



BAYESIAN ESTIMATION AND INFERENCE IN COMPUTATIONAL ANATOMY AND NEUROIMAGING: METHODS & APPLICATIONS

EDITED BY: Xiaoying Tang, Thomas Fletcher and Michael I. Miller
PUBLISHED IN: Frontiers in Neuroscience



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88945-984-1
DOI 10.3389/978-2-88945-984-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

BAYESIAN ESTIMATION AND INFERENCE IN COMPUTATIONAL ANATOMY AND NEUROIMAGING: METHODS & APPLICATIONS

Topic Editors:

Xiaoying Tang, Southern University of Science and Technology, China

Thomas Fletcher, The University of Utah, United States

Michael I. Miller, Johns Hopkins University, United States

Computational Anatomy (CA) is an emerging discipline aiming to understand anatomy by utilizing a comprehensive set of mathematical tools. CA focuses on providing precise statistical encodings of anatomy with direct application to a broad range of biological and medical settings.

During the past two decades, there has been an ever-increasing pace in the development of neuroimaging techniques, delivering in vivo information on the anatomy and physiological signals of different human organs through a variety of imaging modalities such as MRI, x-ray, CT, and PET. These multi-modality medical images provide valuable data for accurate interpretation and estimation of various biological parameters such as anatomical labels, disease types, cognitive states, functional connectivity between distinct anatomical regions, as well as activation responses to specific stimuli.

In the era of big neuroimaging data, Bayes' theorem provides a powerful tool to deliver statistical conclusions by combining the current information and prior experience. When sufficiently good data is available, Bayes' theorem can utilize it fully and provide statistical inferences/estimations with the least error rate. Bayes' theorem arose roughly three hundred years ago and has seen extensive application in many fields of science and technology, including recent neuroimaging, ever since. The last fifteen years have seen a great deal of success in the application of Bayes' theorem to the field of CA and neuroimaging. That said, given that the power and success of Bayes' rule largely depends on the validity of its probabilistic inputs, it is still a challenge to perform Bayesian estimation and inference on the typically noisy neuroimaging data of the real world.

We assembled contributions focusing on recent developments in CA and neuroimaging through Bayesian estimation and inference, in terms of both methodologies and applications. It is anticipated that the articles in this Research Topic will provide a greater insight into the field of Bayesian imaging analysis.

Citation: Tang, X., Fletcher, T., Miller, M. I., eds. (2019). Bayesian Estimation and Inference in Computational Anatomy and Neuroimaging: Methods & Applications. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-984-1

Table of Contents

- 04 Editorial: Bayesian Estimation and Inference in Computational Anatomy and Neuroimaging: Methods and Applications**
Xiaoying Tang and Michael I. Miller
- 06 Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework**
Saurabh Jain, Annemie Ribbens, Diana M. Sima, Melissa Cambron, Jacques De Keyser, Chenyu Wang, Michael H. Barnett, Sabine Van Huffel, Frederik Maes and Dirk Smeets
- 17 Age and Glaucoma-Related Characteristics in Retinal Nerve Fiber Layer and Choroid: Localized Morphometrics and Visualization Using Functional Shapes Registration**
Sieun Lee, Morgan L. Heisler, Karteek Popuri, Nicolas Charon, Benjamin Charlier, Alain Trouvé, Paul J. Mackenzie, Marinko V. Sarunic and Mirza Faisal Beg
- 29 Bayesian Tractography Using Geometric Shape Priors**
Xiaoming Dong, Zhengwu Zhang and Anuj Srivastava
- 44 On the Complexity of Human Neuroanatomy at the Millimeter Morphome Scale: Developing Codes and Characterizing Entropy Indexed to Spatial Scale**
Daniel J. Tward and Michael I. Miller for the Alzheimer's Disease Neuroimaging Initiative
- 54 Brain MRI Pattern Recognition Translated to Clinical Scenarios**
Andreia V. Faria, Zifei Liang, Michael I. Miller and Susumu Mori
- 68 A Hierarchical Bayesian Model for the Identification of PET Markers Associated to the Prediction of Surgical Outcome After Anterior Temporal Lobe Resection**
Sharon Chiang, Michele Guindani, Hsiang J. Yeh, Sandra Dewar, Zulfi Haneef, John M. Stern and Marina Vannucci
- 84 Multivariate Heteroscedasticity Models for Functional Brain Connectivity**
Christof Seiler and Susan Holmes
- 95 A Bayesian Spatial Model to Predict Disease Status Using Imaging Data From Various Modalities**
Wenqiong Xue, F. DuBois Bowman and Jian Kang
- 104 A Fully-Automated Subcortical and Ventricular Shape Generation Pipeline Preserving Smoothness and Anatomical Topology**
Xiaoying Tang, Yuan Luo, Zhibin Chen, Nianwei Huang, Hans J. Johnson, Jane S. Paulsen and Michael I. Miller



Editorial: Bayesian Estimation and Inference in Computational Anatomy and Neuroimaging: Methods and Applications

Xiaoying Tang^{1*} and Michael I. Miller^{2,3,4}

¹ Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China,

² Center for Imaging Science, Johns Hopkins University, Baltimore, MD, United States, ³ Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States, ⁴ Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, United States

Keywords: Bayesian estimation, computational anatomy, medical imaging, shape, prediction

Editorial on the Research Topic

Bayesian Estimation and Inference in Computational Anatomy and Neuroimaging: Methods and Applications

This e-book brings together a total of nine studies focusing on imaging-based Bayesian estimation and computation. Computational tools were developed for various clinical purposes, including white matter (WM) lesion segmentation, statistical shape analysis, fiber tracking, anatomy coding, disease status and pathology detection and prediction, as well as functional connectivity analysis. Most studies focused on MRI whereas two analyzed respectively, OCT and PET. The investigations included a variety of populations, including healthy normal and patients with Multiple Sclerosis (MS), glaucoma, Alzheimer's disease (AD), Ataxia, primary progressive aphasia (PPA), Huntington's disease (HD), temporal lobe epilepsy (TLE), and Parkinson's disease (PD).

Jain et al. proposed a pipeline for segmenting two time point WM lesions in a joint expectation-maximization (EM) framework. The pipeline utilized two-modality MR images (a 3D T1-weighted image and a 3D FLAIR image). It modeled the lesion evolution between the two time points using a Gaussian mixture model and conducted simultaneous tissue and lesion segmentation in images from both time points. The model was optimized using a joint EM algorithm. The proposed pipeline was validated on two datasets, respectively involving 12 and 10 patients with MS.

Lee et al. conducted statistical shape analysis of the retinal nerve fiber layer (RNFL) and choroid in the framework of computational anatomy (CA), with OCT being used. A novel registration technique, namely functional shapes (fshape), was employed to match two retinas and to generate the mean of multiple retinas. In fshape, a diffeomorphism was obtained by a joint optimization of the surface geometry (the retinal surface) and functional signals mapped onto the surface (the retinal layer thickness). Point-wise analyses and visualizations were conducted using the fshape-derived diffeomorphisms. Using this technique, the authors successfully examined age-related and glaucoma-related spatial RNFL thickness patterns in 38 participants.

