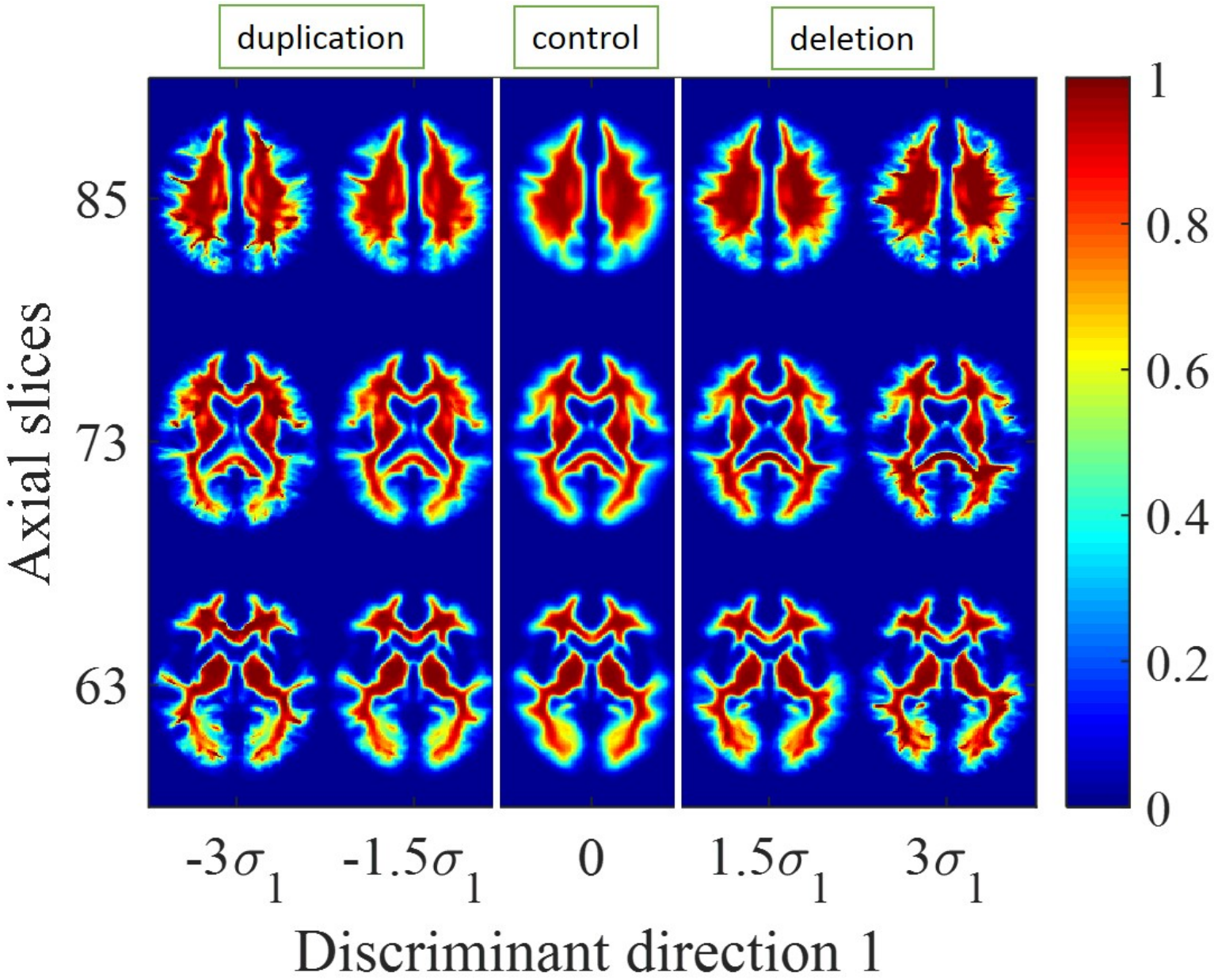


Figure 4.12: Images that show characteristic differences in white matter morphology

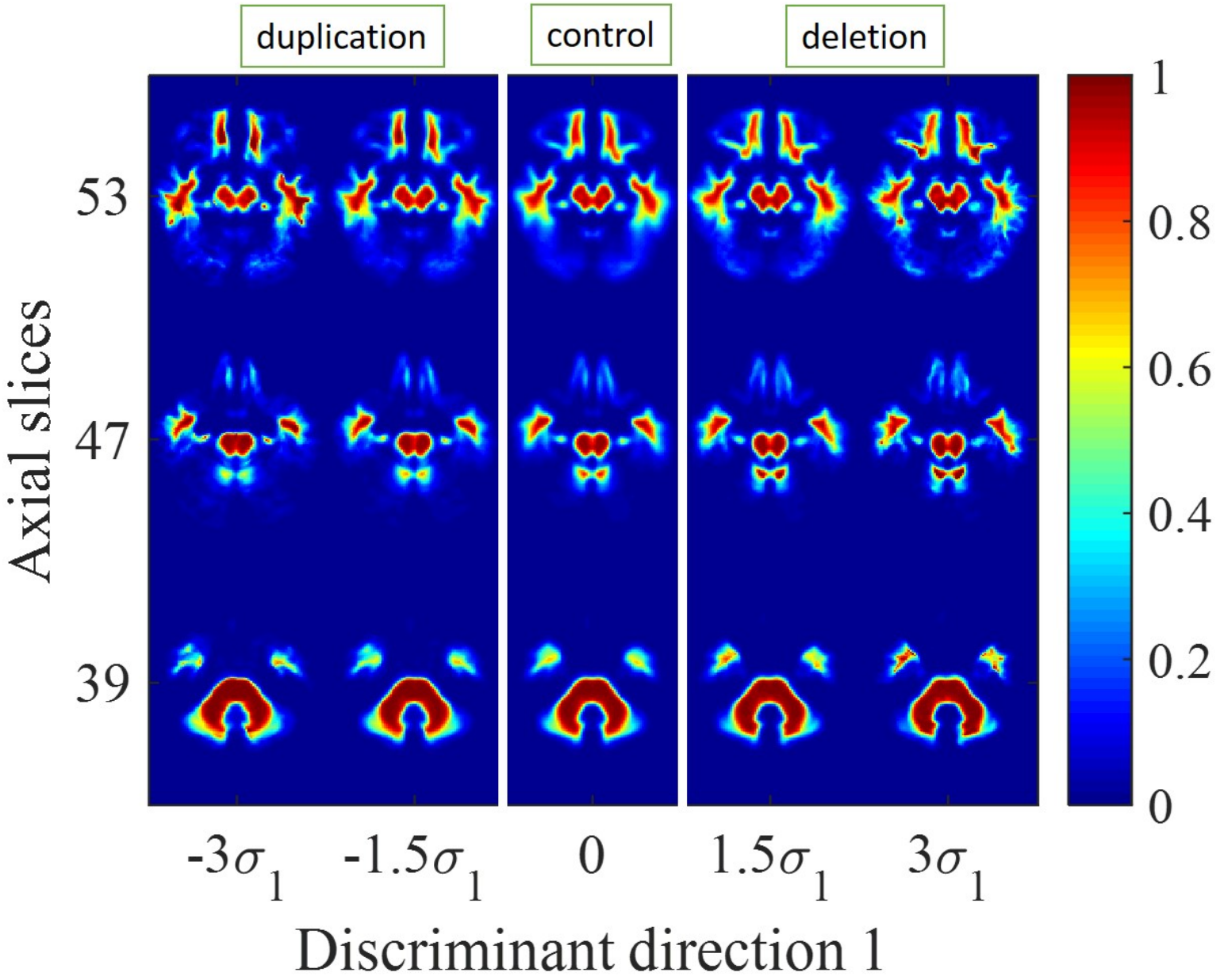


Figure 4.13: Images that show characteristic differences in white matter morphology

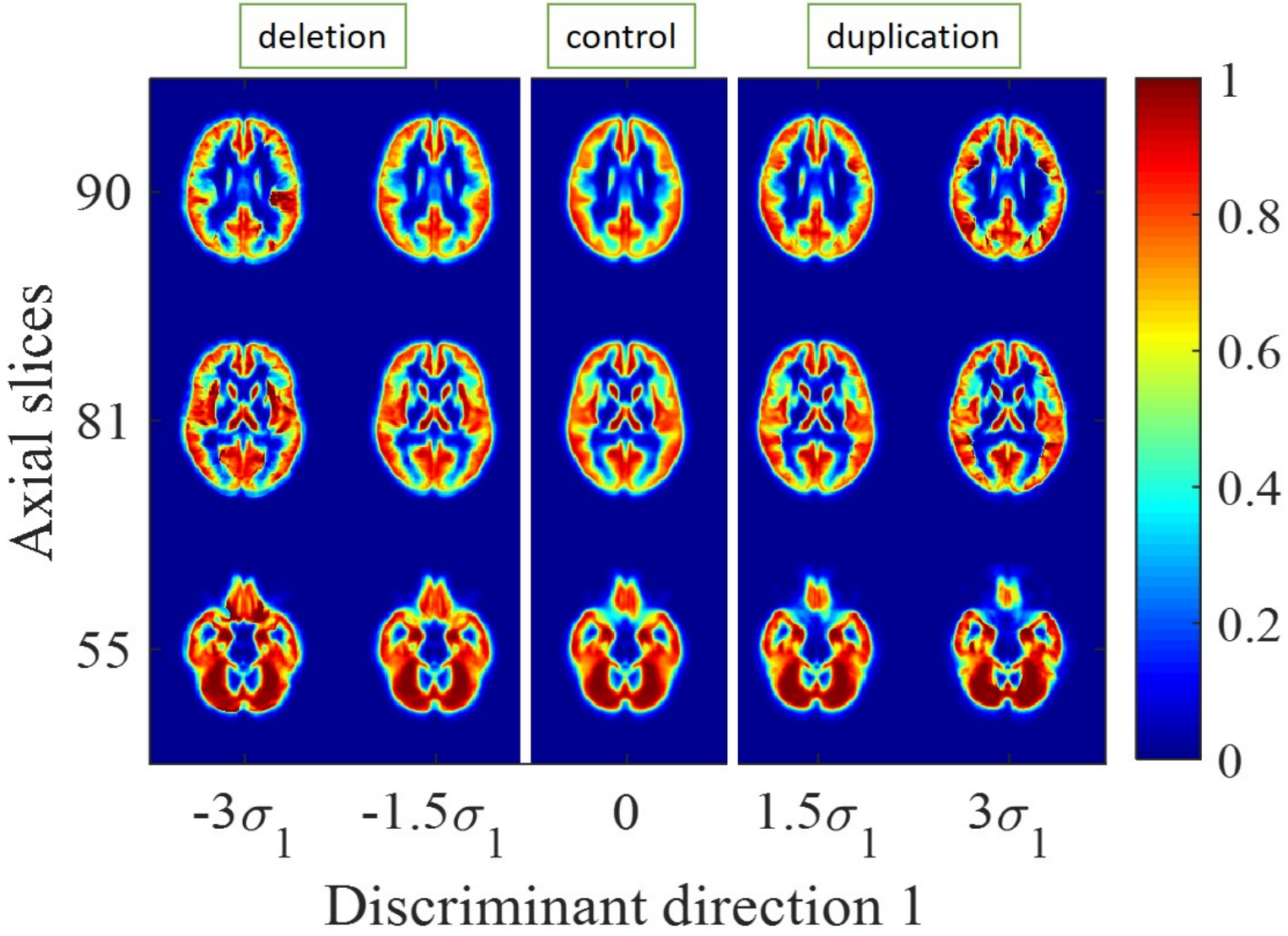


Figure 4.14: Images that show characteristic differences in gray matter morphology

4.3.4 Discussion and Summary

It is known that copy number variations in the 16p11.2 chromosomal locus (BP4-BP5) are associated with a range of neurodevelopmental and neuropsychiatric disorders, and this is the first time that discriminating differences in brain structure have been investigated explicitly. Using the TBM framework for discrimination, we are able to achieve 96% accuracy for white matter and 92.8% accuracy for gray matter in blind predictive studies where the task is to determine whether the subject is a control subject, a duplication carrier, or a deletion carrier. Furthermore, characteristic gray matter and white matter morphology differences that enable sensitive classification are visualized for the first time. This study establishes a direct link between genotype and brain phenotype, which is the first step toward understanding the role of genetics in disorders that involve brain and behavior.

4.4 Summary

In this chapter, we focus on the role that Transport-Based Morphometry can play in facilitating tasks of discrimination and visualization. Constructing a simple linear discriminant in the transport domain, we are able to predict OA 3 years in advance of symptoms with 86% test accuracy. The TBM approach demonstrates extensive subclinical damage in the cartilage of those who progress to osteoarthritis, which is normally obscured by heterogeneous normal variations in cartilage texture.

Given that copy number variations in the 16p11.2 chromosomal locus (BP4-BP5) are associated with a range of developmental and psychiatric disorders, TBM also allowed us to perform a predictive study investigating the influence of genetics on brain structure. We establish that there is a verifiable link between brain structure and 16p11.2 CNVs that enable up to 96% test accuracy in blind predictive studies.

We note that both of these problems have previously been studied with traditional computer-

aided techniques, which were not able to fully close the gap in discovering and visualizing discriminating structural differences. The results reported in this chapter are state of the art for the respective problem areas, both in classification accuracy as well as the ability to visualize the underlying discriminating differences. In the future, TBM has potentially numerous applications in early detection and establishing gene-structure-behavioral relationships with discrete variables.

Chapter 5

TBM for Regression and Correlation

5.1 Overview and Goals

In the previous chapter, we demonstrated how the TBM framework can be used to identify and visualize discriminating differences from MRI. However, not all clinical data can be summarized by discrete variables. In fact, most clinical measurements are continuous. For example, from weight to blood pressure, to enzyme levels and coma scores, we live in a continuous world. Transport-Based Morphometry defines an image transformation framework that can also be used to perform regression tasks. Specifically, we may seek the statistical relationship between a patient's dynamic anatomy and a certain clinical measurement. For tasks such as these, we need to define how regression and correlation experiments can be performed in the transport space. We demonstrate through experimental results on two open problems that TBM can discover the structural changes behind the clinical symptoms.

Mild Traumatic Brain Injury: Cognitive deficits are predictive of poor prognostic outcomes after mild traumatic brain injury, but the underlying structural substrates are poorly understood. Can the TBM technique illuminate a set of white matter tracts on imaging that interact dynamically to produce post-concussive cognitive difficulty?

Effects of Aerobic Exercise on Older Adult Brains: Aerobic fitness has many positive effects

on all aspects of health, and the brain is no exception. Larger cortical and subcortical volumes have been reported in the brains of people who are more aerobically fit, but the tissue characteristics that underlie these changes have never been studied. Can TBM help assess what effect aerobic fitness has on brain tissue distribution in older adults? How do the areas in the brain affected by aerobic fitness compare to the areas that are affected by normal aging?

Here, we add new insights and knowledge in problems related to uncovering how interconnected structural changes result jointly in clinical measurements of interest. What we find is that TBM enables the state of the art in these open problems.

5.2 Assessing Post-Concussive Reaction Time from Diffusion Tensor Imaging

5.2.1 Introduction

Cognitive changes are common after mild traumatic brain injury (mTBI) and predictive of poor post-concussive outcome [4, 19, 66, 67, 77, 140]. However, the underlying structural substrates are poorly understood. The mechanism mediating mTBI-related damage is diffuse axonal injury (DAI) caused by unequal shear-strain forces during injury [99]. DAI is responsible for transient decline in cognitive functions including processing speed, attention and working memory; it may also contribute to persistent post-concussive symptoms [99]. Advanced imaging with diffusion tensor imaging (DTI) holds promise to better detect white matter injuries underlying post-concussive cognitive changes. Unfortunately, the patterns specific to post-concussive cognitive deficits are not easily assessed by visual inspection because they are complex and obscured by other variations (Figure 5.1). However, identifying any patterns of relationship would represent a significant advance in understanding the underlying neurobiology of post-concussive cognitive deficits and may provide an objective assessment of post-concussive sequelae in the future. We investigate

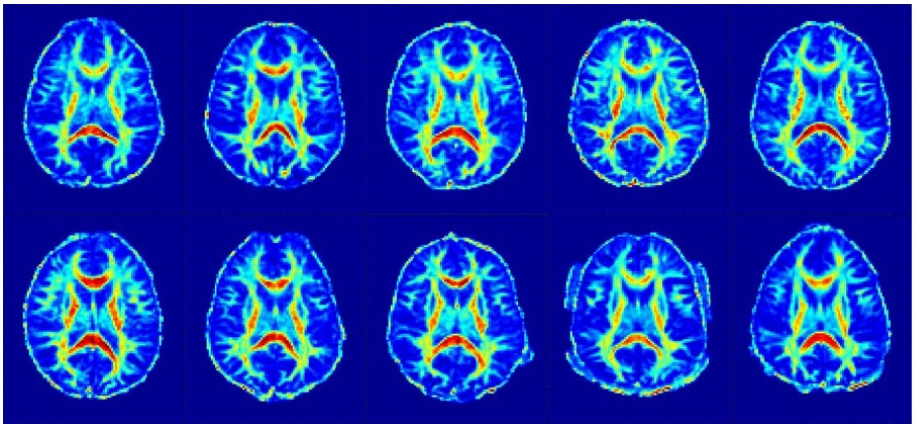


Figure 5.1: FA maps of 10 different post-concussive patients. The same axial slice is shown for patients with the best reaction time percentiles and worst reaction time percentiles. The FA maps corresponding to the best and worst reaction times are not easily differentiated by visual inspection because the abnormal variations related to reaction time are obscured by heterogeneous variations due both to normal and abnormal processes.

computer-aided techniques to discern the structural substrates that escape visual detection.

Traditional computer-aided approaches [99, 140] rely on human input for selecting which variations to examine. Ad hoc selections prevent exhaustive analysis of morphologic patterns. Often, not enough is known about the structural substrates a priori to inform prior input. Thus, a trial and error based approach is common. Typical methods include testing size, volume, or thickness of pre-specified regions of interest (ROI) to test association with post-concussive symptoms. However, testing features sequentially is a tedious process and spatially diffuse patterns are missed, including relationships across multiple regions of the brain [56]. Other approaches involve pixel-wise comparison after aligning images using registration. While registration can capture brain shape variations, it cannot assess variations in tissue texture or topology. Furthermore, the results of pixel-wise analysis depend on the registration algorithm used [24], and it is difficult to ensure structural and functional alignment. Therefore, traditional approaches that compare images through ROI volumes or pixel-wise intensities miss variations that are not adequately assessed by these metrics. If variations are detected, ROI and pixel-wise approaches can at best localize them to specific brain regions but cannot visualize the dynamic influence of changes in volume or pixel

intensity on outcome.