Dong et al. presented a method for fiber tracking in the Bayesian setting with geometric shape priors. The fiber tracts between regions of interest (ROIs) were initialized as Euclidean curves and then iteratively updated via deformations using gradients of a posterior energy. Estimations were performed using an energy function involving three components: the likelihood, the prior knowledge on the geometric shapes of fibers, and a roughness penalty term. The prior on the geometric shapes relied on atlas-based statistical shape models of

OPEN ACCESS

Edited and reviewed by:

Vince D. Calhoun,
University of New Mexico,
United States

*Correspondence:

Xiaoying Tang
tangxy@sustech.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 30 November 2018

Accepted: 15 May 2019

Published: 29 May 2019

Citation:

Tang X and Miller MI (2019) Editorial:
Bayesian Estimation and Inference in
Computational Anatomy and
Neuroimaging: Methods and
Applications. *Front. Neurosci.* 13:562.
doi: 10.3389/fnins.2019.00562

fiber curves between ROIs. The proposed tractography methodology was evaluated on both simulated 2D data and 30 real 3D data from the Human Connectome Project (HCP).

Tward and Miller invented a strategy for anatomy coding using a Bayesian prior model. The entropy of an anatomy of interest was quantified as a function of code rate (number of bits). In this setting, the authors studied the shape of 12 subcortical structures of the human brain through diffeomorphic transformations relating each of them to a population-averaging and structure-specific template. A multivariate Gaussian prior model was trained using 650 MRI data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The authors found that at 1 mm all subcortical structures can be described with <35 bits, and at 1.5 mm all structures can be described with <12 bits.

Faria et al. explored the effectiveness of using MRI-based whole brain segmentations to extract key anatomical phenotypes for characterizing four neurodegenerative diseases (Ataxia, $n = 16$; HD, $n = 52$; AD, $n = 66$; and PPA, $n = 50$), all inducing brain atrophy. Homogeneous clinically-relevant phenotypes were successfully clustered. Using the structural quantification and simple linear classifiers, the authors were able to detect the four diseases with satisfactory accuracies. Moreover, the anatomical features automatically delivered by the classifiers agreed with the patterns of the disease pathologies.

Chiang et al. developed an integrative Bayesian prediction model to identify a brain's pathological status through a selection of fluoro-deoxyglucose PET imaging biomarkers. The proposed model was tested on 19 patients with TLE who subsequently underwent anterior temporal lobe resection. The proposed model successfully identified patient subgroups characterized by latent pathologies that associate differentially to clinical outcomes. It also yielded imaging biomarkers that describe the pathological states of the subjects. The proposed method was shown to achieve good accuracy in predicting post-surgical seizure recurrence.

Seiler and Holmes analyzed functional connectivity using two novel heteroscedasticity covariance models. The first model was low-dimensional, scaling linearly in the total number of brain parcellations. And the second model scaled quadratically. Both models were applied to the functional-resting fMRI data of 820 subjects from HCP, comparing connectivity between short and conventional sleepers. Stronger functional connectivity in short than conventional sleepers were found in brain regions that are consistent with previous findings.

Xue et al. proposed a Bayesian hierarchical model to predict disease status by incorporating information from both functional and structural brain imaging scans. Posterior probabilities were used to perform prediction, with the parameter estimations conducted on samples drawn from the joint posterior distribution using Markov Chain Monte Carlo methods.

Predictions were conducted at both whole-brain and voxel levels, with the disease-related brain regions identified from the voxel-level prediction results. The proposed model was applied to a PD study, with key regions contributing to accurate prediction having been identified.

Tang et al. presented a fully-automated pipeline for generating subcortical and ventricular shapes from brain MR images. The proposed pipeline consisted of three key components: (1) automated structure segmentation; (2) study-specific shape template creation; (3) deformation-based shape filtering. The proposed pipeline was validated on two HD datasets, respectively involving 16 and 1,445 MRI scans. Another independent dataset, consisting of 15 atlas images and 20 testing images, was also used to quantitatively evaluate the proposed pipeline. High accuracy has been observed.

Together, these studies provide evidence for the power of Bayesian estimation theorem in imaging analysis. This e-book contains both methodology developments and scientific applications, with several imaging techniques having been involved (MRI, PET, and OCT). Three papers specifically lie in the CA framework, focusing on statistical shape analysis (Lee et al., Tward and Miller, and Tang et al.). These nine studies provide tools and examples for imaging-based computational analyses addressing various clinical questions and advances future research in this field.

AUTHOR CONTRIBUTIONS

XT wrote the paper and MM revised the manuscript critically for important intellectual content.

FUNDING

XT is supported by the National Natural Science Foundation of China (NSFC 81501546) and the National Key R&D Program of China (2017YFC0112404). MM is supported by NIH P41 EB015909, NIH R01 EB000975, and NIH R01 EB008171.

Conflict of Interest Statement: MM owns an equal share in Anatomyworks LLC. The terms of this arrangement have been reviewed and approved by the Johns Hopkins University, in accordance with its conflict of interest policy.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Tang and Miller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Two Time Point MS Lesion Segmentation in Brain MRI: An Expectation-Maximization Framework

Saurabh Jain^{1*}, Annemie Ribbens¹, Diana M. Sima^{1,2}, Melissa Cambron³, Jacques De Keyser^{3,4}, Chenyu Wang⁵, Michael H. Barnett⁵, Sabine Van Huffel^{2,7}, Frederik Maes⁶ and Dirk Smeets^{1,8}

¹ icometrix, Leuven, Belgium, ² STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium, ³ Department of Neurology, Universitair Ziekenhuis Brussel, Vrije Universiteit Brussel (VUB), Brussel, Belgium, ⁴ Department of Neurology, University Medical Center Groningen (UMCG), Groningen, Netherlands, ⁵ Sydney Neuroimaging Analysis Centre, Brain and Mind Centre, University of Sydney, Sydney, NSW, Australia, ⁶ Medical Image Computing, Processing Speech and Images (PSI), Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium, ⁷ Imec, Leuven, Belgium, ⁸ Biolumaging Lab, Universiteit Antwerpen, Antwerp, Belgium

OPEN ACCESS

Edited by:

Xiaoying Tang,
SYSU-CMU Joint Institute of
Engineering, China

Reviewed by:

Ashish Raj,
Weill Cornell Medical College, USA
Miaomiao Zhang,
Massachusetts Institute of
Technology, USA

*Correspondence:

Saurabh Jain
saurabh.jain@icometrix.com

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 17 August 2016

Accepted: 01 December 2016

Published: 19 December 2016

Citation:

Jain S, Ribbens A, Sima DM,
Cambron M, De Keyser J, Wang C,
Barnett MH, Van Huffel S, Maes F and
Smeets D (2016) Two Time Point MS
Lesion Segmentation in Brain MRI: An
Expectation-Maximization Framework.
Front. Neurosci. 10:576.
doi: 10.3389/fnins.2016.00576

Purpose: Lesion volume is a meaningful measure in multiple sclerosis (MS) prognosis. Manual lesion segmentation for computing volume in a single or multiple time points is time consuming and suffers from intra and inter-observer variability.