The purpose of this study was to investigate white matter injury associated with post-concussive reaction time using the new Transport-Based Morphometry (TBM) [12, 85, 136] technique. Unlike traditional techniques, TBM quantifies all variations in brain morphology by comparing the spatial distribution of pixels in two images using the mathematics of optimal mass transport (OT) [131]. Unlike ROI or pixel-wise approaches, the OT-based metric does not lose information in measuring similarity between images and enables structural correlates to be assessed in a modern, data-driven manner based on machine learning rather than trial and error. Furthermore, unlike traditional approaches, TBM can recover the images enabling statistical association, which allows visualization of the anatomic changes. We hypothesize that the knowledge gap between brain structure and post-concussive reaction time can be closed using the TBM technique.

5.2.2 Methods

Patient cohort

In this retrospective study, Institutional Review Board approval was obtained and informed consent was waived. The electronic medical record was searched for DTI studies performed for mTBI between January 1, 2006 and March 1, 2013. Studies were identified using the keywords concussion, mild traumatic brain injury, and diffusion tensor imaging. Included subjects met the following criteria: 10 to 50 years of age, witnessed closed head trauma, no focal neurologic deficit, loss of consciousness of $<$1 minute, posttraumatic amnesia of $<$30 minutes, and English language proficiency. Subjects were excluded for the following reasons: prior neuropsychiatric illness (2 patients), abnormal CT or conventional brain MR imaging findings (3 patients), history of substance abuse (3 patients), lack of DTI (4 patients), lack of neurocognitive assessment (6 patients), total symptom score of zero (3 patients), and inability to align FA images (2 patients). The final patient cohort consisted of 64 total subjects. Demographics are summarized in Table 5.2 and causes of

Table 5.1: Causes of mTBI in subject cohort

Sports injury (%)	46(65.6%)
Motor vehicle accident (%)	6(9.4%)
Bicycle accident (%)	2(3.1%)
Fell down stairs (%)	2(3.1%)
Fell off horse (%)	2(3.1%)
Other (%)	10(15.6%)

Causes of concussive injury in patient cohort

injury in Table 5.1. Based on the time from injury, most patients in the study were in the sub-acute phase of concussion.

Neuropsychological examination was performed by a neuropsychologist with $>$15 years of experience in treating patients with mTBI. Reaction time was determined as part of a battery of computerized neurocognitive tests in the Immediate Post-Concussion Assessment and Cognitive Testing (ImPACT) test [92]. Reaction time for each subject is a percentile based on normative data from baseline testing of over 17,000 athletes as part of pre-sport participation. Percentile information accounts for both sex and age.

MRI acquisition

All subjects underwent an identical imaging protocol on the same system despite the large time range of the study. Conventional MR imaging and DTI were performed with a 1.5T unit (Signa; GE Healthcare, Milwaukee, Wisconsin) and standard head coil. DTI was performed with a single-shot echo-planar sequence (TR 4000 ms; TE 80 ms; NEX 2; section thickness, 5 mm; 128$\times$128 matrix; FOV, 260 mm). Diffusion gradients were set in 25 noncollinear directions by using 2 b-values (b=0 and 1000 s/mm^2).

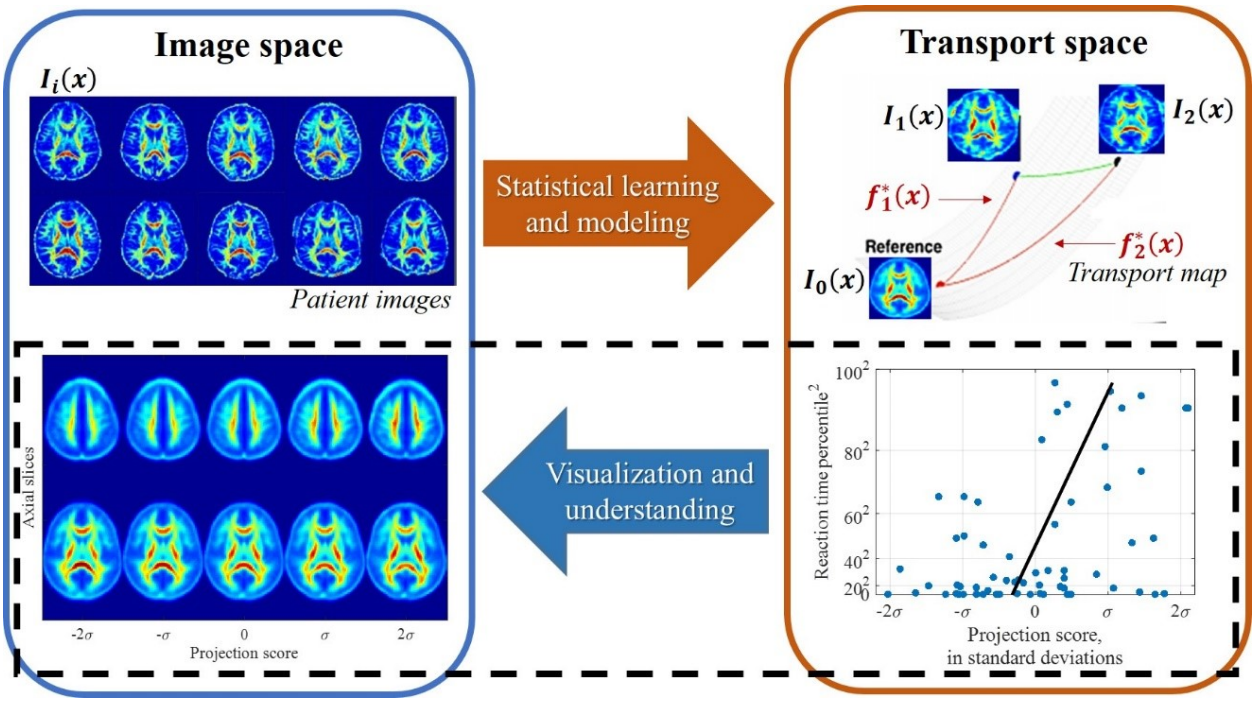


Figure 5.2: System diagram. Transport-Based Morphometry transforms images from a representation in image space ($I_1(x)$ and $I_2(x)$) to an equivalent representation in transport space ($f_1^*(x)$ and $f_2^*(x)$) with respect to a common reference ($I_0(x)$). The representation in transport space facilitates discovery of trends by statistical learning algorithms. Modeling functions constructed in the transport space can be inverted to directly visualize the corresponding clinically interpretable images.

Registration, Segmentation, and Preprocessing

Fractional anisotropy (FA) maps were generated as a measure of white matter integrity using the Brain Diffusion Toolbox [14, 125], which is part of the DTI pipeline of the Functional MRI of the Brain Functional Software Library (FSL) [54, 75]. The FA maps included both gray and white matter and were registered to the Montreal Neurological Institute atlas by using a 12-parameter affine transformation. A study population mean FA map was computed by taking the Euclidean average of all images in the dataset and served as the common reference image.

Analyzing spatial FA distribution using Transport-Based Morphometry

The Transport-Based Morphometry (TBM) [12, 85, 136] technique analyzes FA spatial distribution in each image by measuring its similarity to a common reference image using the mathematics of optimal mass transport (OT) [131]. The variations in FA spatial distribution compared to the reference are quantified in a transport map for each subject. A transport map is a vector field that describes how to transform FA values in each image to match the common reference. The transport map generated by OT is one that minimizes the amount of FA transport, defined as FA value over translated distance, to match the common reference. Figure 5.3 shows FA maps and corresponding transport maps for four sample subjects.

Transport maps facilitate discovery of trends in the data by representing variations more sparsely than in the original image space. Specifically, many complex and nonlinear variations in brain images that challenge humans and computers can be represented as simpler, linear relationships in transport space. Thus, statistical learning on the transport maps can assess complex brain morphology more effectively.

Another key property of TBM is that it guarantees that the mapping from image to transport map is bijective. Invertibility of TBM ensures that variations and trends discovered based on transport maps can be visualized and interpreted directly as images. We use the term image domain representation to refer to the original image and transport domain representation to refer to the corresponding transport map.

Equations (3.2) and (3.3) define the analysis and synthesis equations for the forward and inverse transforms, respectively. All TBM code was implemented in MATLAB (MathWorks, Natick, MA).

Assessing statistical relationship with reaction time

Statistical learning is performed in the transport domain to streamline extraction of image features. First, transport maps are vectorized and concatenated into a standard data matrix $D \in \mathbb{R}^{p \times n}$,

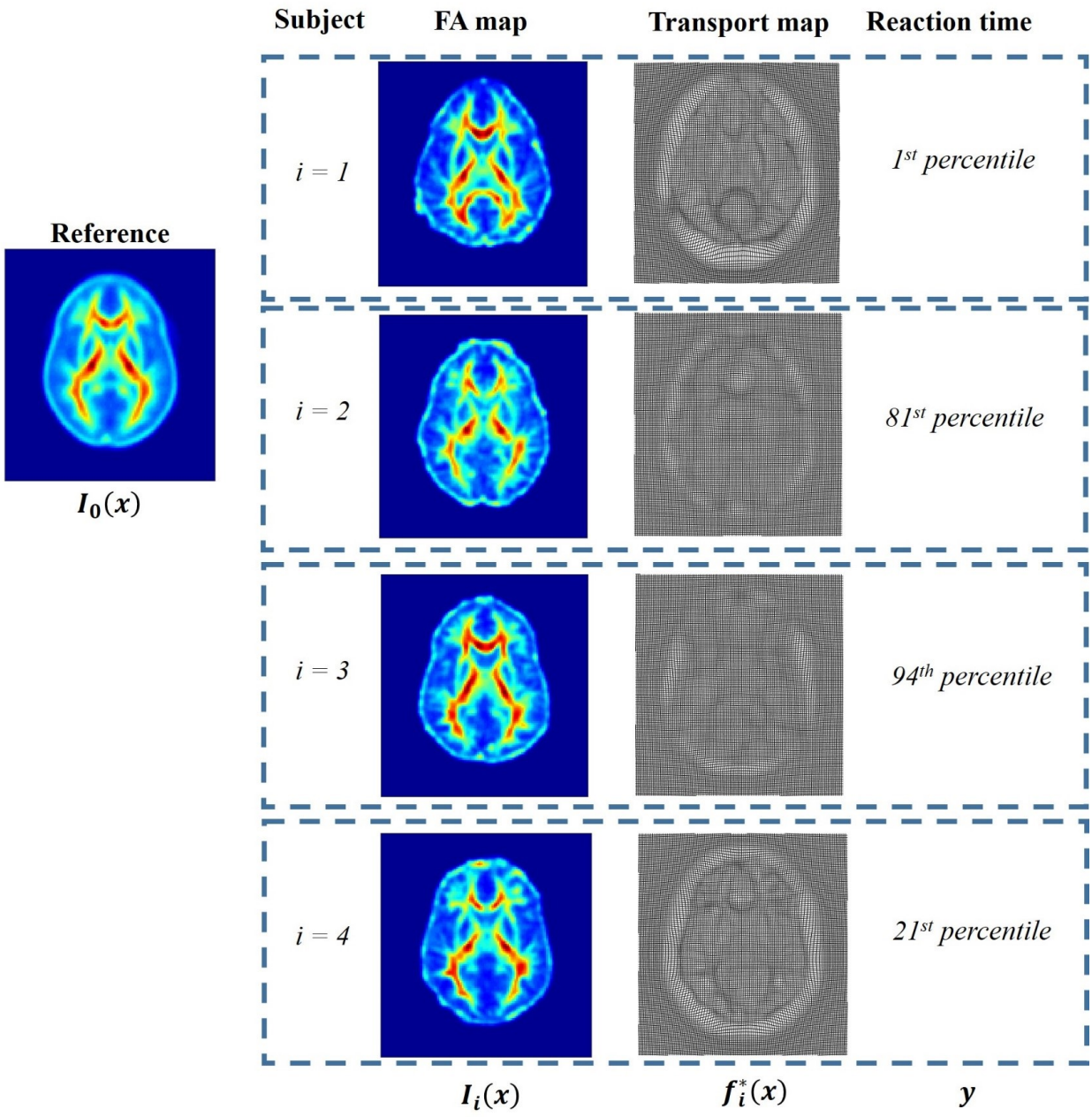


Figure 5.3: Sample FA maps and corresponding transport maps computed with respect to the common reference for four different subjects. Only one axial slice is shown from the 3D stack.

where p is the dimension of each transport map and n is the number of subjects. Then, the standard principal components analysis (PCA) technique (3.28) is used to remove the data dimensions containing little contribution to the overall variance in the dataset. The data matrix is projected onto the topmost d eigenvectors associated with 90% of the variance in transport maps to create the reduced-dimension data matrix $X \in \mathbb{R}^{d \times n}$ $(d << p)$. Let $v \in \mathbb{R}^{n \times 1}$ be reaction time percentile mapped to squared values (to increase separation of lower from higher percentiles), either corrected or uncorrected for age as a covariate. The direction in transport space is computed that results in the strongest linear correlation with reaction time percentile v (Equation (5.1)).

$$w_{corr} = argmax_w \frac{w^T X v}{\sqrt{w^T w}} = \frac{Xv}{\sqrt{v^T X^T X v}} \tag{5.1}$$

Here, w_{corr} is another vector field in the transport domain that quantifies the direction and amount of FA re-distribution that is most correlated with reaction time percentile. These changes can be inverted using Equation (3.3) and visualized as generated FA images.

The effects of covariates are removed according to Equation (5.2).

$$v = y - Z(Z^T Z)^{-1} Z^T y \tag{5.2}$$

Here, v describes the residual from reaction time percentile ($y \in \mathbb{R}^{n \times 1}$) that is decorrelated and orthogonal to the c confounding variables in $Z \in \mathbb{R}^{n \times c}$. This study corrects for age as a covariate.

Statistical significance of the computed direction was assessed using permutation testing with T = 1000 tests. The p-value reported is the fraction of times over T tests that Pearson's correlation coefficient was higher when labels v are randomly assigned to the subject data compared to the original labels.

All statistical analysis code was implemented in MATLAB (MathWorks, Natick, MA).

Table 5.2: Subject demographics

Age (years)	$10 - 28(16.1 \pm 2.9)$
Number Male (percent)	$42(65.5)$
Time from injury to imaging (days)	65 ± 124

Demographic characteristics of subject cohort

5.2.3 Results

Assessing statistical relationship with reaction time

The direction in transport space most correlated with reaction time percentile was statistically significant (Pearson correlation coefficient 0.44, p = 0.004). After correcting for the effect of age as a covariate, the computed direction remained statistically significant (Pearson correlation coefficient 0.39, p = 0.015). There appears to be a positive correlation between FA spatial distribution and reaction time percentile.

Figure 5.4a is the scatter plot of the patient data when it is projected onto the most correlated direction computed in the transport space. Each point in the scatter plot represents a brain image in the dataset. According to the linear regression model constructed in the transport space, increasing reaction time percentile (faster reaction times) is associated with increasing projection scores and decreasing reaction time percentile (slow reaction times) is associated with decreasing projection scores. The mean image $I_0(x)$ maps to a projection score of zero.

Visualizing statistically significant associations

In addition to investigating whether a statistically significant model exists between FA spatial distribution and reaction time percentile, projection scores have direct clinical interpretation through inverse TBM transformation. Figure 5.4b shows visualizations of brain morphology changes that correspond to the different projection scores. It is important to note that the linear relationship with reaction time percentile in the transport domain maps to complex, nonlinear white matter injury

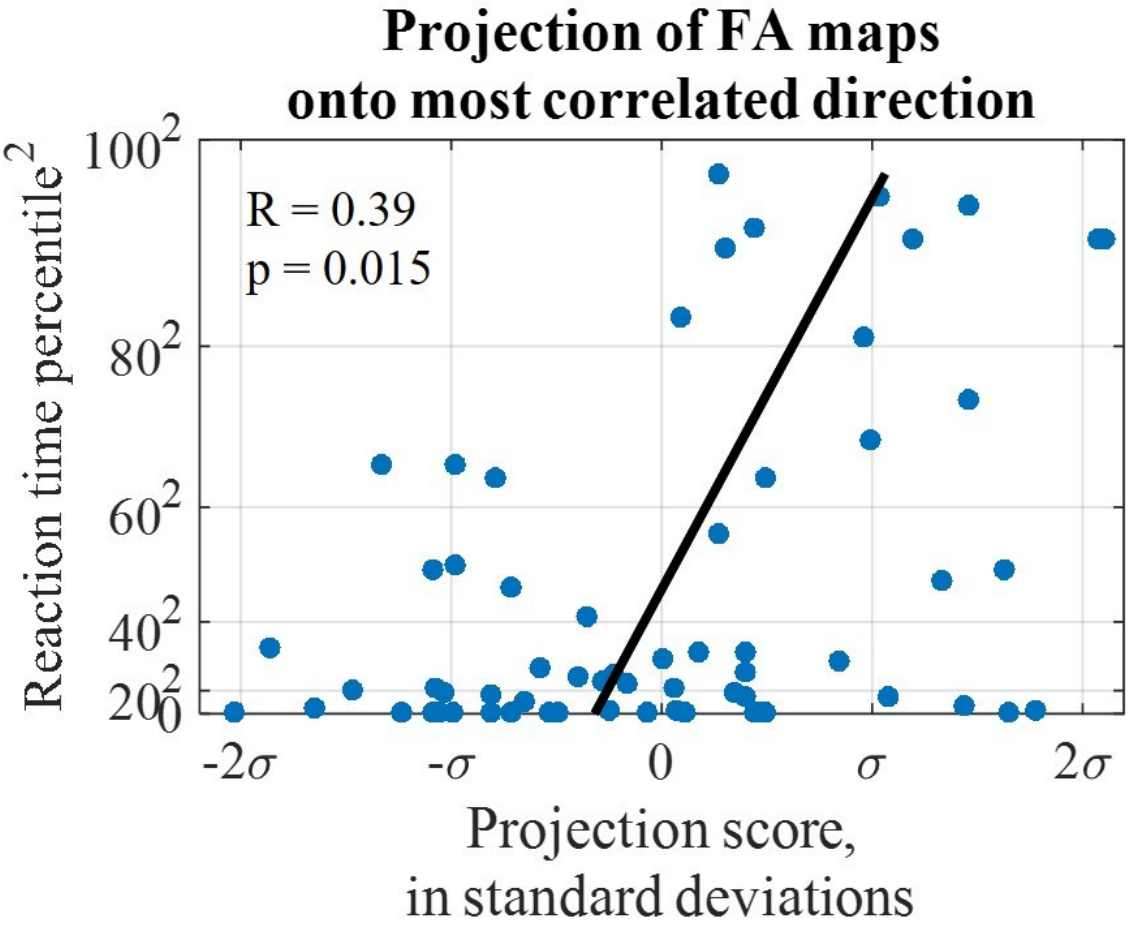


Figure 5.4: Most correlated direction. (a) scatter plot showing projection of FA maps onto most correlated direction.

patterns in the image domain.

The projection scores indicated by the horizontal axis in Figure 5.4b correspond to those in the vertical axis of Figure 4a. The images shown represent points sampled along the line of best fit for which projection score ranges from -2 to $+2$ standard deviations from the mean. Two axial slices are shown that best summarize the morphology differences from low reaction time percentile to high reaction time percentile. The population mean image $I_0(x)$ is indicated by projection score zero in Figure 5.4b. The images at negative projection scores correlate most strongly with low reaction time percentile, while the images at positive projection scores correlate most strongly with high reaction time percentiles, as suggested by Figure 5.4a. As one progresses from low reaction time percentile to high reaction time percentile, the statistical model illustrates that FA in the corpus callosum decreases, and the FA in the optic radiations, corticospinal tract, and anterior thalamic tracts increases.

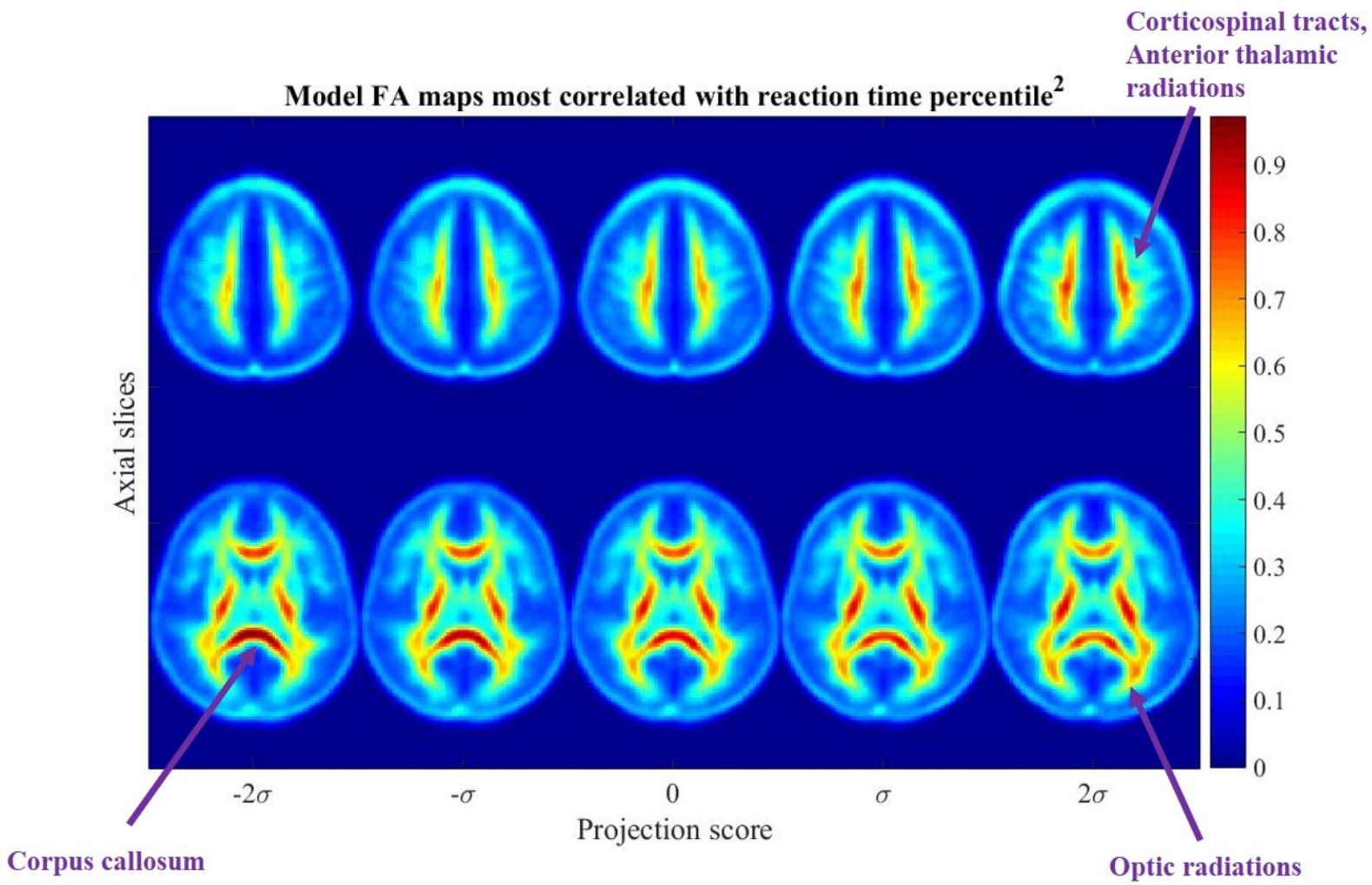


Figure 5.4: (b) images corresponding to most correlated direction in transport space show decreasing FA in the corticospinal tracts, anterior thalamic radiations, and optic radiations with low reaction time percentile corresponding to scatter plot. The FA in the corpus callosum increases as reaction time decreases.

Individual assessment of white matter injury associated with reaction time

In addition to identifying structural correlates of reaction time on a group level, TBM also facilitates identification of white matter injury patterns in individual patients. The direction computed in transport space provides a way to assess whether the patterns of white matter injury identified by Figure 5.4b are present in an individual FA image. Projecting the computed transport map onto this direction yields a projection score for each patient. The projection scores can be interpreted based on the relationship in Figure 5.4a to prognosticate whether cognitive deficits may be present. Figure 5.5 illustrates that the FA map of a patient with a fast reaction time in the 94th percentile maps to a high positive projection score and that of a patient with slow reaction time in the 21st percentile maps to a low negative projection score, providing a metric by which to differentiate these patients. Thus, a key strength of TBM is its ability to assess spatially distributed and complex white matter patterns that are not easily identified visually, as Figure 5.1 illustrates, or by an ROI-based perspective.

Furthermore, Figure 5.5 illustrates that the transport map can capture information not readily accessible from the FA maps. The determinant of the Jacobian is computed from the transport maps and measures the joint changes in FA concentration across multiple white matter tracts. Increase in FA concentration is indicated by value > 1 and decreases are indicated by value < 1. We see that the direction of FA change compared to the common reference in white matter tracts for a patient with high and low reaction time corroborate those identified by Figure 5.4b.

5.2.4 Discussion and Summary

This study investigates structural correlates of reaction time percentile in sub-acute post-concussive patients. The new Transport-Based Morphometry (TBM) [12, 85, 136] technique was used because it overcomes a major limitation of traditional approaches in visualizing dependent brain morphology changes associated with post-concussive outcome. Current techniques do not enable a full

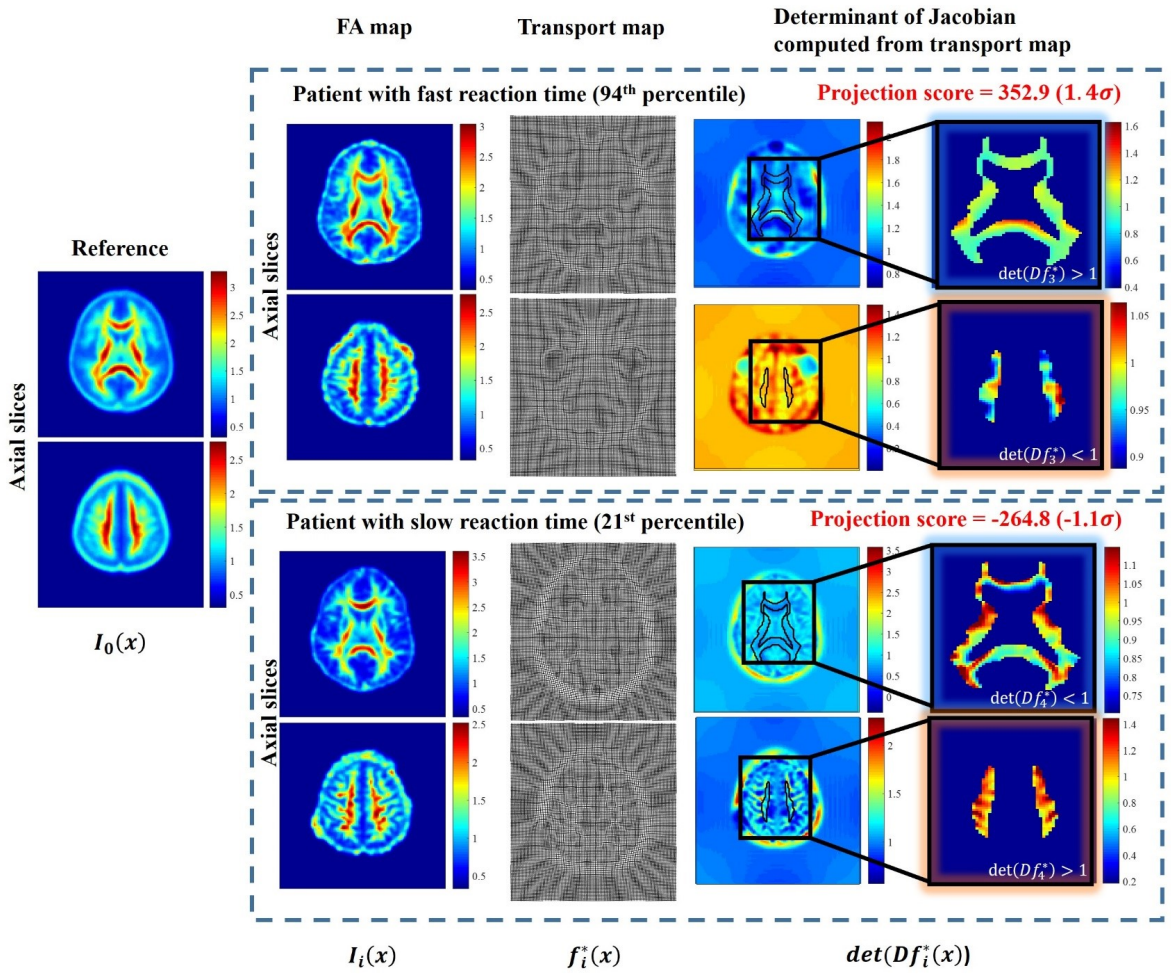


Figure 5.5: TBM can help assess cognitive deficits in individual patients by using the computed direction to assess whether the complex and spatially distributed patterns of white matter injury identified by Figure 5.4b are present. When transport maps are projected onto this direction, a projection score is yielded that can be used to differentiate patients according to the relationship in Figure 5.4a. The subjects shown correspond to i=3 and 4 in Figure 5.3. The direction of FA change with respect to the reference in individual patients corroborate the patterns identified in Figure 5.4b.

data-driven analysis free from human bias or easy visual interpretation of complex, interconnected white matter injury patterns found to be statistically dependent on reaction time percentile. A significant correlation was discovered between FA spatial distribution and reaction time percentile. The correlation remained robust after adjustment for age. A set of clinically interpretable FA images were generated through inverse TBM transformation illustrating patterns of white matter injury associated with a range of reaction time percentiles. Images reveal not only the location of abnormalities associated with post-concussive reaction time, but also the relationship across these regions to identify how bidirectional changes in FA affect reaction time changes. Thus, the study hypothesis is affirmed that TBM can overcome the limitations of traditional computer-aided methods to enable a broader analysis and produce clinically interpretable image findings.

Prior studies have suggested that certain metrics quantified from images may correlate with reaction time in post-concussive patients [99, 140], however, these study methods suffer several limitations. At the most basic level of investigation, histogram analysis does not provide regionally specific answers [58, 100]. ROI-based techniques are not exhaustive [100], and Voxel-Based Analysis (VBA) is only as accurate as the degree of structural/functional alignment [24, 58, 100]. By analyzing variations in FA spatial distribution, TBM addresses limitations of many current techniques. First, TBM is able to assess the interconnected nature of white mater injuries by analyzing their joint influence on clinical outcome. Second, the fully automated approach is exhaustive and free from human bias because statistical machine learning is used to discover morphology changes. Third, TBM allows the influence of white matter injury on outcome to be directly visualized, rather than simply localized. Unlike deformation-based approaches (i.e. Voxel-Based Morphometry [7] and Deformation-Based Morphometry [8]), TBM can analyze variations in both shape and texture simultaneously. Deformation-based approaches analyze shape differences only and in fact, deformation fields cannot match two images with error zero unlike the transport maps generated by TBM. Furthermore, TBM provides advantages in mixed-age patient populations, where differences in head size may confound traditional voxel-based techniques. Unlike voxel-based methods,

TBM quantifies FA spatial distribution and is therefore robust to the effects of head size influencing misalignment.

Our findings of abnormalities in the corpus callosum, optic radiations, corticospinal tract, and anterior thalamic radiations correlate well with the known anatomic circuits key to reaction time beginning with visual-spatial interpretation, executive attention, interhemispheric transmission, and finally, response selection. Beginning with the initial visual spatial interpretation, the optic radiations are key to interpreting the initial stimulus and could contribute to inter-individual variability in reaction time [127]. Next, frontally-mediated supervisory circuits related to attention focus help to process the observed stimulus, centered in the anterior thalamic radiations. Anterior thalamic radiations have been implicated as the major frontalsubcortical circuit mediating cognitive impairment [110]. Indeed, damage to the anterior thalamic radiations impairs cognitive function [25] and is associated with reaction time deficits [110]. Executive outcome is translated into a physical response through interhemispheric transmission of the data. Integrity of the corpus callosum, a key mediator of interhemispheric transmission, is correlated with reaction time in both normal and cognitively impaired individuals [3]. Finally, corticospinal tract activation generates a motor response as the final step in cortical processing of stimulus interpretation and response selection [86].

TBM reveals that the direction of FA changes with reaction time is different among the corpus callosum and visual tracts, corticospinal tracts, and anterior thalamic radiations. The sub-acute time period during which imaging was performed may be responsible for this finding. White matter injury in mTBI is a dynamic pathological process, demonstrating increased FA early, presumably related to cytotoxic edema, and decreased FA late, reflecting decreased white matter integrity and axonal loss [44]. Our finding of mixed increased and decreased FA likely reflects a transitional time period, in which some regions of injury have progressed to axonal loss while others still remain in the earlier stages of edema and swelling. This mixed response has been seen in both the heterogeneity of findings during the semi acute interval, [13, 71, 91, 94] as well as within

individual studies performed during this time period [118].

There are several limitations of this study. We conducted a retrospective, single center study with only 64 patients. Future multi-center studies are needed to build on our results. Additionally, our patients ranged in age from pediatric to adult. Although aging-related changes in reaction time have been reported [42, 43], our model remained significant even after accounting for age as a covariate. Furthermore, as many mTBI patients do not seek medical attention, there is a possibility of selection bias towards more seriously injured or symptomatic patients. However, this bias may help direct towards the patients that would benefit most from imaging biomarkers. Finally, quantitative use of DTI metrics in the clinical setting is currently limited by the lack of universally accepted normative data. Demonstrating injured regions associated with cognitive impairment is the first step towards an individual diagnostic tool.

We demonstrate that it is possible to bridge the knowledge gap between brain structure and clinical outcomes in post-concussive injury using Transport-Based Morphometry. In the future, the application of TBM to analyze DTI imaging in prospective longitudinal studies may help provide prognostic markers to assess either recovery or progression to chronic stages [58, 100].

5.3 Aerobic Fitness and Brain Tissue Distribution in Older Adults

5.3.1 Introduction

Despite the apparent ubiquity of age-related changes in brain morphology and function including atrophy and cerebral shrinking, there remain several promising approaches for mitigating these declines and promoting brain health in late life. Aerobic exercise interventions remain one of the most promising approaches. For example, 12-month exercise interventions increase the size of the hippocampus in older adults [48, 98, 124], a brain region susceptible to age-related losses in size and associated with conversion to dementia. Also, 6 and 12-month interventions increase

the size of the prefrontal cortex [35] and increase its functional connectedness with other areas [29, 36, 133]. Unfortunately, exercise interventions are often expensive and challenging, leading many to use cross-sectional markers of exercise participation including measures of actigraphy and aerobic fitness. Along these lines, several studies have demonstrated that cross-sectional associations between aerobic fitness and brain morphology and function closely overlap with those that are found in randomized interventions [34, 46, 47, 101, 130, 137, 139].

The brain morphology differences associated with aerobic fitness are robust, but there remains many unanswered questions about the relationship between morphology and the underlying tissue characteristics that account for volumetric differences in older adults. Unfortunately, traditional computer-aided techniques have not been adequate to provide answers to some of these questions as they suffer from well-known limitations in biological interpretation of the morphologic patterns. One limitation of traditional region-of-interest (ROI) approaches is that they do not enable a simultaneous analysis of multiple brain regions. For example, a commonly-used approach is to measure the volume of pre-specified ROIs (e.g., hippocampus) and test which of these pre-specified regions are association with aerobic fitness levels. There are certainly benefits of this approach as it requires a theoretical position about regions that may and may not be associated with the variable of interest. However, this is also a potential limitation - ROI approaches are only able to test directed hypotheses and could potentially miss spatially diffuse effects happening outside the ROI. On the other hand, voxelwise comparison techniques (e.g., Voxel-Based Morphometry, Deformation-Based Morphometry, Tensor-Based Morphometry) provide a whole-brain analysis of associations and do not require a priori assumptions about particular regions associated with the variable of interest (i.e., fitness), they do require proper alignment to a standard template in order to produce meaningful results [24]. This has the potential of inflating or masking statistical patterns testing associations between morphology and fitness. Furthermore, voxel-based approaches using non-rigid registration techniques can capture variations in brain shape and volume, but not those related to tissue topology or texture. Thus, traditional approaches have provided important

information about the role of fitness in brain aging, but have been critiqued based on some of the methods used to assess morphology.

This study investigated the association between aerobic fitness and brain morphology in older adults using the Transport-Based Morphometry (TBM) technique [12, 85, 136], which overcomes some of the limitations of current ROI and voxel-based methods and also enables visual interpretation of fitness-related associations with brain morphology. The TBM technique is fully automated and uses T1-weighted brain images. TBM produces visualizations that illustrate the differences in brain morphology most strongly associated with aerobic fitness levels (if one exists). Furthermore, TBM addresses some of the limitations of traditional morphometry techniques by enabling simultaneous analysis of the entire brain on a voxel-wise basis while avoiding the use of registration techniques typically employed in traditional voxel-based approaches. TBM utilizes shape and texture information simultaneously when providing an assessment of whether the variable of interest (e.g., fitness) is related to morphology. Using the TBM technique, we examined whether higher aerobic fitness levels would be associated with differences in gray and white matter tissue distribution across the brain. Based on prior studies of brain morphology and fitness we predicted that higher fitness levels would be associated with differences in tissue distribution in prefrontal and temporal brain regions.