Methods: In this paper, we present MSmetrix-long: a joint expectation-maximization (EM) framework for two time point white matter (WM) lesion segmentation. MSmetrix-long takes as input a 3D T1-weighted and a 3D FLAIR MR image and segments lesions in three steps: (1) cross-sectional lesion segmentation of the two time points; (2) creation of difference image, which is used to model the lesion evolution; (3) a joint EM lesion segmentation framework that uses output of step (1) and step (2) to provide the final lesion segmentation. The accuracy (Dice score) and reproducibility (absolute lesion volume difference) of MSmetrix-long is evaluated using two datasets.

Results: On the first dataset, the median Dice score between MSmetrix-long and expert lesion segmentation was 0.63 and the Pearson correlation coefficient (PCC) was equal to 0.96. On the second dataset, the median absolute volume difference was 0.11 ml.

Conclusions: MSmetrix-long is accurate and consistent in segmenting MS lesions. Also, MSmetrix-long compares favorably with the publicly available longitudinal MS lesion segmentation algorithm of Lesion Segmentation Toolbox.

Keywords: MSmetrix, multiple sclerosis, longitudinal lesion segmentation, expectation-maximization, MRI

1. INTRODUCTION

Accurate and reliable lesion segmentation based on brain MRI scans is valuable for the diagnosis and monitoring of disease activity in patients with Multiple Sclerosis (MS) (Blystad et al., 2016; Deeks, 2016). The availability of longitudinal MRI data permits an analysis of lesion evolution over time, a potential biomarker of disease progression and treatment efficacy. **Figure 1** shows bias corrected FLAIR images of a MS subject scanned twice with an interval of approximately 1 year, along with the expert lesion segmentation followed by the lesion evolution, i.e., the new,

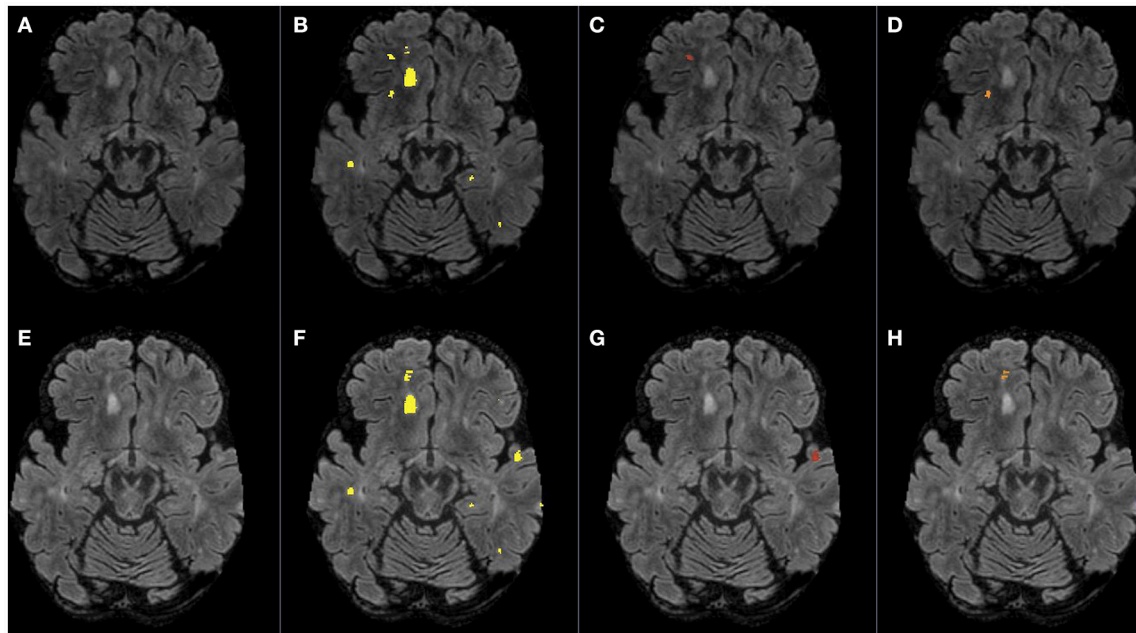


FIGURE 1 | Bias corrected FLAIR images (A,E) followed by super-imposed lesion segmentations from: (B,F) the expert, (C) disappearing lesion, (D) shrinking lesion, (G) new lesion, and (H) enlarging lesion. The first row corresponds to time point 1 and the second row corresponds to time point 2.

disappearing, enlarging, and shrinking lesions. Although expert manual delineation of lesions is considered as the gold standard, it is time consuming and often suffers from intra and inter observer variability (Erbayat Altay et al., 2013). To alleviate this problem, several automatic methods have been proposed in the literature to segment MS lesions. Interestingly, the vast majority of automatic methods are based on a single time point (cross-sectional) and relatively few methods take into account multiple time points (longitudinal) (Llado et al., 2012; Garcia-Lorenzo et al., 2013). Executing a cross-sectional method for each time point would indeed produce the longitudinal measures of interest, but such measures are less reliable as each time point is processed independently. Longitudinal methods incorporate both spatial and temporal information and are expected to be more reliable. Based on the underlying approach, longitudinal methods could be categorized in three different groups: change detection (Gerig et al., 2000; Welte et al., 2001; Prima et al., 2002; Rey et al., 2002; Bosc et al., 2003; Elliott et al., 2013), 4D connectivity (Metcalf et al., 1992; Bernardis et al., 2013) and outlier detection (Solomon and Sood, 2004; Ait-Ali et al., 2005) in multiple time points. Pre-processing of input MR images in these three groups is generally performed and consists of registration to a reference image or a common space, skull stripping, bias field correction and intensity normalization.

Change detection methods primarily aim to detect MS activity by statistical analysis of image features or by measuring local volume variation. Statistical analysis can be performed in an unsupervised or supervised manner. Unsupervised approaches detect significant changes in the intensities between consecutive scans by either analysing the corresponding patches of two time points (Bosc et al., 2003), or performing clustering on

the extracted spatial and temporal features from longitudinal images (Gerig et al., 2000; Welte et al., 2001; Prima et al., 2002). The main drawback with unsupervised approaches is that they assume perfect registration and intensity normalization. Supervised approaches learn the desired change from a training dataset; for instance, in Elliott et al. (2013), a random forest discriminative classifier was trained to learn relevant features (intensity, size, and contextual information) related to new lesions and then use these features to segment them. The main drawback with this approach is that it often requires that the training dataset is large enough in order to capture all the distinctive features of the lesions to be segmented. To avoid the need for extracting image features, changes between consecutive images could be directly detected by measuring local volume variations. To this end, a Jacobian operator could be applied to the local deformation field obtained after non-rigid registration between the two time points. Although this approach has proven to be invariant to registration errors, it has given poor results for lesion segmentation (Rey et al., 2002).

Four-dimensional connectivity methods use voxel association in space and time to simultaneously segment and track lesion evolution. For example, Metcalf et al. (1992) segments the lesions in two time points by clustering voxels that are both spatially and temporally adjacent to each other. The main disadvantage of this approach is that it often results in substantial false lesion segmentation. A more advanced method from the same family is based on spectral graph partitioning Bernardis et al. (2013). It constructs a 3D graph in which spatial pairwise affinities characterize lesions and background, and temporal affinities between adjacent time points represent lesion evolution direction. This graph is segmented into lesions and non-lesions

via spectral clustering by maximizing the force within-group attraction and between-group repulsion. The drawback of this approach is that it cannot discriminate between consistent artifacts and lesions.

Outlier detection methods are based on the fact that MS lesions are hyper-intense on T2-weighted and fluid-attenuated inversion recovery (FLAIR) brain MRI scans and thus could be detected as an outlier to normal tissue class intensities distribution. For example, a joint expectation-maximization (EM) based approach such as in Ait-Ali et al. (2005) models the healthy brain tissue classes across the time points as a Gaussian mixture model (GMM) using a 4D (3D + time) intensity histogram. The parameters of the model are optimized via a modified version of the EM algorithm referred to as STREAM. After convergence, the lesions are extracted as outliers to healthy tissue classes using Mahalanobis distance and some prior information. In this approach the lesion segmentation is largely dependent on the choice of the Mahalanobis distance parameter and does not target lesion evolution, which is clinically relevant (Ait-Ali et al., 2005). Another approach using outlier detection is based on the hidden Markov model (HMM) technique as in Solomon and Sood (2004). Initially, EM segments the first time point into different tissue classes including lesions, which are then manually corrected. Subsequently, using a lesion growth transition model and outlier detection sensor model, lesions are segmented in the following time points. The transition model enforces consistent lesion segmentation; however, it was validated only on simulations with exponential lesion growth.

In this paper, we present MSmetrix-long: an iterative white matter (WM) lesion segmentation method based on a joint EM framework that takes as input clinically acquired 3D T1-weighted and 3D FLAIR images of two time points. The proposed framework is fully automated, unsupervised and models the lesion evolution as GMM between two time points, thereby simultaneously segmenting new, enlarging, disappearing, shrinking and static lesions. The method is validated for accuracy and reproducibility on two different datasets that are representative for clinically feasible acquisition protocols.

2. METHODS

The MSmetrix-long pipeline analyses the MS lesions evolution between two time points based on 3D T1-weighted and 3D FLAIR image acquired at each time point. The pipeline has four steps: (1) Cross-sectional analysis, that segments the individual time points into gray matter (GM), WM, cerebro-spinal fluid (CSF), and lesions, (2) FLAIR based difference image, which is created by subtracting the FLAIR images of both time points after bias correction, co-registration and intensity normalization, (3) Joint lesion segmentation, that aims to improve the individual time point lesion segmentation using the other time point information on tissue and lesion segmentation (initialized using step-1 results) and difference image obtained from step-2, (4) a pruning step, that refines the lesion segmentation obtained in the step-3 to eliminate non-lesions candidates. **Figure 2** presents an illustrative explanation of these steps. Steps (3) and (4) are performed sequentially in both directions, by using one time

point as reference and then the other. These steps are also iterated, by changing the input lesion segmentation used as prior. Only for the first iteration, the lesion segmentations priors come from the cross-sectional pipeline in step-3, while from the second iteration onwards lesion segmentations from previous iteration are used to initialize the lesion priors for the current iteration. The convergence of our method is decided when the relative lesion segmentation difference between the current and previous iteration is negligible. It takes generally three iterations for the algorithm to converge. The following sections explain the different steps in more detail.

2.1. Cross-Sectional Analysis

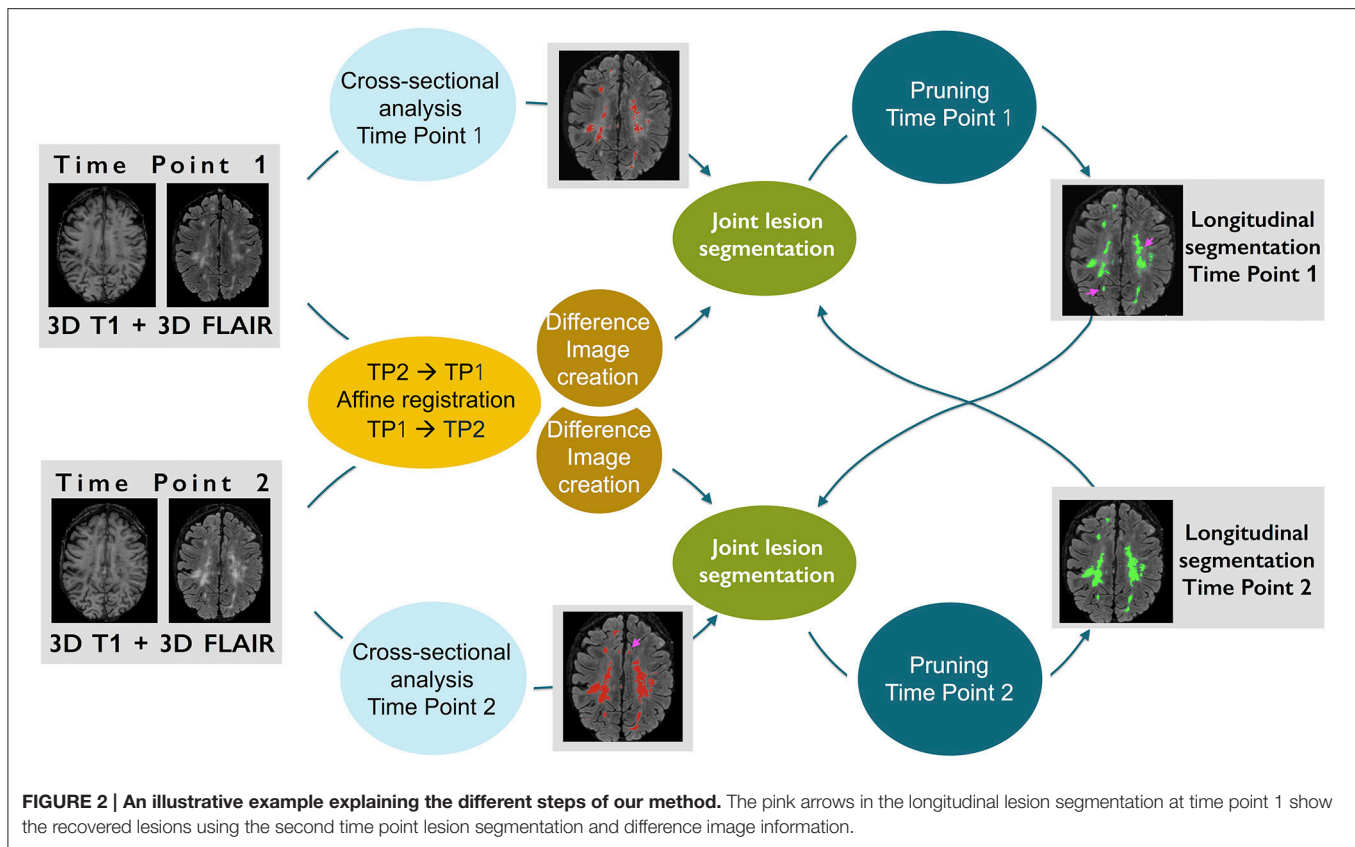
Image segmentation is performed independently for each time point using the cross-sectional pipeline referred to as MSmetrix-cross (Jain et al., 2015). The cross-sectional method iteratively segments the T1-weighted image into GM, WM, and CSF, segments the WM lesions on the FLAIR image as an outlier to normal brain using Mahalanobis distance, and performs lesion filling in the T1-weighted image to improve tissue segmentation at next iteration. After convergence, segmentations of WM, GM, CSF and lesions are created. In addition, bias corrected T1-weighted and FLAIR images are also produced. The segmentation tasks of the MSmetrix-cross are optimized using an EM algorithm (Van Leemput et al., 1999) as implemented in NiftySeg (Cardoso, 2012).