5.3.2 Methods

Participant characteristics

In this cross-sectional study, we recruited 172 healthy older adult subjects, and an informed consent approved by the University of Illinois Institutional Review Board was obtained for each subject. Subject demographics are summarized in Table 5.3. Inclusion criteria included being physically inactive as defined by engaging in $\leq$ 30 min of exercise each week in the prior 6 months [48], demonstrated strong right handedness determined by score of $\geq 75\%$ on the Edinburgh Handedness

Questionnaire [102], score ≥ 51 on modified Mini Mental Status Examination [120] to rule out clinical cognitive impairment, normal color vision, visual acuity of at least 20/40, and no history of neuropsychiatric conditions or neurological diseases or infarcts including Parkinsons disease, multiple sclerosis, Alzheimers disease, or stroke. Subjects were excluded from the study on the basis of having a score greater than 3 on the Geriatric Depression Scale [116], history of Type II diabetes and cardiovascular disease, evidence of chronic inflammation (i.e., severe arthritis, psoriasis, inflammatory bowel disease, asthma, polyneuropathies, Lupus), and metal implants or contraindications for MRI.

MRI Acquisition

High-resolution T1-weighted images of the brain were collected on a 3T head-only Siemens Allegra MRI scanner. Images were acquired using a 3D MPRAGE (Magnetization Prepared Rapid Gradient Echo Imaging) protocol in which 144 contiguous axial slices collected in ascending fashion parallel to anterior posterior commissures. Scan parameters were as follows: echo time $(TE) = 3.87$ ms, repetition time $(TR) = 1800$ ms, field of view $(FOV) = 256$ mm, acquisition matrix 192×192, slice thickness$= 1.3$ mm, flip angle$= 8^{\circ}$ [132].

Aerobic fitness assessment

All participants in the study obtained required consent from their personal physician before cardiorespiratory fitness testing was conducted. Graded maximal exercise testing on a motor-driven treadmill was used to assess aerobic fitness (VO_2 max). Each participant walked at a speed slightly faster than their normal walking pace (approximately 30–100m per minute) with increasing grade increments every two minutes of 2%. A cardiologist and a nurse continuously monitored measurements of heart rate, blood pressure, and oxygen uptake. Oxygen uptake (VO_2) was measured from expired air samples taken at 30-second intervals until a maximal VO_2 was attained or to the point of test termination due to symptom limitation and/or volitional exhaustion. VO_2 max was

defined as the highest recorded VO_2 value when two of following three criteria were satisfied: (1) a plateau in VO_2 peak between two or more workloads; 2) a respiratory exchange ratio >1.00; and (3) a heart rate equivalent to their age predicted maximum (i.e. $220 - age$).

Registration, segmentation, preprocessing

T1-weighted images were co-registered using a 12-parameter affine transformation with respect to the native space of the first image, then segmented into gray matter (GM) and white matter (WM) channels using the Statistical Parametric Mapping (SPM12) software [52]. Images were then smoothed using a Gaussian filter with $\sigma = 1$, and intensities were normalized to sum to 1 in order to enable analysis of T1 spatial distribution. The Euclidean average of all GM and WM images were taken separately in order to generate study population sample mean images that served as the common reference for each channel.

Analyzing brain tissue spatial distribution using Transport-Based Morphometry

Gray matter and white matter were analyzed separately. The Transport-Based Morphometry (TBM) [12, 85, 136] approach described by Kundu *et al* [85] was used to measure similarity between each subject image and the common reference using the mathematics of optimal mass transport (OT) (Figure 5.6). A transport map was generated for both gray and white matter images that quantified the individual variation in T1 tissue distribution compared to the common reference image in native space. Transport maps are vector fields describing how to transform individual images in order to match the reference image. The transport maps generated by TBM are optimal in terms of minimizing T1 transport (amount of T1 intensity over distance moved). Each element of the transport map contains the coordinates to which the intensity value at the present location should be mapped.

The transport maps facilitate information extraction by quantifying variations from a mean image. Euclidean operations in the transport space correspond to complex and nonlinear morphologic

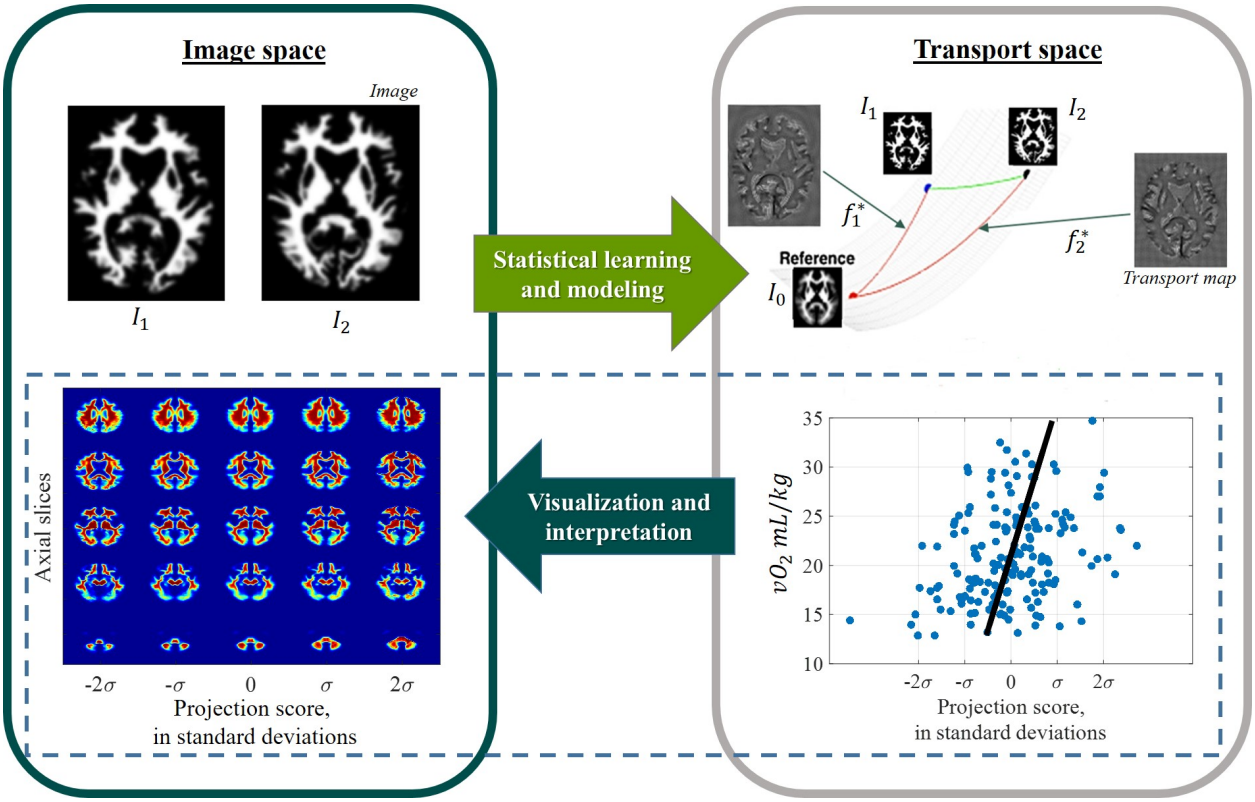


Figure 5.6: Transport-Based Morphometry system diagram. Images are mapped to corresponding transport maps through a unique, one-to-one transformation using optimal mass transport to facilitate statistical learning and modeling. Modeling functions computed in transport space can be inverted to visualize and interpret the corresponding images.

shifts in tissue distribution in the image space. Therefore, statistical analysis in the transport space facilitates discovery of dependent morphologic features.

Furthermore, another key property of TBM is that the mapping requiring the minimum mass transport is unique because TBM satisfies several constraints imposed by OT theory [27, 85]. OT theory ensures that TBM is an invertible transformation. Thus, statistical analyses performed in the transport space can be directly inverted to visualize the MR images corresponding to varying levels of aerobic fitness.

Equations for analysis (Equation (3.2)) and synthesis (Equation (3.3)) are described in Chapter 3. All TBM codes were implemented in MATLAB (MathWorks, Natick, MA).

Assessing statistical relationship with clinical variables

Statistical learning was performed using the set of transport maps as the feature set. First, the transport maps are vectorized and concatenated as column vectors into a standard data matrix for subjects $1, ..., n$. The standard principal components analysis (PCA) technique is used to remove the data dimensions containing little or no data before computing the most correlated direction with reaction time. Then, the data matrix is projected onto the d topmost eigenvectors associated with 90% of the variance in the dataset to create the reduced-dimension data matrix $X \in \mathbb{R}^{d \times n}$. The direction in transport domain that best describes the relationship between an independent variable $v \in \mathbb{R}^{n \times 1}$ and the T1 spatial variation is computed according to Equation (5.1) [12].

This direction w_{corr} is the best in the sense of highest linear correlation coefficient between independent variable v and T1 spatial distribution as characterized in transport domain.

Assessing statistical relationship with age

The statistical relationship between GM and WM morphology and age was assessed by computing the direction in transport space most correlated with age. Let $y \in \mathbb{R}^{n \times 1}$ be the column vector representing the ages for the subjects in the study. The most correlated direction in transport space is computed by setting $v = y$ in Equation (5.1).

Assessing statistical relationship with aerobic fitness

We also assess the relationship between aerobic fitness and GM and WM morphology using Equation (5.1). In this case, y is set as the VO_2 mL/kg for the subjects, and we correct for the influences of confounding variables before performing the regression analysis using the formula in Equation (5.2), where $Z \in \mathbb{R}^{n \times c}$ is the matrix containing c covariates. In this study, the covariates were age, gender, and level of education. The variable v represents the component of y that is decorrelated and orthogonal to the confounding variables in Z.

The analyses were performed separately for GM and WM channels. Statistical significance of the computed directions w_{corr} was assessed by permutation testing with $T = 1000$ tests.

Assessing interaction between age and aerobic fitness

The interaction effect between age and fitness was assessed by performing regression on brain morphology in the transport space and setting $v = (age - \overline{age})(vO_2$ mL/kg - $\overline{vO_2}$ mL/kg), where the mean-subtracted age and vO_2 mL/kg were multiplied to test whether aerobic fitness significantly modifies the effect of age on brain tissue distribution. Covariates of gender and years of education were included in the model.

To visualize the associations and interaction terms, subjects were categorized into high fitness (fitness above the median level) and low fitness (fitness below the median level) to perform separate regressions with age, correcting for gender and years of education.

5.3.3 Results

Variation captured by transport-based morphometry

For both white matter and gray matter images, fewer principal components are needed to capture the same amount of variance in the transport domain compared to the image domain (Figure 5.7). Therefore, an image representation based on the spatial distribution of T1 (transport domain) better captures the natural structure in the dataset than examining a combination of the individual pixels

Table 5.3: Subject demographics

Age (years)	$58 - 81(66.5 \pm 5.7)$
Gender (M:F)	$60 : 112$
Education level	$2 - 8(5.7 \pm 1.3)^{\dagger}$
VO_2 **mL/kg**	$12.9 - 34.7(21.1 \pm 4.7)$

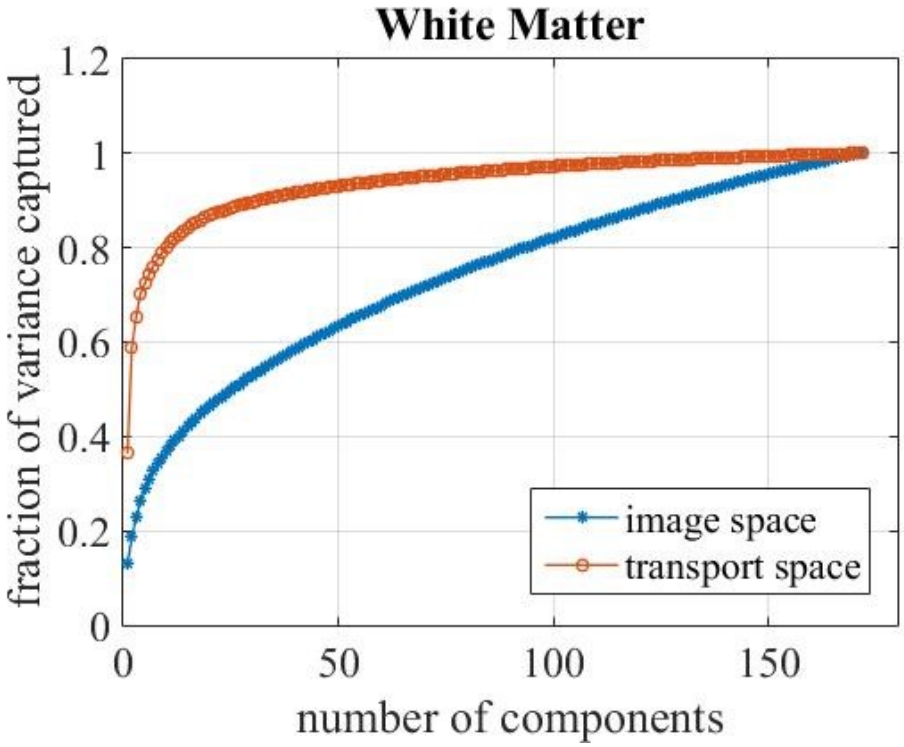
Demographic characteristics of subject cohort.† :1 = less than 7th grade; 2 =9th grade (junior high); 3 = partial high school; 4 = high school graduate; 5 =1-3 years of college or 2 year college ; 6 = college/university degree; 7 = master's degree; 8 = PhD or equivalent

(image domain representation). Thus, representation in the transport space can lead to a more natural representation of the data requiring fewer principal components than in the image domain.

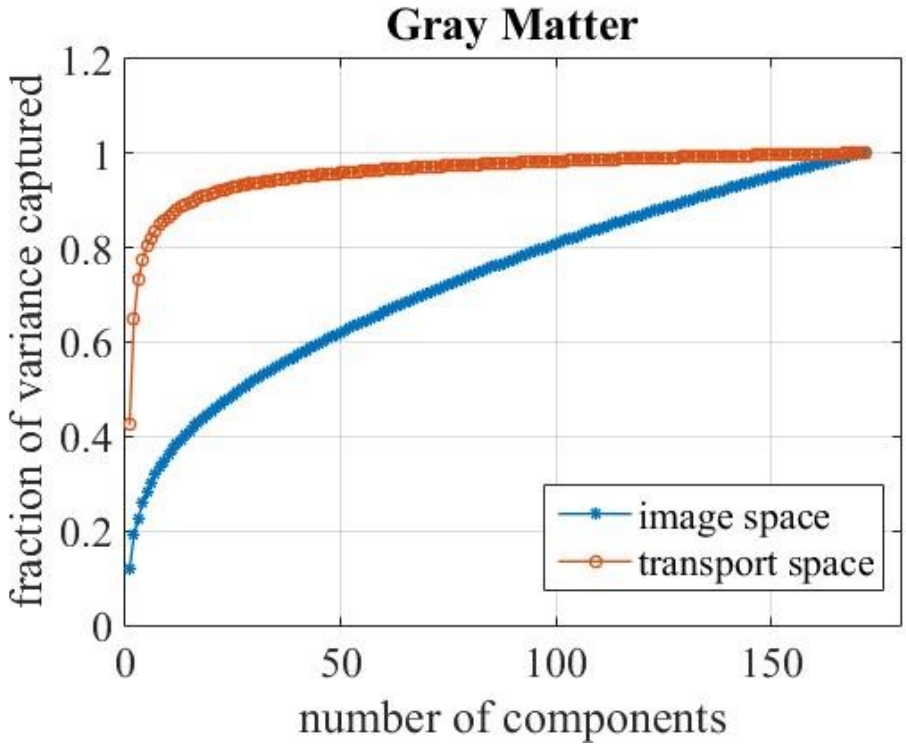
Age-related brain tissue distribution

There was a significant positive correlation between gray matter and white matter tissue distribution assessed in the transport space with age. Pearson's r = 0.36 and p = 0.00013 for white matter images and Pearson's r = 0.36 and $p < 0.00001$ for gray matter images. Figure 5.8a and 5.9a show the scatter plots corresponding to subject projection scores when the data is projected onto the direction computed in transport space. Each data point represents a subject's brain image. The mean image is mapped to projection score zero in the transport space. A line of best fit is shown that best captures the linear relationship between brain tissue morphology and age in the dataset.

The nature of the relationship with age can be interpreted through visualization of the brain morphology differences through inverse TBM transformation. Figures 5.8b and 5.9b show selected axial slices of brain images generated by sampling along the line of best fit. The axial slices that best summarize the morphology differences are shown, with the mean image represented by projection score zero. The changes correspond to an average 0.66 mm shift in GM distribution and 0.31 mm shift in WM distribution across the brain with every 10 year increase in age.



(a)



(b)

Figure 5.7: Scree plots showing number of principal components needed to capture the variation in the data for (a) white matter channels and (b) gray matter channels. Far fewer components are needed to capture the variance in the transport space compared to the image space, suggesting that the transport domain features capture structure in the data better.