2.2. FLAIR Based Difference Image

A FLAIR based difference image is created by image co-registration and intensity normalization. Image co-registration is performed using affine registration, which comprises a rigid registration based on the whole T1-weighted image, followed by a skull based affine registration to avoid small scaling differences, and a final whole brain rigid registration (Smeets et al., 2016). The rigid registration and skull based affine registration use an inverse consistent registration algorithm (Modat et al., 2010). Subsequently, the GM, WM, CSF, lesion segmentation and the bias corrected FLAIR images obtained from the cross-sectional analysis are propagated using the final affine transformation. The matched bias corrected FLAIR images are then corrected for differential bias field as described in Lewis and Fox (2004). Subsequently, the differential bias field corrected images are intensity normalized using a cumulative histogram matching technique Castleman (1995) with the image of time point 1 as reference. A FLAIR based difference image is now created in time point 1 space. To avoid bias toward a specific time point, a second difference image is created, using time point 2 space as reference.

2.3. Joint Lesion Segmentation

The joint lesion segmentation model aims at simultaneous tissue class label segmentation of the images from both time points (see the blocks denoted by “Joint lesion segmentation” in **Figure 2**). The model is optimized using a joint EM algorithm. In this section we present the model formulation, for more details please see Supplementary Material. We now describe the notations, variables and assumptions used, followed by the model definition and its optimization using joint EM.



2.3.1. Notations, Variables, and Model Assumptions

We assume that image 1, image 2 and difference image are co-registered and have the same voxel size. Additionally, image 1 and in image 2 have identical tissue classes. We denote the set of image intensities for image 1 as I_1 and similarly for image 2 as I_2 and for the directional difference image as D . $k^{(1)}$ and $k^{(2)}$ denote tissue class indices for image 1 and image 2, respectively. The tissue class labels in image 1 and in image 2 are denoted by L_1 and L_2 respectively.

We now specify our model assumptions. A Gaussian mixture model is used on the image intensities of each time point where a Gaussian model is used for each tissue class. Let θ_1 denote the Gaussian mixture model parameters for the intensities of image 1 and $P(I_1|L_1, \theta_1)$ denotes the probabilistic model for image 1. Analogously, the probabilistic model for image 2 is denoted by $P(I_2|L_2, \theta_2)$.

We make the underlying assumption that the “difference image” might be independently generated as an image that captures anatomical changes including new lesions or atrophy. The image created by subtracting image 1 from image 2 or vice-versa (after intensity normalization) is one such instance of the difference image. The intensity model of image 1 and image 2 can therefore be reinforced by including a tissue transition model defined on the difference image. As our method focuses on two time point WM lesion segmentation, we only model the transformations between WM and lesions. We assume that the difference image has three different transformations: “static,”

“growth,” and “shrinkage.” The static transformation class is defined as a set of voxels in the difference image that are either labeled as WM in both images or lesions. The growth transformation class (describing the new and enlarging lesions) is defined as a set of voxels in the difference image that are labeled as WM in image 1 and lesion in image 2. The shrinkage transformation class (describing the disappearing and shrinking lesions) is defined as a set of voxels in the difference image that are labeled as lesion in image 1 and WM in image 2. For all other possible tissue transformations from image 1 and image 2 a uniform distribution is assumed. **Figure 3** shows an illustrative example of the difference image and the histograms of its classes with corresponding Gaussian fitting. Under these assumptions, a Gaussian mixture model for the difference image intensities is used where each transformation class (static, growth, shrinkage) is modeled as Gaussian. The probabilistic model for the difference image is denoted by $P(D|L_1, L_2, \zeta)$, where ζ stands for the Gaussian mixture model parameters for the difference image intensities.

Finally, we assume that we have no prior knowledge on the relationship of the tissue class labels between both images. Therefore, we define the prior probabilities independently for each image. Often these prior probabilities are given by a probabilistic atlas. However, our cross-sectional model provided us with more specific knowledge and hence, we use the probabilistic cross-sectional tissue class segmentations. The prior probabilities on tissue class labels

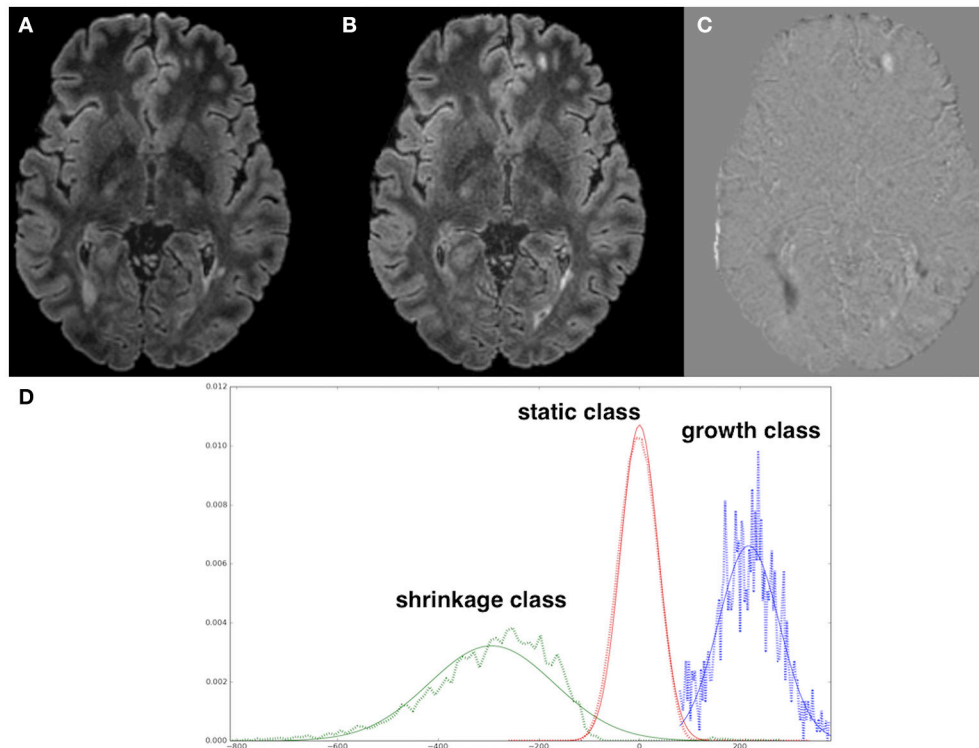


FIGURE 3 | (A) Normalized FLAIR image of time point 1, (B) FLAIR image of time point 2, (C) difference image (A,B), (D) histograms of difference image classes with corresponding Gaussian fitting, normalized per class. Note the artifactual difference values at the brain contour (due to subtle differences in brain mask extraction) are excluded by only including WM voxels in the analysis.

for image 1 and image 2 are denoted by $P(L_1)$ and $P(L_2)$, respectively.