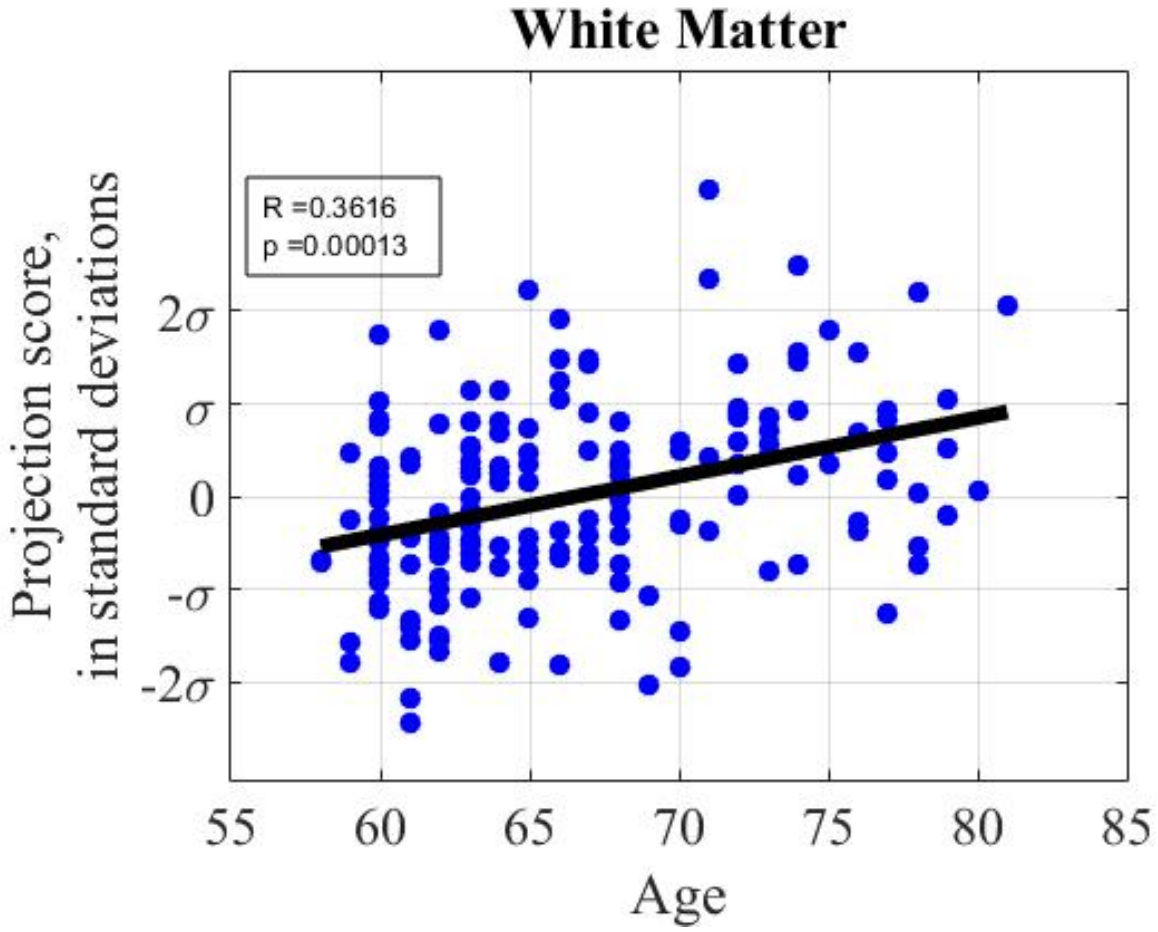


Figure 5.8: (a) Scatter plot showing relationship between subject age and white matter morphology.

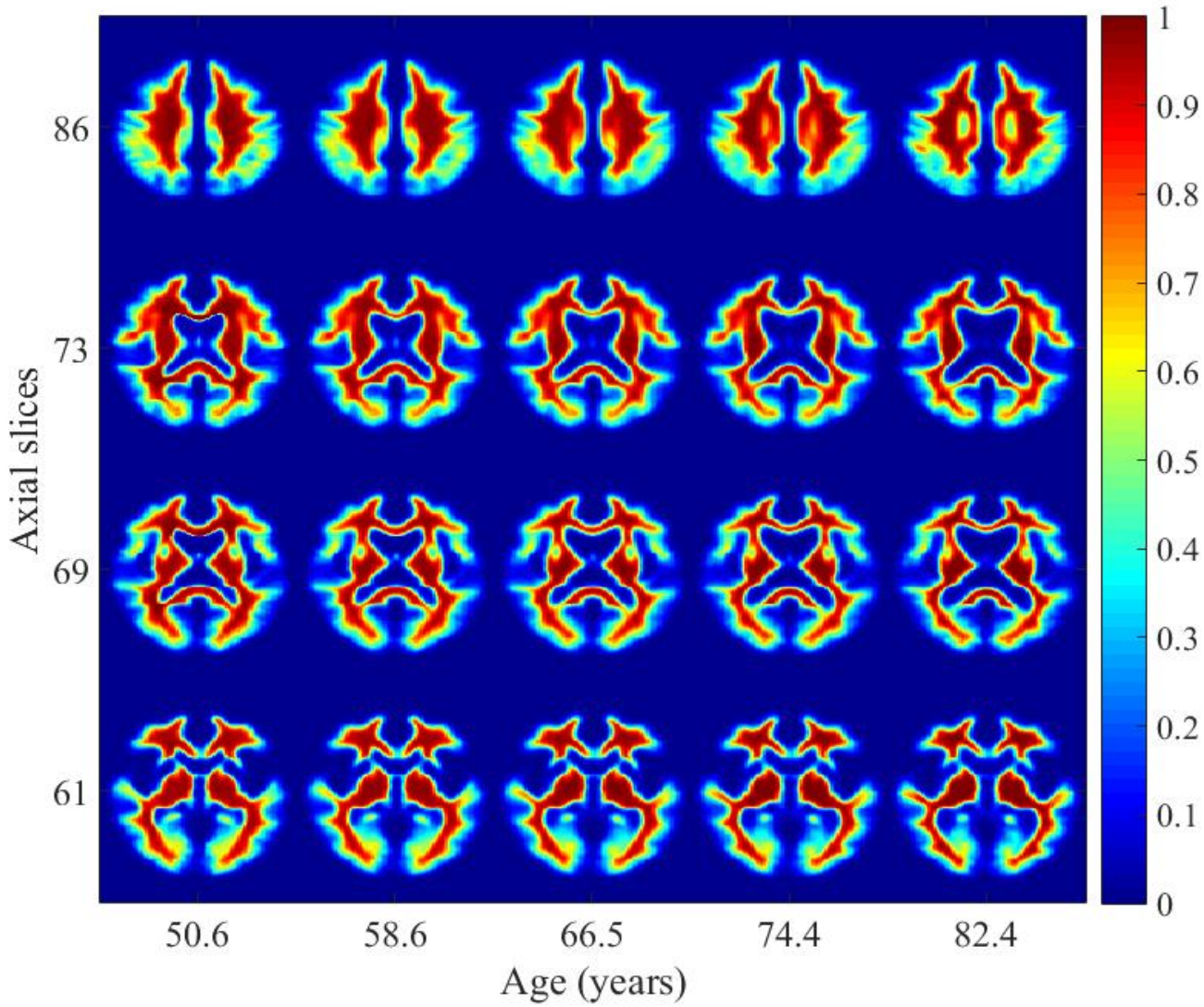


Figure 5.8: (b) images generated by Transport-Based Morphometry modeling white matter morphology changes. Select images in the z-plane are shown that best summarize the morphology differences.

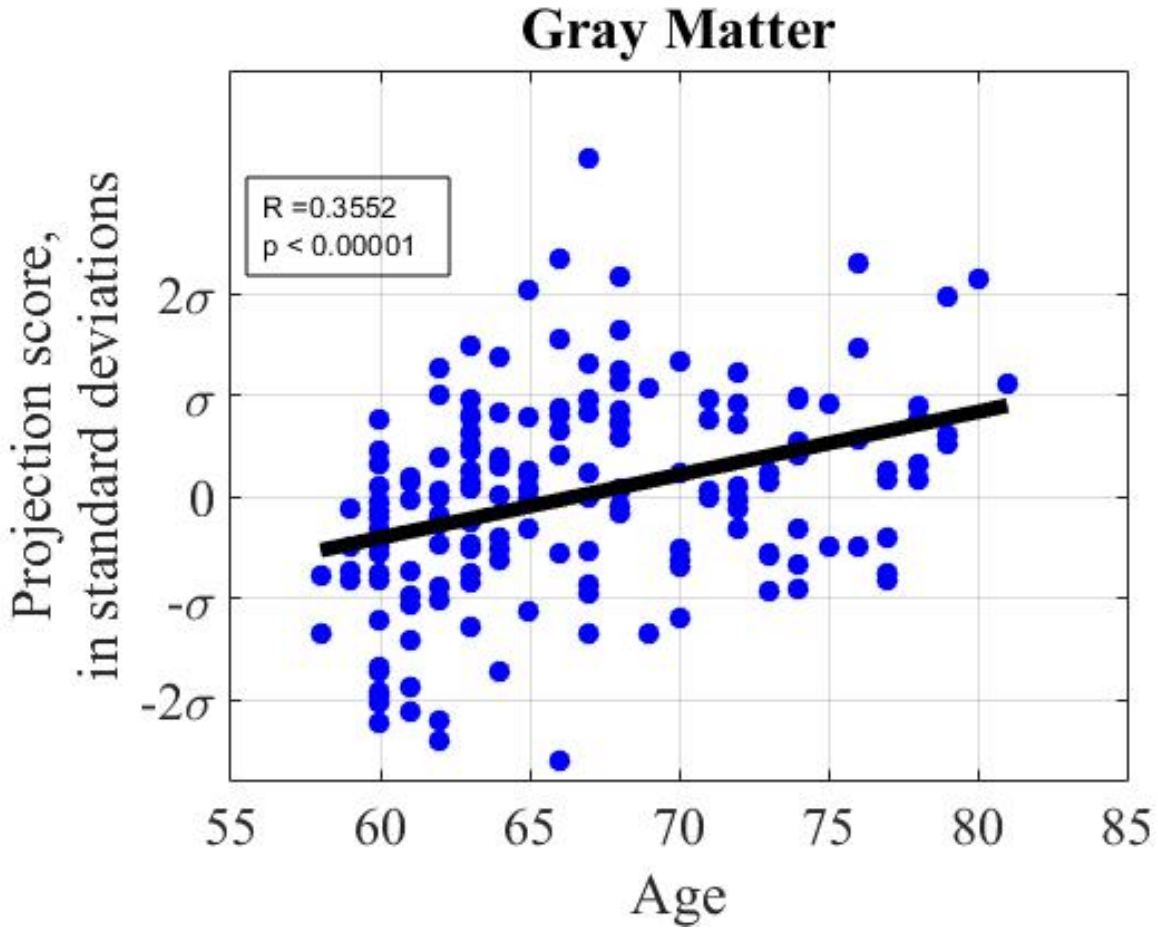


Figure 5.9: (a) Scatter plot showing relationship between subject age and gray matter morphology.

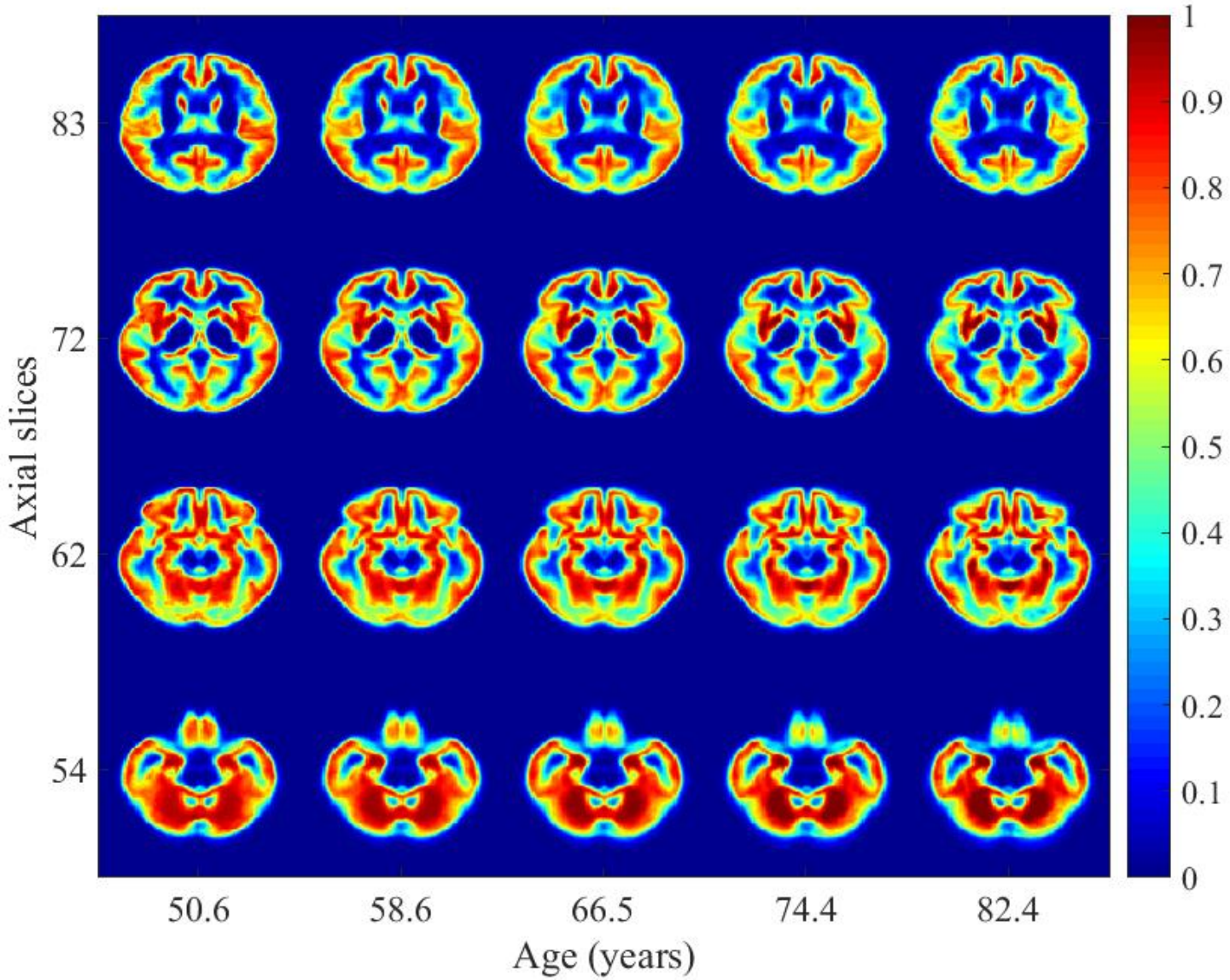


Figure 5.9: (b) images generated by Transport-Based Morphometry modeling gray matter morphology changes. Select images in the z-plane are shown that best summarize the morphology differences.

Aerobic fitness mediated effects on brain tissue distribution

The relationship between brain tissue distribution and aerobic fitness was significant even after correcting for covariates of age, gender, and years of education. The relationship for white matter (Pearson's r = 0.30, p = 0.004) as well as gray matter (Pearson's r = 0.40, $p < 0.00001$) with VO_2 mL/kg were significant.

Figures 5.10a and 5.11a show the data when it is projected onto the most correlated direction in the transport domain. The line of best fit represents the relationship between WM and GM morphology with VO_2 mL/kg in the transport domain. Interpretable images corresponding to the linear relationship in transport domain computed are shown in Figures 5.10b and 5.11b. The differences correspond to an average 0.31 mm shift in GM tissue distribution and 0.21 mm shift in WM tissue distribution for every 10 VO_2 mL/kg difference in aerobic fitness.

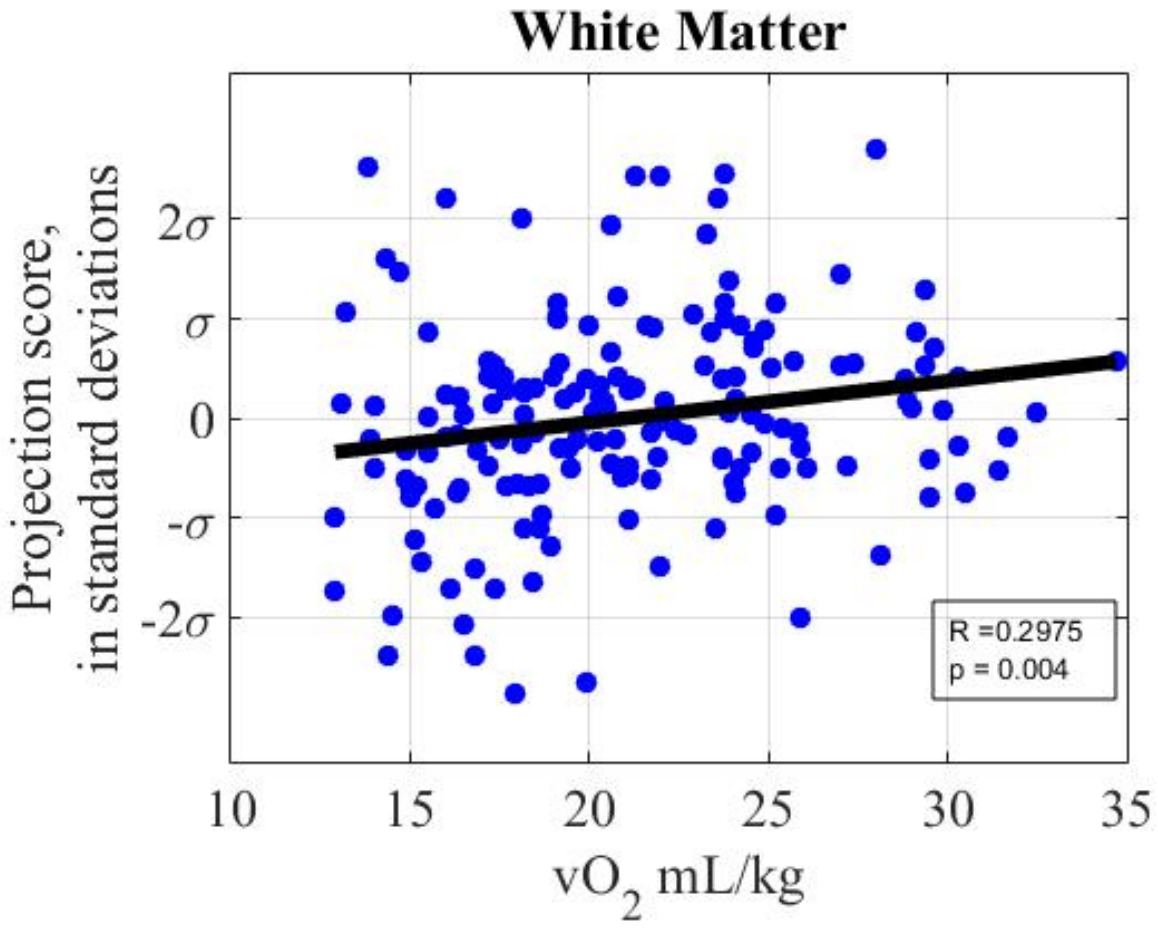


Figure 5.10: (a) Scatter plot showing relationship between subject VO_2 mL/kg and white matter morphology.

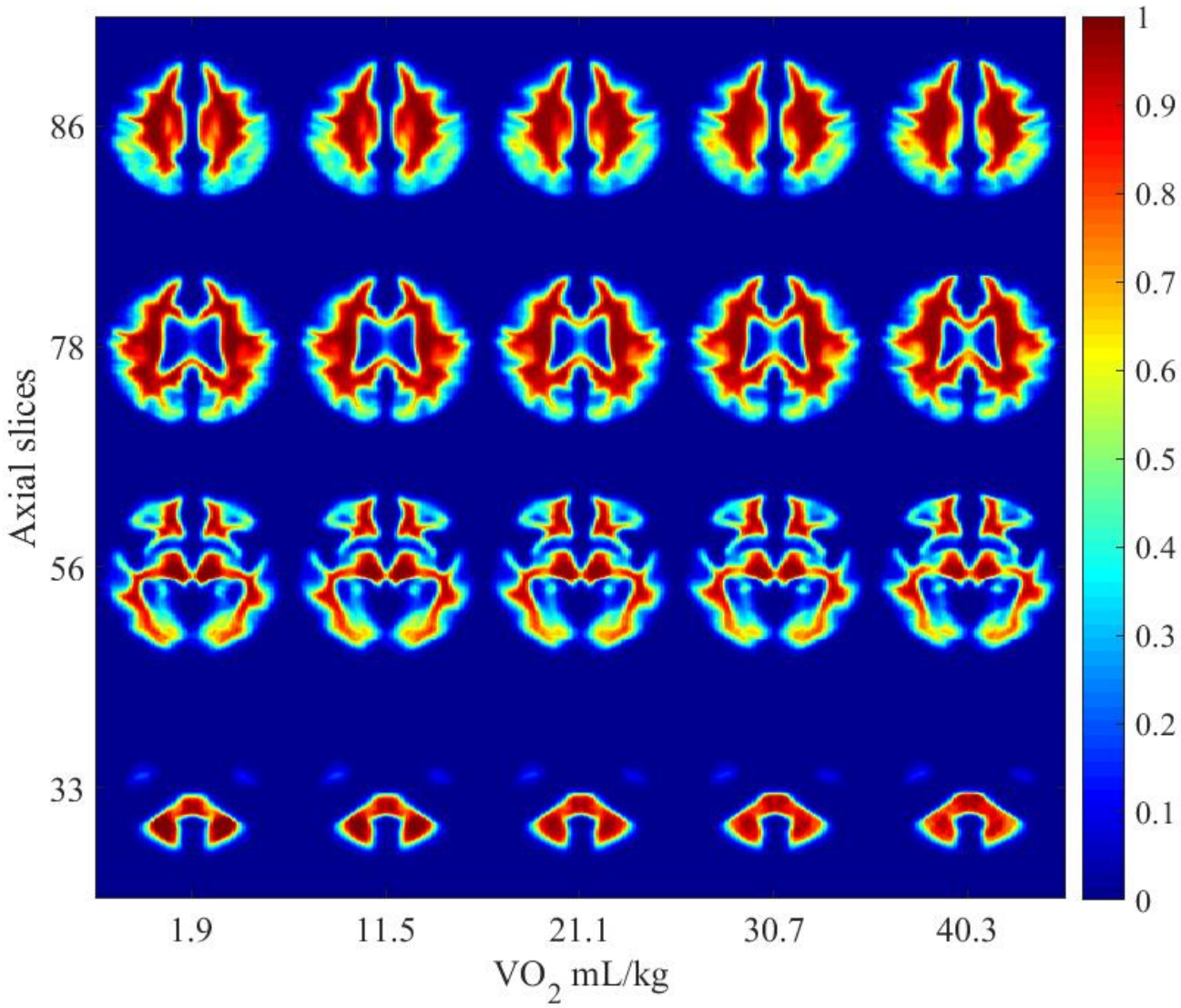


Figure 5.10: (b) images generated by Transport-Based Morphometry modeling white matter morphology changes. Select images in the z-plane are shown that best summarize the morphology differences.

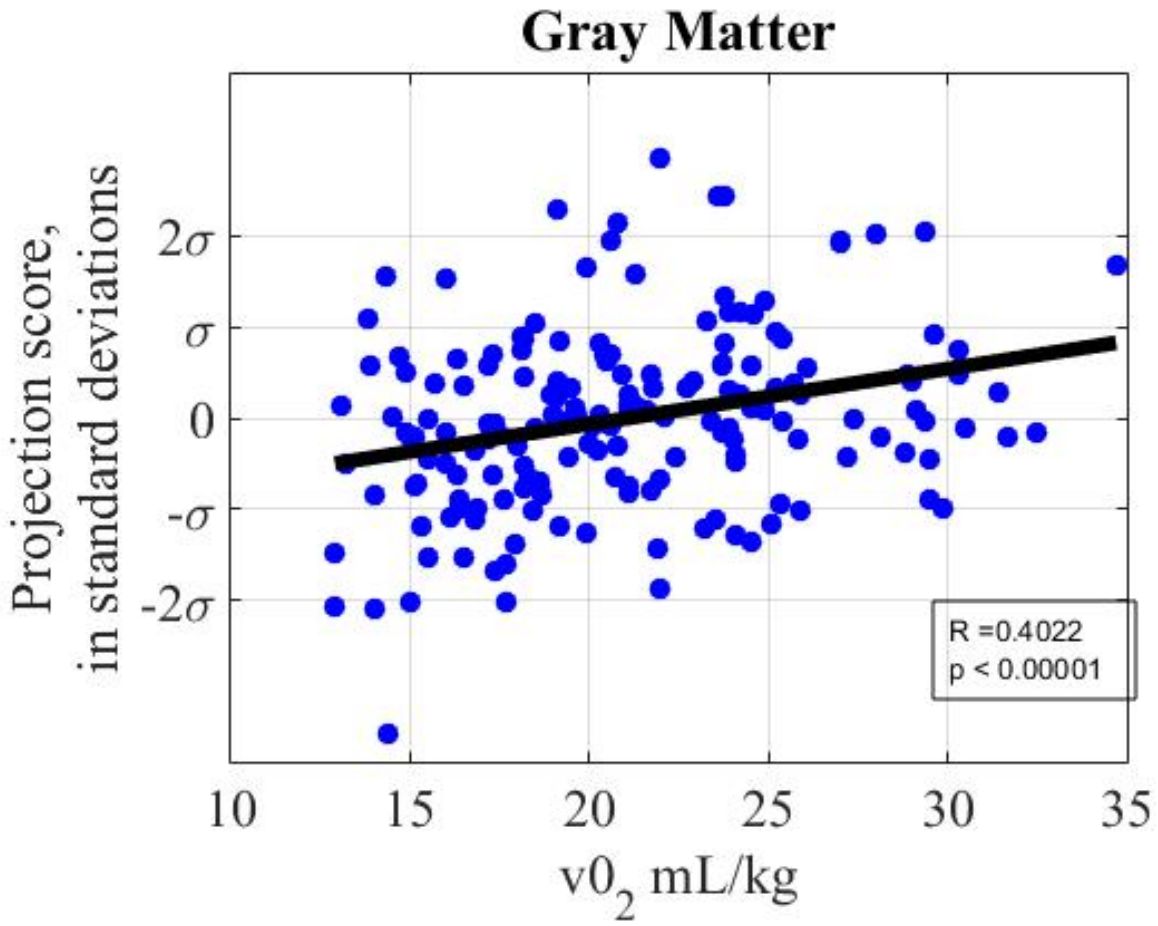


Figure 5.11: (a) Scatter plot showing relationship between subject VO_2 mL/kg and gray matter morphology.

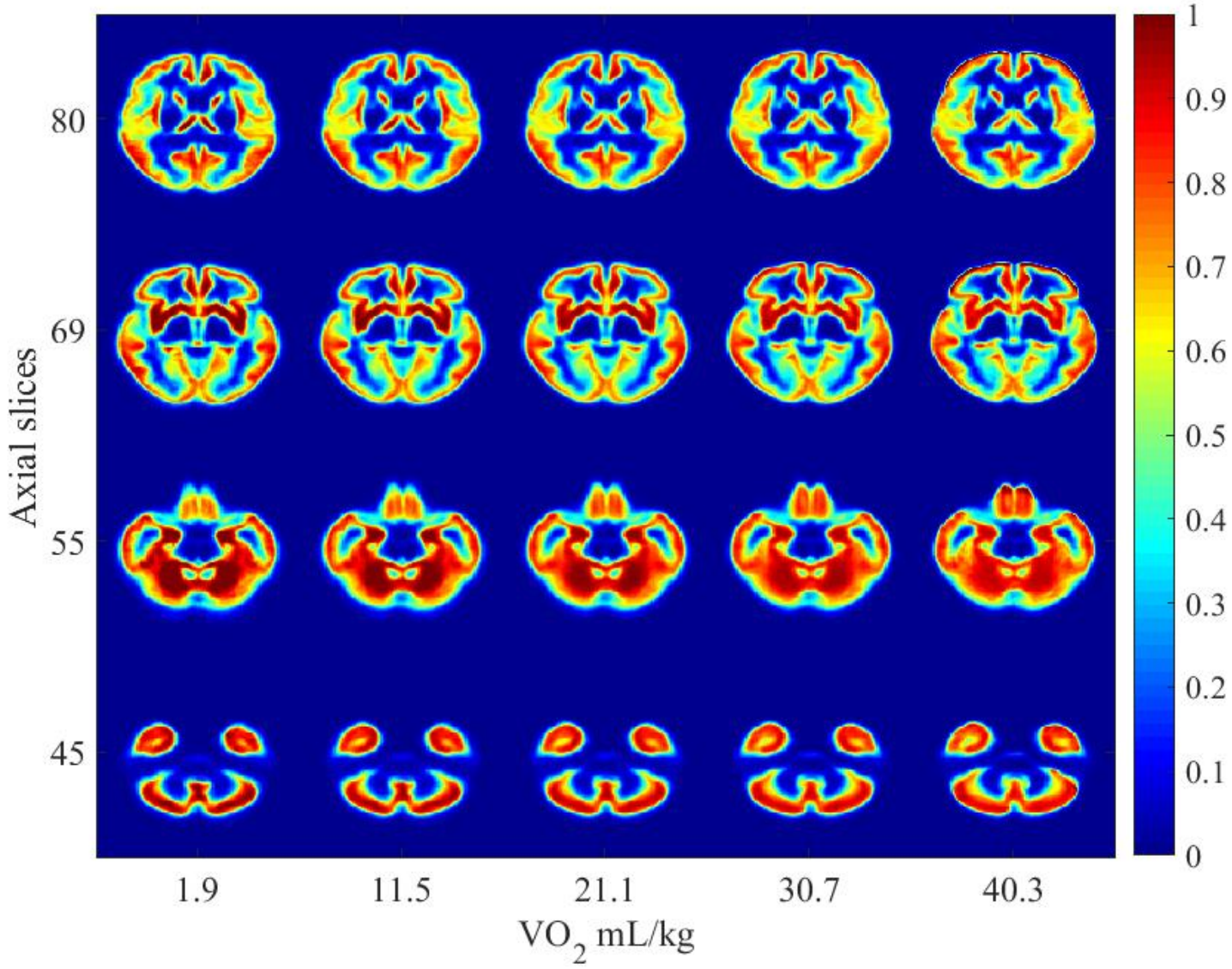


Figure 5.11: (b) images generated by Transport-Based Morphometry modeling gray matter morphology changes. Select images in the z-plane are shown that best summarize the morphology differences.

Aging by aerobic fitness effect on brain tissue distribution

The interaction effect between aerobic fitness and age was not significant for either gray matter or white matter when using VO_2 mL/kg as a measure of fitness. The GM interaction had Pearson's r = 0.1750 and p = 0.2180 when corrected for gender and education and the WM interaction had Pearson's r = 0.1607 and p = 0.5640 when corrected for gender and education.

5.3.4 Discussion and Summary

In this paper, we investigated the associations between brain morphology and aerobic fitness in older adults using the Transport-Based Morphometry (TBM) technique. While higher fitness levels have been linked with greater cortical and subcortical volumes, limitations of current computer-aided approaches have prevented analysis of tissue characteristics underlying the volume changes. Consistent with our predictions, we found that both gray matter and white matter tissue distribution are significantly associated with aerobic fitness levels in older adults using TBM, yielding both a statistical model and images illustrating the effects.

In addition to visualization, analyzing tissue spatial distribution using TBM has other advantages over traditional techniques. TBM enables a fully automated analysis of structural substrates without human biases. Approaches that rely on human input, such as manually traced ROI approaches, etc. do not enable a whole-brain analysis and may miss associations with other brain areas. Additionally, these methods do not offer insight into the morphology underlying the volume differences. Compared to deformation-based techniques, such as VBM, DBM, etc., the Transport-Based Morphometry technique can capture variations in both brain shape and tissue topology. VBM and DBM can only capture variations in brain shape and in fact, deformation fields yielded by these techniques cannot match two images with zero error, unlike TBM. Furthermore, methods that rely on non-rigid registration such as VBM and DBM are reliable only if there is accurate structural and functional alignment [24] a requirement that is difficult to achieve in practice. Voxelwise

comparison can result in heat maps identifying which voxels are found statistically significant. As we demonstrate, a nonlinear model based on T1 spatial distribution (transport domain representation) better captures the natural structure in the data than a model based on pixel-wise comparison (image domain representation). Furthermore, analyzing tissue spatial distribution has the advantage of assessing how different regions are jointly associated with aerobic fitness. At best, ROI and deformation-based methods can localize statistical associations to certain regions, but cannot visualize the nature of the changes underlying these statistical associations. Furthermore, Neither ROI nor deformation-based methods allow insight into the dynamic relationship among the regions that are jointly involved. Lastly, although these methods may enable regions to be identified in the form of heat maps, information related to the morphology of tissue is not recoverable. TBM is advantageous as an analysis tool because it does not lose any image information in the statistical analysis.

Our results add to the growing literature on the promising associations between fitness, exercise, and brain health. Previous studies in this literature have focused on gray matter volume [34, 35, 47, 48, 137], white matter microstructure [101], and cortical thickness. However, as described in the previous paragraph, a full analysis of tissue characteristics has not been performed to date due to limitations of traditional techniques. Our results represent a major advance in knowledge, demonstrating that specific morphology patterns are associated with aerobic fitness in older adults, even after correcting for the effects of age, sex, and years of education. The morphology pattern visually appears to overlap with regions affected in aging, although in opposite directions. The results suggest that aerobic fitness is important for maintaining brain health across all ages.

There are several limitations to this study. First, our study was cross-sectional in nature and so we cannot determine from this study that changing fitness levels causes the changes in tissue distribution found here. Future randomized control trials are needed to assess the causal influence of exercise on brain morphology. Future studies can also examine how these associations reported here are related to cognitive, affective, or other behavioral outcomes.

In summary, we demonstrate that the Transport-Based Morphometry technique can be used to bridge the knowledge gap between brain tissue distribution and aerobic fitness in older adults. Although prior studies have determined that cortical and subcortical volumes are increased with higher levels of aerobic fitness, TBM is the first method to overcome limitations in traditional computer-aided techniques to analyze tissue characteristics underlying volume differences by studying tissue spatial distribution.

5.4 Summary

In this chapter, we studied two application areas related to TBM and regression with a continuous variable. Specifically, we study how TBM can be used to yield a set of objective findings relating cognitive deficits post-concussion to diffusion tensor imaging findings. We find that white matter tracts in the brain that are associated with efferent and afferent aspects of reaction time are dynamically correlated with post-concussive reaction time. This is the first time that a set of objective, clinically useful markers have been described based on data-driven techniques.

We studied how aerobic fitness influences tissue distribution in the brains of older adults. We find that the areas of the brain that are affected by healthy aging overlap with areas that are influenced by aerobic fitness. Ours is the first study to yield visualization of the relationship with brain tissue distribution. Future randomized control trials can be designed to understand whether aerobic fitness can modify or change the trajectory of changes due to healthy aging.

Chapter 6

Conclusion

6.1 Summary

Medical care has come a long way since the invention of magnetic resonance imaging (MRI) in the 1970s by Paul Lauterbur and Sir Peter Mansfield. Today, imaging studies are a vital part of accurate medical diagnosis and treatment for every part of the body. While advances in imaging hardware have enabled us to collect images faster, more accurately, and measure many different properties of tissue, paralleled progress in image analysis software is needed to make sense of this data and extract relevant information. Visual inspection is not sufficient to capture complex, spatially diffuse, and subtle differences, which motivates the work in this thesis. The goal is to design an automated technique to discover and visualize clinically discriminating structural markers.

In Chapter 1, we highlighted the challenges and obstacles that needed to be overcome to accomplish the outlined goal. The first challenge is what to quantify from the images. Choosing a finite set of descriptors requires *a priori* knowledge about the most important features, which we often do not have. Second, even if there were a way to ensure that the most informative quantities were measured from the image, machine learning on high-dimensional, low-sample size (HDLSS) data is encumbered by poor generalizability that comes from high variance, especially as the complexity of classification and regression models increase. Third, another challenge is that the signals

often lie in a nonlinear metric space, where simple Euclidean operations are not sufficient to model the underlying signal space. Nonlinear modeling techniques are paramount to understanding the signal structure, but how to design a nonlinear model that explains the data is a challenging task. Furthermore, how to design a nonlinear *generative* model, which enables new data instances to be generated from the model is an even harder task. We described some current methods and their shortcomings in addressing the listed challenges. Finally, we offered that a technique based on the mathematics of optimal mass transport (OT) may be have the potential to close the gap.

In Chapter 2, we delved into the mathematical foundations of OT and reviewed some of its applications in image analysis today. We described some promising prior work demonstrating that 1D and 2D signals can become more separable when they are transformed by OT, as well as how nonlinear operations can be modeled as simple Euclidean operations in the OT-defined transform space. However, we also described that there are several computational challenges in solving OT for 3D signals such as MRI. We postulated that if a feasible OT solver for 3D images were developed, its mathematics could be extended to develop a Transport-Based Morphometry (TBM) approach for MRI. We hypothesized that a new TBM framework for MRI could meet the dual goals of discovering and visualizing clinically discriminating structural markers in a fully automated manner.

Chapter 3 described the approach we propose for solving numerical OT for large 3D images and validation of the approach on MR images of older adult brains. We show that our approach is able to overcome computational difficulties that affect other numerical OT techniques. We state that a solver that is able to produce viable transport maps for 3D images, such as ours, can enable TBM to be performed for the first time on MRI. We demonstrate several examples showing how the proposed approach could enable TBM-assisted tasks of regression, discrimination, and unsupervised learning on brain MRI.

In Chapters 4 and 5 concerned application of the new TBM framework for tasks of classification (Chapter 4) and regression (Chapter 5) analysis. In Chapter 4, we see how the TBM framework

can be applied to enable early detection of osteoarthritis 3 years in advance of symptoms with 86% accuracy. Furthermore, we see how gene-brain-behavior relationships can be explored using TBM. Given that 16p11.2 copy number variants are associated with increased risk of several developmental and psychiatric disorders, we were able to elucidate the structural changes caused by duplications and deletions of the gene, compared to control subjects. We see opposing, spatially diffuse effects that enable determination of the 16p11.2 duplication carrier, control, or deletion carrier genotype with 92.8% accuracy using gray matter appearance and 96% accuracy using white matter appearance. Furthermore, discriminant directions along which the diseases under study in this chapter shifted the morphology of the normal population were readily visualized. These are both state of the art results in the respective problems.