2.3.2. The Model

Under these assumptions, the joint probabilistic model is formulated as follows:

$$P(I_1, I_2, D, L_1, L_2, \gamma) = P(I_1|L_1, \theta_1) \cdot P(I_2|L_2, \theta_2) \cdot P(D|L_1, L_2, \zeta) \cdot P(L_1) \cdot P(L_2) \quad (1)$$

where $\gamma = \{\theta_1, \theta_2, \zeta\}$. Our model is optimized by the maximum a posteriori (MAP) problem shown in Equation (2). Since the knowledge of tissue class labels helps in finding the model parameters and vice-versa, we reformulate our MAP problem as presented in Equation (3).

$$\hat{\gamma}_{\text{MAP}} = \underset{\gamma}{\operatorname{argmax}} \ln P(\gamma|I_1, I_2, D) = \underset{\gamma}{\operatorname{argmax}} \ln P(I_1, I_2, D, \gamma) \quad (2)$$

$$= \underset{\gamma}{\operatorname{argmax}} \ln \sum_{L_1, L_2} P(I_1, I_2, D, L_1, L_2, \gamma) \quad (3)$$

$$\geq \underset{\gamma}{\operatorname{argmax}} \sum_{L_1, L_2} P(L_1, L_2|I_1, I_2, D, \bar{\gamma}) \cdot \ln \frac{P(I_1, I_2, D, L_1, L_2, \gamma)}{P(L_1, L_2|I_1, I_2, D, \bar{\gamma})} \quad (4)$$

Finally, a lower bound of our model is derived using Jensen's inequality and optimized by the EM algorithm. The Q-function, which is the log likelihood function whose expected value is computed in the E-step can now be written as:

$$Q(\gamma|\bar{\gamma}) = E_{L_1, L_2|I_1, I_2, D, \bar{\gamma}} [\ln P(I_1, I_2, D, L_1, L_2, \gamma)] \quad (5)$$

with the joint posterior distribution $P(L_1, L_2|I_1, I_2, D, \bar{\gamma})$. The sum over all possible tissue classes $k^{(2)}$ of the joint posterior distribution gives us the soft segmentation of the tissue class at time point 1. Similarly, the sum over all possible tissue classes $k^{(1)}$ of the joint posterior distribution gives us the soft segmentation of the tissue class at time point 2.

In the M-step, a new set of values for model parameter γ is computed by maximizing the Q-function (see Supplementary Material for closed form solutions).

2.4. Pruning

The soft lesion segmentations obtained from the E-step of the joint EM algorithm are pruned to eliminate non-lesions (such as partial volume effects, artifacts) that share intensities and locations with the potential lesions. Thereto, a priori information on the appearance, location and volume of lesions is incorporated: (1) the lesion intensities should be hyper-intense compared to the WM intensities on bias field corrected FLAIR image, (2) the lesions are in the WM region, and (3) the lesion

needs to have a minimum volume of 0.005 ml (empirically determined) to avoid spurious lesion detection. The hyper-intensity is defined as the mean plus two times the standard deviation of WM intensities. The intensities and location of WM region are computed using the WM segmentation from the MSmetrix-cross pipeline. In addition, a priori defined binary mask (defined in the MNI space and consisting of the cerebral cortex and WM in-between the ventricles) is warped to the subject space to remove lesion candidates from these regions that are likely to result in a false lesion segmentation. After the pruning, the soft lesion segmentations are binarized using a threshold of 0.9 (empirically determined) on the posterior probabilities.

2.5. Performance Tests

2.5.1. Comparison with State-of-the-Art Methods

We compare MSmetrix-long pipeline with the MSmetrix-cross pipeline to know the gain over the cross-sectional method. Furthermore, we also compare against the longitudinal pipeline of the Lesion Segmentation Toolbox (LST) software package (LST¹), version 2.0.12, which is implemented in SPM12 (SPM²). The longitudinal pipeline of LST, which is referred to as LST-long in this paper, performs individual time point lesion segmentation using the lesion growth algorithm described in Schmidt et al. (2012). The obtained lesion segmentation maps of different time points are coregistered to the baseline scan and are corrected by comparing the relative differences of FLAIR intensities in all lesion maps to produce the final lesion segmentation at each time point (see LST documentation, LST¹).

For comparison, all three methods were executed on the same datasets and default parameter settings were used. Thus, no parameter tuning was performed at dataset or subject level.

2.5.2. Data

Dataset 1 contains scans from 12 relapsing remitting MS patients on a GE 3T scanner (Discovery MR750), each scanned twice at an interval of approximately 1 year. Therefore, the sample size of dataset 1 equals 24. Each time point contained two a 3D sequences: a CUBE FLAIR (TR: 8000 ms, TE: 165 ms, TI: 2179

ms) and a 3D T1-weighted IR-FSPGR sequence (TR 7.2 ms, TE 450 ms, TI 2.8 ms). Both 3D sequences have voxel resolution close to 1 mm³. Expert WM lesion segmentations were created on the baseline FLAIR scan by the experienced neuro-imaging analyst using JIM software tool (JIM³), version 6.0. For follow-up scans, baseline lesion segmentation was overlaid on rigidly registered follow-up scan at the beginning, and then the lesion segmentation was adapted according to lesion activities. This study was reviewed and conducted within the guidelines set out in the National Statement on Ethical Conduct in Human Research (2007) in Australia, and approved by University of Sydney Human Research Ethics Committee. All subjects gave written informed consent.

The second dataset, dataset 2 contains scans from 10 MS patients scanned twice, with re-positioning (time interval between two scans is 5~10 min), on each of three different 3T scanners from GE (Discovery MR750w), SIEMENS (Skyra) and PHILIPS (Achieva). Therefore, the sample size of dataset 2 equals 60. The protocol contained two 3D sequences: T1-weighted and FLAIR, and their details are described in **Table 1**. For this dataset, no expert segmentations were available. This study was carried out in accordance with the recommendations of the “International Conference on Harmonization of Good Clinical Practice (ICH-GCP),” and the applicable Belgian and Dutch legislation. The study was approved by the UZ Brussels ethical committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

2.5.3. Accuracy and Reproducibility Assessment

The agreement between the expert segmentation and automatic methods on dataset 1 is evaluated at three levels: voxel-by-voxel, lesion-wise and volumetric. Voxel-by-voxel metric includes the Dice similarity index which is defined as the ratio of total number of lesion voxels where both the expert reference and the automatic segmentation agree (true positives) to the mean number of voxels labeled as lesion by the two methods. The lesion-wise metrics include lesion-wise true positive rate (LTPR), false positive rate, F1 score, absolute lesion change difference and Pearson correlation coefficient (PCC). LTPR is defined as the

¹ www.statistical-modelling.de/lst.html

² <http://www.fil.ion.ucl.ac.uk/spm/software/spm12>

³ <http://www.xinapse.com>

TABLE 1 | Dataset-2 sequences description for all three scanners.