Finally, Chapter 5 concerned regression tasks to understand statistical relationships between anatomy and continuous clinical measurements. We studied the relationship between the appearance of the brain post-concussion, as measured by DTI, and cognitive deficits. With TBM, we discovered clinical markers on imaging that are predictive of poor cognitive outcome after concussion as well as dynamic visualization of the changes spanning multiple brain regions. We also examined the effects of aerobic exercise on brain tissue distribution of older adults and found that the areas affected overlap with those that are affected in normal aging. This is the first time for both problems that visualization of the underlying structural changes has been possible.

More broadly, the results in this chapter confirm our hypothesis through experimentation that complex, nonlinear changes across the images can be adequately captured by using the Transport-Based Morphometry technique we describe. Examining images in the transport domain enhances discovery of trends as is the case for 1D and 2D signals previously described. Furthermore, a unique contribution of TBM is that it can offer the ability to visualize the shifts in morphologic profile that enable sensitive classification or regression. The TBM approach described has broad application to multiple image modalities and supervised problems to advance both scientific understanding and yield potential clinical biomarkers to aid objective diagnosis and treatment monitoring

in the future.

6.2 New Directions

This work opens several new areas for investigation, both on technical and applied aspects of TBM.

6.2.1 Theoretical Properties of TBM for 3D

We leave it as an open problem to ”convexify” the formulation of the numerical OT solver we present in this thesis. Or, we propose coupling the TBM framework with alternative numerical OT solvers if they exist for faster processing times and more accurate numerical solutions.

In the 1D and 2D cases, important theoretical results have been proven showing that under certain assumptions, signals that are not separable in the image space become separable in the transform space. Experimentally, by analyzing MRI data in this thesis, we see that this principle generalizes in the sense that many 3D image classes that are not well-discriminated in the image space become more separable in the transform space. A theoretical characterization of this experimental observation in 3D is the subject for future work.

6.2.2 Exploration of Other Modalities and Problems

We note that the TBM framework is fully general to multiple image modalities and problems. Here, we focus on MRI applications, but in the future, applying the foundations in this thesis to many other problems has the potential for significant impact in these problem areas. TBM opens the door to many more predictive clinical studies, not just those that analyze group differences. Availability of more data samples offers the opportunity for an expanded study using training and testing data to fully validate the markers found.

6.2.3 TBM for Subtype Analysis

The applications in this thesis focus on supervised learning, or, problems where the data comes with labels. In classification tasks, the data is discretely labeled and in regression tasks, it is continuously labeled. However, there are problems where the data do not come with labels. One of these problems is in subtype analysis, where a heterogeneous population must be categorized into smaller subtypes that share characteristics in etiology, responsiveness to treatment, etc. In many psychiatric diseases, common symptoms are due to many genotypes and etiologies. A future direction for this research is in exploring the applications of TBM in the unsupervised realm, to discover subtypes that can validated clinically. For example, being able to identify several subtypes of pre-osteoarthritis would aid in designing directed therapy for different subgroups of patients, aiding the movement toward personalized medicine. We briefly touched on the potential of PCA analysis to yield meaningful visualizations of the variations in the dataset in Chapter 3.

Bibliography

[1] Luigi Ambrosio. Transport equation and Cauchy problem for non-smooth vector fields. In *Calculus of variations and nonlinear partial differential equations*, pages 1–41. Springer, 2008. 1.3.4

[2] Sigurd Angenent, Steven Haker, and Allen Tannenbaum. Minimizing flows for the Monge–Kantorovich problem. *SIAM journal on mathematical analysis*, 35(1):61–97, 2003. 2.2, 2.4

[3] Kaarin J Anstey, Holly A Mack, Helen Christensen, Shu-Chen Li, Chantal Reglade-Meslin, Jerome Maller, Rajeev Kumar, Keith Dear, Simon Easteal, and Perminder Sachdev. Corpus callosum size, reaction time speed and variability in mild cognitive disorders and in a normative sample. *Neuropsychologia*, 45(8):1911–1920, 2007. 5.2.4

[4] E Arcia and CT Gualtieri. Neurobehavioural performance of adults with closed-head injury, adults with attention deficit, and controls. *Brain Injury*, 8(5):395–404, 1994. 5.2.1

[5] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007. (document), 1.7, 3.6.4

[6] John Ashburner and Karl J Friston. Voxel-based morphometry the methods. *Neuroimage*, 11(6):805–821, 2000. 1.3.2, 4.2.4

[7] John Ashburner, Catriona Good, and Karl J Friston. Tensor based morphometry. *NeuroImage*, 11(5):S465, 2000. 1.3.2, 5.2.4

[8] John Ashburner, Chloe Hutton, Richard Frackowiak, Ingrid Johnsrude, Cathy Price, Karl Friston, et al. Identifying global anatomical differences: deformation-based morphometry. *Human brain mapping*, 6(5-6):348–357, 1998. 1.3.2, 4.2.4, 5.2.4

[9] BG Ashinsky, CE Coletta, M Bouhrara, VA Lukas, JM Boyle, DA Reiter, CP Neu, IG Goldberg, and RG Spencer. Machine learning classification of oarsi-scored human articular cartilage using magnetic resonance imaging. *Osteoarthritis and Cartilage*, 23(10):1704–1712, 2015. 4.2.1, 4.2.1, 4.2.4

[10] Alexi Assmus. Early history of x rays, 1995. 1.1

[11] Dana H Ballard and Christopher M Brown. Computer vision, 1982. *Prenice-Hall, Englewood Cliffs, NJ*, 1982. 1.2.1

[12] Saurav Basu, Soheil Kolouri, and Gustavo K Rohde. Detecting and visualizing cell phenotype differences from microscopy images using transport-based morphometry. *Proceedings of the National Academy of Sciences*, 111(9):3448–3453, 2014. (document), 1.3.4, 1.4, 2.3, 2.6, 2.3.4, 3.1, 3.2, 3.5, 3.5.1, 3.5.3, 3.8, 4.2.1, 4.2.4, 5.2.1, 5.2.2, 5.2.4, 5.3.1, 5.3.2, 5.3.2

[13] Jeffrey J Bazarian, Jianhui Zhong, Brian Blyth, Tong Zhu, Voyko Kavcic, and Derick Peterson. Diffusion tensor imaging detects clinically important axonal damage after mild traumatic brain injury: a pilot study. *Journal of neurotrauma*, 24(9):1447–1459, 2007. 5.2.4

[14] TEJ Behrens, MW Woolrich, M Jenkinson, H Johansen-Berg, RG Nunes, S Clare, PM Matthews, JM Brady, and SM Smith. Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magnetic resonance in medicine*, 50(5):1077–1088, 2003. 5.2.2

[15] N Bellamy. Validation study of womac: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. *J Orthop Rheumatol*, 1:95–108, 1988. 4.2.2

[16] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. 2.2, 2.4

[17] Jean-David Benamou, Brittany D Froese, and Adam M Oberman. Numerical solution of the optimal transportation problem using the monge–ampere equation. *Journal of Computational Physics*, 260:107–126, 2014. 2.2, 2.4

[18] Ewert Bengtsson and Patrik Malm. Screening for cervical cancer using automated analysis of pap-smears. *Computational and mathematical methods in medicine*, 2014, 2014. 1.1, 1.2.1

[19] Erin D Bigler. Neuropsychological results and neuropathological findings at autopsy in a case of mild traumatic brain injury. *Journal of the International Neuropsychological Society*, 10(05):794–806, 2004. 5.2.1

[20] EK Bijlsma, ACJ Gijsbers, JHM Schuurs-Hoeijmakers, A Van Haeringen, DE Fransen Van De Putte, B-M Anderlid, J Lundin, P Lapunzina, LA Perez Jurado, BARBARA Delle Chiaie, et al. Extending the phenotype of recurrent rearrangements of 16p11. 2: deletions in mentally retarded patients without autism and in normal individuals. *European Journal of Medical Genetics*, 52(2):77–87, 2009. 4.3.1

[21] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999. 1.3.3

[22] Nicolas Bonnotte. From Knothe's rearrangement to Brenier's optimal transport map. *SIAM Journal on Mathematical Analysis*, 45(1):64–87, 2013. 2.2

[23] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deforma-

tions. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 6:567–585, 1989. 4.2.2

[24] Fred L Bookstein. "Voxel-based morphometry" should not be used with imperfectly registered images. *Neuroimage*, 14(6):1454–1462, 2001. 1.3.2, 3.8, 5.2.1, 5.2.4, 5.3.1, 5.3.4

[25] Tom Booth, Mark E Bastin, Lars Penke, Susana Muñoz Maniega, Catherine Murray, Natalie A Royle, Alan J Gow, Janie Corley, Ross D Henderson, Maria del C Valdés Hernández, et al. Brain white matter tract integrity and cognitive abilities in community-dwelling older people: the lothian birth cohort, 1936. *Neuropsychology*, 27(5):595, 2013. 5.2.4

[26] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991. 2.2, 2.2, 3.2

[27] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991. 5.3.2

[28] Tom Brosch, Roger Tam, Alzheimers Disease Neuroimaging Initiative, et al. Manifold learning of brain mris by deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 633–640. Springer, 2013. 1.3.3

[29] Jonathan H Burdette, Paul J Laurienti, Mark A Espeland, Ashley R Morgan, Qawi Telesford, Crystal D Vechlekar, Satoru Hayaska, Janine J Jennings, Jeff A Katula, Robert A Kraft, et al. Using network science to evaluate exercise-associated brain changes in older adults. *Frontiers in aging neuroscience*, 2:23, 2010. 5.3.1

[30] D Burstein and ML Gray. Is mri fulfilling its promise for molecular imaging of cartilage in arthritis? *Osteoarthritis and Cartilage*, 14(11):1087–1090, 2006. 4.2.1

[31] Nicholas Carlson. Fluorescence guided resection, 2011. `https://engineering.`

dartmouth.edu/brainidb/fluorescence_guided_resection_background.html. (document), 1.5

[32] Knee replacement explained. http://www.knee-replacement-explained.com/Knee_Cartilage.html. Accessed: 2016-06-01. (document), 4.2

[33] Rick Chartrand, K Vixie, B Wohlberg, and Erik Bollt. A gradient descent solution to the Monge-Kantorovich problem. *Applied Mathematical Sciences*, 3(22):1071–1080, 2009. (document), 2.2, 2.2, 2.4, 3.6.2, 3.6.3, 3.7, 3.4

[34] Stanley Colcombe and Arthur F Kramer. Fitness effects on the cognitive function of older adults a meta-analytic study. *Psychological science*, 14(2):125–130, 2003. 5.3.1, 5.3.4

[35] Stanley J Colcombe, Kirk I Erickson, Paige E Scalf, Jenny S Kim, Ruchika Prakash, Edward McAuley, Steriani Elavsky, David X Marquez, Liang Hu, and Arthur F Kramer. Aerobic exercise training increases brain volume in aging humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 61(11):1166–1170, 2006. 5.3.1, 5.3.4

[36] Stanley J Colcombe, Arthur F Kramer, Kirk I Erickson, Paige Scalf, Edward McAuley, Neal J Cohen, Andrew Webb, Gerry J Jerome, David X Marquez, and Steriani Elavsky. Cardiovascular fitness, cortical plasticity, and aging. *Proceedings of the National academy of Sciences of the United States of America*, 101(9):3316–3321, 2004. 5.3.1

[37] Simons VIP Consortium et al. Simons variation in individuals project (simons vip): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron*, 73(6):1063–1067, 2012. 4.3.2

[38] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001. 1.3.3

[39] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013. 2.2, 2.2

[40] Christos Davatzikos. Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *Neuroimage*, 23(1):17–20, 2004. 1.3.2

[41] MA Davis. Epidemiology of osteoarthritis. *Clinics in Geriatric Medicine*, 4(2):241–255, 1988. 4.2.1, 4.2.1

[42] Mark A Eckert. Slowing down: age-related neurobiological predictors of processing speed. *Front Neurosci*, 5(25):1–13, 2011. 5.2.4

[43] Mark A Eckert, Noam I Keren, Donna R Roberts, Vince D Calhoun, and Kelly C Harris. Age-related changes in processing speed: unique contributions of cerebellar and prefrontal cortex. *Frontiers in Human Neuroscience*, 4:10, 2010. 5.2.4

[44] Cyrus Eierud, R Cameron Craddock, Sean Fletcher, Manek Aulakh, Brooks King-Casas, Damon Kuehl, and Stephen M LaConte. Neuroimaging after mild traumatic brain injury: review and meta-analysis. *NeuroImage: Clinical*, 4:283–294, 2014. 5.2.4

[45] Luther Pfahler Eisenhart. *Riemannian geometry*. Princeton university press, 1997. 1.2.3

[46] Kirk I Erickson, Regina L Leckie, and Andrea M Weinstein. Physical activity, fitness, and gray matter volume. *Neurobiology of aging*, 35:S20–S28, 2014. 5.3.1

[47] Kirk I Erickson, Ruchika S Prakash, Michelle W Voss, Laura Chaddock, Liang Hu, Katherine S Morris, Siobhan M White, Thomas R Wójcicki, Edward McAuley, and Arthur F Kramer. Aerobic fitness is associated with hippocampal volume in elderly humans. *Hippocampus*, 19(10):1030–1039, 2009. 3.6.1, 5.3.1, 5.3.4

[48] Kirk I Erickson, Michelle W Voss, Ruchika Shaurya Prakash, Chandramallika Basak, Amanda Szabo, Laura Chaddock, Jennifer S Kim, Susie Heo, Heloisa Alves, Siobhan M

White, et al. Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences*, 108(7):3017–3022, 2011. 5.3.1, 5.3.2, 5.3.4

[49] Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014. 2.3.3

[50] Bruce Fischl and Anders M Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055, 2000. 1.3.1

[51] Centers for Disease Control and Prevention. Osteoarthritis (oa), 2015. 4.1, 4.2.4

[52] Wellcome Trust Centre for Neuroimaging. Spm12, 2016. 1.3.2, 3.6.2, 4.3.2, 5.3.2

[53] KJ Friston and J Ashburner. Generative and recognition models for neuroanatomy. *Neuroimage*, 23(1):21–24, 2004. 1.3, 1.3.2, 4.3.1

[54] Fmrib software library v5.0, Oct 2015. 5.2.2

[55] Yi Gao, Liang-Jia Zhu, Sylvain Bouix, and Allen Tannenbaum. Interpolation of longitudinal shape and image data via optimal mass transport. In *SPIE Medical Imaging*, pages 90342X–90342X. International Society for Optics and Photonics, 2014. 2.3

[56] A Ghodadra, L Alhilali, and S Fakhran. Principal component analysis of diffusion tensor images to determine white matter injury patterns underlying postconcussive headache. *American Journal of Neuroradiology*, 2015. 5.2.1

[57] Ulf Grenander and Michael I Miller. Computational anatomy: An emerging discipline. *Quarterly of applied mathematics*, 56(4):617–694, 1998. 1.3.2

[58] Elan J Grossman, Matilde Inglese, and Roland Bammer. Mild traumatic brain injury: is diffusion imaging ready for primetime in forensic medicine? *Topics in magnetic resonance imaging: TMRI*, 21(6):379, 2010. 5.2.4

[59] Eldad Haber, Kilian M Pohl, Steven Haker, Michael Halle, Florin Talos, Lawrence L Wald, Ron Kikinis, Allen R Tannenbaum, et al. Multimodal registration of white matter brain data via optimal mass transport. In *The MIDAS Journal-Computational Biomechanics for Medicine (MICCAI 2008 Workshop)*, pages 27–36, 2008. 2.3, 2.3.2

[60] Eldad Haber, Tauseef Rehman, and Allen Tannenbaum. An efficient numerical method for the solution of the L_2 optimal mass transfer problem. *SIAM Journal on Scientific Computing*, 32(1):197–211, 2010. (document), 2.2, 2.3, 2.3.2, 2.4, 2.4

[61] Steven Haker, Allen Tannenbaum, and Ron Kikinis. Mass preserving mappings and image registration. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001*, pages 120–127. Springer, 2001. 2.3.2

[62] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004. 2.2, 2.3, 2.3.2, 2.4

[63] Steven Haker, Lei Zhu, Allen Tannenbaum, and Sigurd Angenent. Optimal mass transport for registration and warping. *International Journal of Computer Vision*, 60(3):225–240, 2004. 3.6.2, 3.6.3, 3.7

[64] Ellen Hanson, Raphael Bernier, Ken Porche, Frank I Jackson, Robin P Goin-Kochel, LeeAnne Green Snyder, Anne V Snow, Arianne Stevens Wallace, Katherine L Campe, Yuan Zhang, et al. The cognitive and behavioral phenotype of the 16p11. 2 deletion in a clinically ascertained population. *Biological psychiatry*, 77(9):785–793, 2015. 4.3.1

[65] Mark M Harrison, T Derek V Cooke, Bryn S Fisher, and Malcolm P Griffin. Patterns of knee arthrosis and patellar subluxation. *Clinical orthopaedics and related research*, 309:56–63, 1994. 4.2.4

[66] Deborah J Hellawell, Robert Taylor, and Brian Pentland. Cognitive and psychosocial outcome following moderate or severe traumatic brain injury. *Brain injury*, 13(7):489–504, 1999. 5.2.1

[67] Leena Himanen, Raija Portin, Heli Isoniemi, Hans Helenius, Timo Kurki, and Olli Tenovuo. Cognitive functions in relation to mri findings 30 years after traumatic brain injury. *Brain injury*, 19(2):93–100, 2005. 5.2.1

[68] DH Hoch, AJ Grodzinsky, TJ Koob, ML Albert, and DR Eyre. Early changes in material properties of rabbit articular cartilage after meniscectomy. *Journal of Orthopaedic Research*, 1(1):4–12, 1983. 4.2.1

[69] Michal Horemuz. Morphable brain model for monitoring disease related brain changes. *Master's Thesis*, 2015. 1.3.3

[70] Hu Huang, Akif Burak Tosun, Jia Guo, Cheng Chen, Wei Wang, John A Ozolek, and Gustavo K Rohde. Cancer diagnosis by nuclear morphometry using spatial information. *Pattern recognition letters*, 42:115–121, 2014. 2.3, 3.8

[71] Thierry AGM Huisman, Lee H Schwamm, Pamela W Schaefer, Walter J Koroshetz, Neetha Shetty-Alva, Yelda Ozsunar, Ona Wu, and A Gregory Sorensen. Diffusion tensor imaging as potential biomarker of white matter injury in diffuse axonal injury. *American Journal of Neuroradiology*, 25(3):370–376, 2004. 5.2.4

[72] `http://brain-development.org/ixi-dataset/`, accessed 3-21-16. 3.6.1

[73] Sébastien Jacquemont, Alexandre Reymond, Flore Zufferey, Louise Harewood, Robin G Walters, Zoltán Kutalik, Danielle Martinet, Yiping Shen, Armand Valsesia, Noam D Beckmann, et al. Mirror extreme bmi phenotypes associated with gene dosage at the chromosome 16p11. 2 locus. *Nature*, 478(7367):97–102, 2011. 4.3.1

[74] Douglas R Jeffery. The use of vaccinations in patients with multiple sclerosis. *Infections in medicine*, 19(2):73–79, 2002. (document), 1.5

[75] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012. 5.2.2

[76] Sarang C Joshi and Michael I Miller. Landmark matching via large deformation diffeomorphisms. *Image Processing, IEEE Transactions on*, 9(8):1357–1370, 2000. 1.3.2

[77] Candia Post Kaplan and John D Corrigan. The relationship between cognition and functional independence in adults with traumatic brain injury. *Archives of physical medicine and rehabilitation*, 75(6):643–647, 1994. 5.2.1

[78] Claude Kauffmann, Pierre Gravel, Benoît Godbout, Alain Gravel, Gilles Beaudoin, Jean-Pierre Raynauld, Johanne Martel-Pelletier, Jean-Pierre Pelletier, and Jacques A De Guise. Computer-aided method for quantification of cartilage thickness and volume changes using mri: validation study using a synthetic model. *Biomedical Engineering, IEEE Transactions on*, 50(8):978–988, 2003. 4.2.1, 4.2.1

[79] JH Kellgren and JS Lawrence. Radiological assessment of osteo-arthrosis. *Annals of the rheumatic diseases*, 16(4):494, 1957. 4.2.2

[80] Soheil Kolouri, Se Rim Park, and Gustavo K Rohde. The radon cumulative distribution transform and its application to image classification. *IEEE Transactions on Image Processing*, 25(2):920–934, 2016. 2.3.4, 3.5

[81] Soheil Kolouri and Gustavo K Rohde. Quantifying and visualizing variations in sets of images using continuous linear optimal transport. In *SPIE Medical Imaging*, pages 903438–903438. International Society for Optics and Photonics, 2014. (document), 1.6, 1.3.4, 1.3.4, 1.4

[82] Soheil Kolouri and Gustavo K Rohde. Temporal information inference from static high-content fluorescent microscopy. In *Northeast Bioengineering Conference (NEBEC), 2014 40th Annual*, pages 1–2. IEEE, 2014. 2.3

[83] Soheil Kolouri and Gustavo K Rohde. Transport-based single frame super resolution of very low resolution face images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4876–4884, 2015. (document), 2.5, 2.3.3, 3.8

[84] Soheil Kolouri, Akif B Tosun, John A Ozolek, and Gustavo K Rohde. A continuous linear optimal transport approach for pattern analysis in image datasets. *Pattern Recognition*, 2015. 1.3.4, 2.3, 2.3.4, 3.2, 3.2, 3.2, 3.5, 3.8

[85] S Kundu, S Kolouri, KI Erickson, E McAuley, A Kramer, and GK Rohde. Multiscale variational optimal transport for medical image modeling and analysis. *Under revision.*, xx(xx):xx, 2016. 4.2.1, 4.2.4, 5.2.1, 5.2.2, 5.2.4, 5.3.1, 5.3.2, 5.3.2

[86] Letizia Leocani, Leonardo G Cohen, Eric M Wassermann, Katsunori Ikoma, and Mark Hallett. Human corticospinal excitability evaluated with transcranial magnetic stimulation during different reaction time paradigms. *Brain*, 123(6):1161–1173, 2000. 5.2.4

[87] Peihua Li, Qilong Wang, and Lei Zhang. A novel earth mover's distance methodology for image matching with gaussian mixture models. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1689–1696, 2013. 2.3.1

[88] Jan Maas, Martin Rumpf, Carola Schönlieb, and Stefan Simon. A generalized model for optimal transport of images including dissipation and density modulation. *arXiv preprint arXiv:1504.01988*, 2015. 2.2

[89] David JC MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003. 1.2.2

[90] Dgradient. http://www.mathworks.com/matlabcentral/fileexchange/29887-dgradient, 2016. 3.6.2

[91] AR Mayer, J Ling, MV Mannell, C Gasparovic, JP Phillips, D Doezema, R Reichard, and RA Yeo. A prospective diffusion tensor imaging study in mild traumatic brain injury. *Neurology*, 74(8):643–650, 2010. 5.2.4

[92] Lester B Mayers and Thomas S Redick. Clinical utility of impact assessment for postconcussion return-to-play counseling: psychometric issues. *Journal of clinical and experimental neuropsychology*, 34(3):235–242, 2012. 5.2.2

[93] Shane E McCarthy, Vladimir Makarov, George Kirov, Anjene M Addington, Jon McClellan, Seungtai Yoon, Diana O Perkins, Diane E Dickel, Mary Kusenda, Olga Krastoshevsky, et al. Microduplications of 16p11. 2 are associated with schizophrenia. *Nature genetics*, 41(11):1223–1227, 2009. 4.1, 4.3.1

[94] Laura Miles, Robert I Grossman, Glyn Johnson, James S Babb, Leonard Diller, and Matilde Inglese. Short-term dti predictors of cognitive dysfunction in mild traumatic brain injury. *Brain Injury*, 22(2):115–122, 2008. 5.2.4

[95] R. Miller. IBM Buying Merge Healthcare For $1B to bring Medical Image Analysis to Watson Health. http://techcrunch.com/2015/08/06/ibm-buying-merge-healthcare-for-1b-to-bring-medical-image-analysis-to-2015. [Online; accessed 2015-December-08]. 1.1

[96] Yurii Nesterov et al. Gradient methods for minimizing composite objective function. Technical report, UCL, 2007. 3.3.3

[97] M Nevitt, D Felson, and G Lester. The osteoarthritis initiative: protocol for the cohort study. *URL: http://oai. epi-ucsf. org/datarelease/docs/StudyDesignProtocol. pdf*, 2006. 4.2.2

[98] Claudia Niemann, Ben Godde, and Claudia Voelcker-Rehage. Not only cardiovascular, but also coordinative exercise increases hippocampal volume in older adults. *Frontiers in aging neuroscience*, 6, 2014. 5.3.1

[99] SN Niogi, P Mukherjee, J Ghajar, C Johnson, RA Kolster, R Sarkar, H Lee, M Meeker, RD Zimmerman, GT Manley, et al. Extent of microstructural white matter injury in post-concussive syndrome correlates with impaired cognitive reaction time: a 3t diffusion tensor imaging study of mild traumatic brain injury. *American Journal of Neuroradiology*, 29(5):967–973, 2008. 5.2.1, 5.2.1, 5.2.4

[100] Sumit N Niogi and Pratik Mukherjee. Diffusion tensor imaging of mild traumatic brain injury. *The Journal of head trauma rehabilitation*, 25(4):241–255, 2010. 5.2.4

[101] Lauren E Oberlin, Timothy D Verstynen, Agnieszka Z Burzynska, Michelle W Voss, Ruchika Shaurya Prakash, Laura Chaddock-Heyman, Chelsea Wong, Jason Fanning, Elizabeth Awick, Neha Gothe, et al. White matter microstructure mediates the relationship between cardiorespiratory fitness and spatial working memory in older adults. *NeuroImage*, 2015. 5.3.1, 5.3.4

[102] Richard C Oldfield. The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971. 5.3.2

[103] =http://www.oai.ucsf.edu/datarelease/. 4.2.2

[104] `https://www.andrew.cmu.edu/user/gustavor/software.html`, accessed 3-21-16. 3.8

[105] Julia P Owen, Yi Shin Chang, Nicholas J Pojman, Polina Bukshpun, Mari LJ Wakahiro, Elysa J Marco, Jeffrey I Berman, John E Spiro, Wendy K Chung, Randy L Buckner, et al. Aberrant white matter microstructure in children with 16p11. 2 deletions. *The Journal of Neuroscience*, 34(18):6214–6223, 2014. 4.3.1

[106] Se Rim Park, Soheil Kolouri, Shinjini Kundu, and Gustavo Rohde. The cumulative distribution transform and linear pattern classification. *arXiv preprint arXiv:1507.05936*, 2015. 1.4

[107] Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *Computer vision, 2009 IEEE 12th international conference on*, pages 460–467. IEEE, 2009. 2.3.1

[108] CG Peterfy, E Schneider, and M Nevitt. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage*, 16(12):1433–1441, 2008. 4.2.2

[109] Abid Y Qureshi, Sophia Mueller, Abraham Z Snyder, Pratik Mukherjee, Jeffrey I Berman, Timothy PL Roberts, Srikantan S Nagarajan, John E Spiro, Wendy K Chung, Elliott H Sherr, et al. Opposing brain differences in 16p11. 2 deletion and duplication carriers. *The Journal of Neuroscience*, 34(34):11199–11211, 2014. 4.3.1

[110] Ruthger Righart, Marco Duering, Mariya Gonik, Eric Jouvent, Sonia Reyes, Dominique Hervé, Hugues Chabriat, and Martin Dichgans. Impact of regional cortical and subcortical changes on processing speed in cerebral small vessel disease. *NeuroImage: Clinical*, 2:854–861, 2013. 5.2.4

[111] S. Rim Park, S. Kolouri, S. Kundu, and G. Rohde. The Cumulative Distribution Transform and Linear Pattern Classification. *ArXiv e-prints 1507.05936*, July 2015. 2.3.4, 3.5, 3.8

[112] Jill A Rosenfeld, Justine Coppinger, Bassem A Bejjani, Santhosh Girirajan, Evan E Eichler, Lisa G Shaffer, and Blake C Ballif. Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11. 2 microdeletions and microduplications. *Journal of neurodevelopmental disorders*, 2(1):26, 2009. 4.3.1

[113] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric

for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000. (document), 2.3, 2.3.1, 2.2

[114] Louis-Philippe Saumier, Martial Agueh, and Boualem Khouider. An efficient numerical algorithm for the L2 optimal transport problem with periodic densities. *IMA Journal of Applied Mathematics*, 80(1):135–157, 2015. 2.2, 2.4

[115] Lior Shamir, Nikita Orlov, D Mark Eckley, Tomasz Macura, Josiah Johnston, and Ilya G Goldberg. Wndchrm–an open source utility for biological image analysis. *Source code for biology and medicine*, 3(1):1, 2008. 1.3.1

[116] JA Sheikh and JA Yesavage. Geriatric depression scale (gds): Recent findings and development of a shorter version in: Brink tl, ed. clinical gerontology: a guide to assessment and intervention, 1986. 5.3.2

[117] Marwan Shinawi, Pengfei Liu, Sung-Hae L Kang, Joseph Shen, John W Belmont, Daryl A Scott, Frank J Probst, William J Craigen, Brett H Graham, Amber Pursley, et al. Recurrent reciprocal 16p11. 2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *Journal of medical genetics*, 47(5):332–341, 2010. 4.3.1

[118] Marion Smits, Gavin C Houston, Diederik WJ Dippel, Piotr A Wielopolski, Meike W Vernooij, Peter J Koudstaal, MG Myriam Hunink, and Aad van der Lugt. Microstructural brain injury in post-concussion syndrome after minor head injury. *Neuroradiology*, 53(8):553–563, 2011. 5.2.4

[119] Justin Solomon, Fernando de Goes, Pixar Animation Studios, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (Proc. SIGGRAPH 2015), to appear*, 2015. 2.2

[120] Y Stern, M Sano, J Paulson, and R Mayeux. Modified mini-mental state examination: validity and reliability. *Neurology*, 37(suppl 1):179, 1987. 5.3.2

[121] Zhengyu Su, Yalin Wang, Rui Shi, Wei Zeng, Jian Sun, Feng Luo, and Xianfeng Gu. Optimal mass transport for shape matching and comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(11):2246–2259, 2015. 2.2, 2.4

[122] Jimeng Sun and Chandan K Reddy. Siam, 2013. (document), 1.2

[123] Jay A Swedberg and Jeffrey R Steinbauer. Osteoarthritis-epidemiology, pathophysiology, diagnosis, and treatment. *Am Fam Physician*, 45:557–568, 1992. 4.2.1, 4.2.1

[124] Lisanne F Ten Brinke, Niousha Bolandzadeh, Lindsay S Nagamatsu, Chun Liang Hsu, Jennifer C Davis, Karim Miran-Khan, and Teresa Liu-Ambrose. Aerobic exercise increases hippocampal volume in older women with probable mild cognitive impairment: a 6-month randomised controlled trial. *British journal of sports medicine*, 49(4):248–254, 2015. 5.3.1

[125] FMRIB's Diffusion Toolbox. Fdt. `http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FDT`, mar 2016. 1.3.2, 5.2.2

[126] Akif Burak Tosun, Oleksandr Yergiyev, Soheil Kolouri, Jan F Silverman, and Gustavo K Rohde. Novel computer-aided diagnosis of mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens. In *SPIE Medical Imaging*, pages 90410Z–90410Z. International Society for Optics and Photonics, 2014. 2.3, 3.8

[127] David S Tuch, David H Salat, Jonathan J Wisco, Alexandra K Zaleta, Nathanael D Hevelone, and H Diana Rosas. Choice reaction time performance correlates with diffusion anisotropy in white matter pathways supporting visuospatial attention. *Proceedings of the National Academy of Sciences of the United States of America*, 102(34):12212–12217, 2005. 5.2.4

[128] Tauseef ur Rehman, Eldad Haber, Gallagher Pryor, John Melonakos, and Allen Tannenbaum. 3D nonrigid registration via optimal mass transport on the GPU. *Medical image analysis*, 13(6):931–940, 2009. (document), 2.2, 2.3, 2.3.2, 2.3, 2.4

[129] Kenneth L Urish, Matthew G Keffalas, John R Durkin, David J Miller, Constance R Chu, and Timothy J Mosher. T2 texture index of cartilage can predict early symptomatic oa progression: data from the osteoarthritis initiative. *Osteoarthritis and Cartilage*, 21(10):1550–1557, 2013. 4.2.1, 4.2.1, 4.2.4

[130] Timothy D Verstynen, Andrea M Weinstein, Walter W Schneider, John M Jakicic, Dana L Rofey, and Kirk I Erickson. Increased body mass index is associated with a global and distributed decrease in white matter microstructural integrity. *Psychosomatic medicine*, 74(7):682, 2012. 5.3.1

[131] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 2.1, 2.2, 5.2.1, 5.2.2

[132] Michelle W Voss, Kirk I Erickson, Ruchika Shaurya Prakash, Laura Chaddock, Jennifer S Kim, Heloisa Alves, Amanda Szabo, Siobhan M Phillips, Thomas R Wójcicki, Emily L Mailey, et al. Neurobiological markers of exercise-related brain plasticity in older adults. *Brain, behavior, and immunity*, 28:90–99, 2013. 5.3.2

[133] Michelle W Voss, Ruchika S Prakash, Kirk I Erickson, Chandramallika Basak, Laura Chaddock, Jennifer S Kim, Heloisa Alves, Susie Heo, Amanda Szabo, Siobhan M White, et al. Plasticity of brain networks in a randomized intervention trial of exercise training in older adults. *Frontiers in aging neuroscience*, 2:32, 2010. 5.3.1

[134] Wei Wang, Yilin Mo, John A Ozolek, and Gustavo K Rohde. Penalized fisher discriminant analysis and its application to image-based morphometry. *Pattern recognition letters*, 32(15):2128–2135, 2011. 1.4, 3.5.2

[135] Wei Wang, John A Ozolek, Dejan Slepcev, Ann B Lee, Cheng Chen, and Gustavo K Rohde. An optimal transportation approach for nuclear structure-based pathology. *Medical Imaging, IEEE Transactions on*, 30(3):621–631, 2011. 1.3.4, 1.4, 2.3.4, 3.5, 3.8

[136] Wei Wang, Dejan Slepcev, Saurav Basu, John A Ozolek, and Gustavo K Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *International journal of computer vision*, 101(2):254–269, 2013. 1.3.4, 1.4, 2.3, 2.3.4, 3.1, 3.2, 3.2, 3.5, 3.8, 4.2.1, 4.2.4, 5.2.1, 5.2.2, 5.2.4, 5.3.1, 5.3.2

[137] Andrea M Weinstein, Michelle W Voss, Ruchika Shaurya Prakash, Laura Chaddock, Amanda Szabo, Siobhan M White, Thomas R Wojcicki, Emily Mailey, Edward McAuley, Arthur F Kramer, et al. The association between aerobic fitness and executive function is mediated by prefrontal cortex volume. *Brain, behavior, and immunity*, 26(5):811–819, 2012. 5.3.1, 5.3.4

[138] Lauren A Weiss, Yiping Shen, Joshua M Korn, Dan E Arking, David T Miller, Ragnheidur Fossdal, Evald Saemundsen, Hreinn Stefansson, Manuel AR Ferreira, Todd Green, et al. Association between microdeletion and microduplication at 16p11. 2 and autism. *New England Journal of Medicine*, 358(7):667–675, 2008. 4.1

[139] Chelsea N Wong, Laura Chaddock-Heyman, Michelle W Voss, Agnieszka Z Burzynska, Chandramallika Basak, Kirk I Erickson, Ruchika S Prakash, Amanda N Szabo-Reed, Siobhan M Phillips, Thomas Wojcicki, et al. Brain activation during dual-task processing is associated with cardiorespiratory fitness and performance in older adults. *Frontiers in aging neuroscience*, 7, 2015. 5.3.1

[140] Jeffrey R Wozniak, Linda Krach, Erin Ward, Bryon A Mueller, Ryan Muetzel, Sarah Schnoebelen, Andrew Kiragu, and Kelvin O Lim. Neurocognitive and neuroimaging correlates

of pediatric traumatic brain injury: a diffusion tensor imaging (dti) study. *Archives of Clinical Neuropsychology*, 22(5):555–568, 2007. 5.2.1, 5.2.1, 5.2.4

[141] Yang Xia, Jonathan B Moody, and Hisham Alhadlaq. Orientational dependence of t2 relaxation in articular cartilage: A microscopic mri (μmri) study. *Magnetic resonance in medicine*, 48(3):460–469, 2002. 4.2.4

[142] BS Yulish, JM Lieberman, AJ Newman, PJ Bryan, GP Mulopulos, and MT Modic. Juvenile rheumatoid arthritis: assessment with mr imaging. *Radiology*, 165(1):149–152, 1987. 4.2.1, 4.2.4