Sequence	TR (ms)	TE (ms)	TI (ms)	FOV (mm ²)	Voxel size (mm ³)	No. of slices (sagittal)
GE						
3D T1-weighted FSPGR	7.32	3.14	NA	220 × 220	0.43 × 0.43 × 0.50	328
Fat saturated 3D FLAIR	9500	135.78	2428	240 × 240	0.47 × 0.47 × 0.70	232
SIEMENS						
3D T1-weighted MPRAGE	2300	2.29	NA	240 × 240	0.94 × 0.94 × 0.94	176
Fat saturated 3D FLAIR	5000	387	1800	230 × 230	0.45 × 0.45 × 0.90	192
PHILIPS						
3D T1-weighted FSPGR	4.93	2.3	NA	230 × 230	0.53 × 0.53 × 0.50	310
Fat saturated 3D FLAIR	4800	276	1650	240 × 240	1.04 × 1.04 × 0.56	321

NA, Not available.

ratio of the total number of lesions where the expert reference and the automatic segmentation intersect to the total number of lesions in the expert reference segmentation. Lesion-wise false positive rate (LFPR) is defined as the ratio of the total number of lesions that are present only in the automatic segmentation to the total number of lesions in the automatic segmentation. Lesion-wise F1 score is defined as the harmonic mean of LTPR and LFPR. Absolute lesion-wise change difference is defined as the absolute difference between the overall lesion-wise change (number of new lesions minus number of disappearing lesions) in the expert lesion segmentation and the automatic segmentation. In this paper, we consider new, disappearing, enlarging and shrinking lesions that have size more than 20 voxels and at least one slice which encompasses the lesion presents a minimum of 5 lesion voxels.

Volumetric metrics measure the total lesion volume agreement and consist of the PCC and the absolute volume difference. The absolute volume difference is computed as the absolute difference between the total volume reported by the expert reference segmentation and the corresponding value derived from the automatic method.

The reproducibility of the method is evaluated on dataset 2 by the Dice similarity index of the lesion segmentations at both times points. Moreover, the estimated number of new lesions and the absolute total lesion volume difference is also calculated between time points, which are both expected to be zero in this test-retest scenario.

To determine if there is a statistical difference between MSmetrix-long and LST-long and between MSmetrix-cross and MSmetrix-long methods' performance, two tailed paired Wilcoxon signed-rank test is performed.

3. RESULTS

3.1. Accuracy Results on Dataset 1

Figure 4 shows a representative example of lesion segmentation obtained by MSmetrix-cross, MSmetrix-long and LST-long on a patient from dataset 1. By comparing against expert delineations, it can be observed that MSmetrix-long has improved in accuracy over MSmetrix-cross and that LST-long has missed lesions.

The volumetric correlation of MSmetrix-long and LST-long to the expert reference segmentation can be visualized in **Figure 5**. MSmetrix-long has a better correlation ($PCC = 0.96$) with expert reference segmentation compared to LST-long ($PCC = 0.88$).

Table 2 summarizes the cross-sectional lesion segmentation performance of MSmetrix-cross, MSmetrix-long and LST-long on dataset 1 ($n = 24$) in a quantitative way. MSmetrix-long has improved over MSmetrix-cross in the median Dice, F1 score and LFPR. Compared to LST-long, MSmetrix-long has a higher median Dice, F1 score, LTPR, and PCC, together with lower LFPR and absolute lesion volume difference.

Table 3 summarizes the lesion-wise change accuracy performance of MSmetrix-cross, MSmetrix-long and LST-long on dataset 1 in a quantitative way. In case of new lesions, MSmetrix-long has improved over MSmetrix-cross in the median F1 score and LFPR. Compared to LST-long, MSmetrix-long has a higher median F1 score and LTPR. In case of enlarging

lesions, MSmetrix-long has improved over MSmetrix-cross in the median LFPR, with marginally better F1 score. Compared to LST-long, MSmetrix-long has a higher median F1 score, LTPR, and LFPR. When new and enlarging lesions are combined, MSmetrix-long has better correlation ($PCC = 0.77$) with the expert segmentations compared to MSmetrix-cross ($PCC = 0.63$) and LST-long ($PCC = 0.53$). In case of absolute lesion-wise change difference, MSmetrix-long has marginally better performance over MSmetrix-cross and LST-long, however, with better correlation with the lesion-wise change difference of the expert segmentations ($PCC = 0.84$) compared to MSmetrix-cross (0.65) and LST-long (0.72).

3.2. Reproducibility Results on Dataset 2

Figure 6 shows an example of lesion segmentation obtained by MSmetrix-cross, MSmetrix-long and LST-long on a patient from dataset 2 ($n = 60$). Both MSmetrix-long and LST-long are more consistent in lesion segmentation compared to MSmetrix-cross. Compared to LST-long, MSmetrix-long also shows better reproducibility in segmenting small lesions. Quantitatively, LST-long has the best median Dice with zero error in detecting new lesions and absolute volume difference between both time points. MSmetrix-long has improved in the median Dice, with median error in detecting new lesions and absolute volume difference over MSmetrix-cross. The reproducibility of LST-long is highest because it segments the most certain hyper-intense lesions in both time points at the expense of missing substantial amount of less hyper-intense lesions as shown in **Figure 6**.

4. DISCUSSION AND CONCLUSIONS

Accurate and consistent lesion segmentation is very important in monitoring the MS disease progression. As manual lesion segmentation is time consuming and suffers from inter- and intra-rater variability, automated methods have the advantage of being fast and consistent. The vast majority of automatic methods are cross-sectional in nature and the average accuracy (Dice) of these methods is sufficiently high, however, these cross-sectional methods seldom report results on the lesion evolution accuracy and this hinders a fair comparison of our method against them. Moreover, another factor to consider is whether the segmentation method is supervised or unsupervised. We compare our unsupervised method with other unsupervised methods only because supervised methods often require a representative training dataset, including expert segmentation, in order to build a model that can be used on new patients for lesion segmentation. This training dataset is very difficult to build because MS lesions have all possible shapes, intensities and are heterogeneously distributed in the WM. Moreover, the new image to be segmented should be well represented in the training dataset which is not always possible. Two well-known publicly available unsupervised MS lesion segmentation tools are Lesion-TOADS (Shiee et al., 2010) and LST. We choose LST because of two reasons: (1) in a previous paper (Jain et al., 2015), we have shown that our cross-sectional method (MSmetrix-cross) had a better performance compared to Lesion-TOADS in terms of accuracy and reproducibility. Since in this

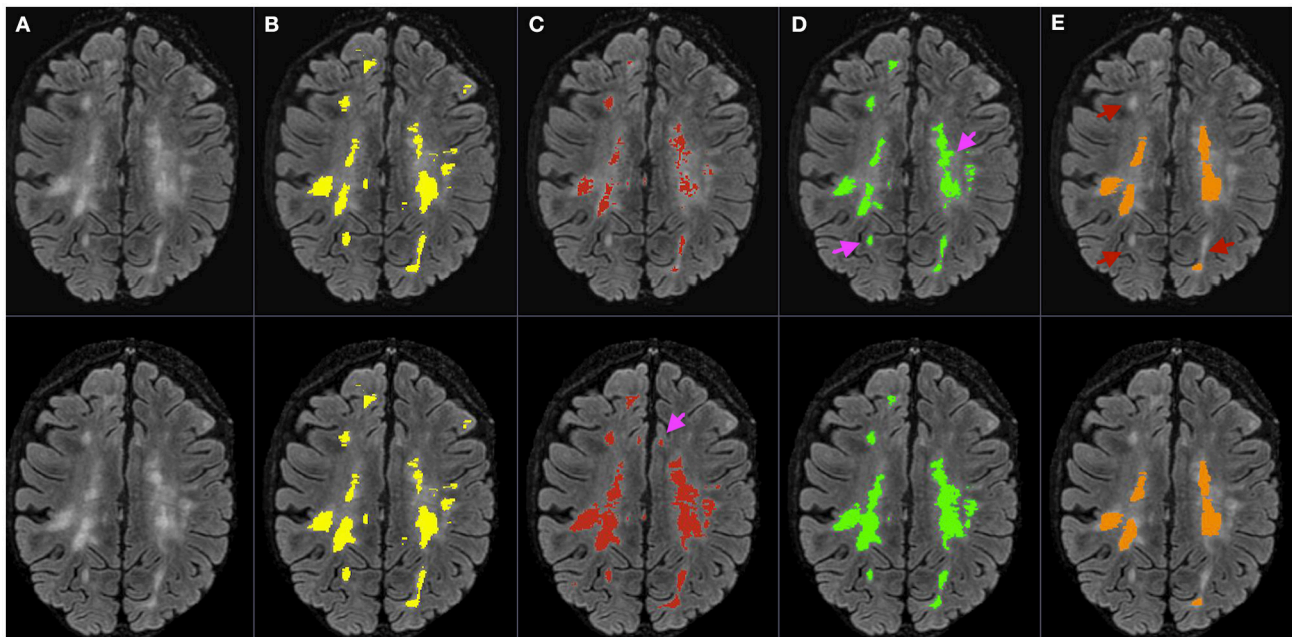


FIGURE 4 | Bias corrected FLAIR image (A) followed by super-imposed lesion segmentation from: **(B)** expert segmentation, **(C)** MSmetrix-cross (version 1.4), **(D)** MSmetrix-long, and **(E)** LST-long. The first row corresponds to the lesion segmentation of time point 1 and the second row corresponds to the lesion segmentation of time point 2. Pink arrows specify places where MSmetrix-long has improved in accuracy over MSmetrix-cross and red arrows indicate regions where LST-long has missed lesions.

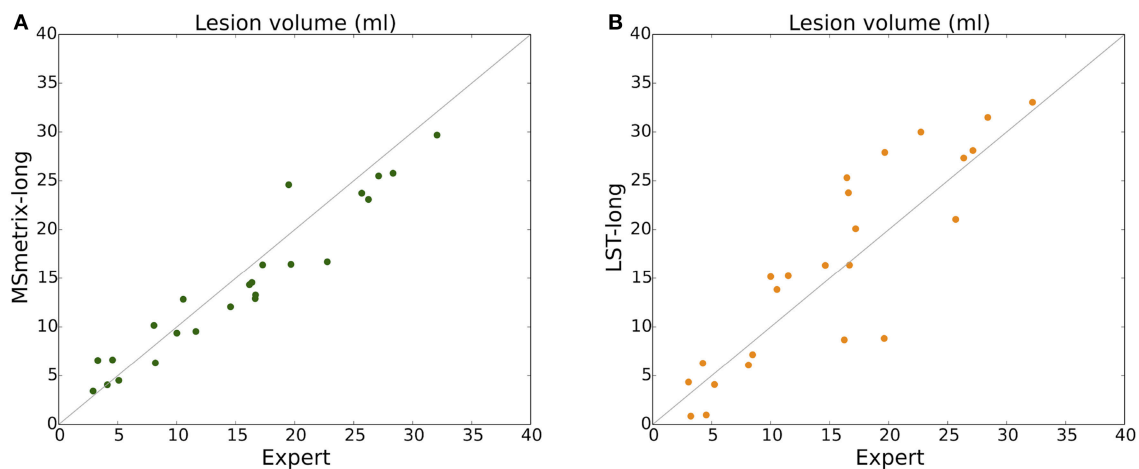


FIGURE 5 | Scatter plot of total lesion volume (ml) for reference expert segmentation vs. (A) MSmetrix-long and (B) LST-long.

paper we also report results from our cross-sectional method, we decided that the comparison with Lesion-TOADS is not required, (2) only LST tool has a longitudinal MS lesion segmentation pipeline. Thus it is logical to compare MSmetrix-long with LST-long as both methods are unsupervised and longitudinal in nature.

In this paper, MSmetrix-long pipeline combines both spatial and temporal relationships of lesions for accurate and consistent

lesion segmentation. The spatial relationship is based on Markov Random Field and is incorporated in MSmetrix-cross. The temporal relationship is modeled in a joint lesion segmentation, which uses difference image and cross-sectional lesion segmentations of two time points. The difference image models the growth and shrinkage of lesions and thus helps in recovering those lesions that are missed by the cross-sectional lesion segmentation. In addition, if a lesion is present in both

TABLE 2 | Quantitative metrics (voxel-by-voxel, lesion and volumetric level) for measuring the cross-sectional accuracy of the automatic methods MSmetrix-long, MSmetrix-cross and LST-long with respect to expert segmentations on dataset 1 ($n = 24$).

Automatic method	Dice	F1 score	LTPR	LFPR	Absolute volume difference (ml)	PCC
MSmetrix-long	0.63 (0.49–0.68)	0.61 (0.54–0.63)	0.50 (0.43–0.59)	0.25 (0.20–0.37)	2.09 (1.77–3.18)	0.96
MSmetrix-cross	0.60 (0.46–0.66)**	0.56 (0.52–0.61)*	0.57 (0.52–0.65)**	0.48 (0.36–0.55)**	1.48 (0.81–2.59)	0.95
LST-long	0.60 (0.47–0.65)*	0.48 (0.37–0.53)**	0.42 (0.30–0.52)*	0.40 (0.30–0.47)	2.66 (1.52–4.84)*	0.88

Except PCC, all metrics are reported in median (first quartile–third quartile). LTPR, lesion-wise true positive rate; LFPR, lesion-wise false positive rate; PCC, Pearson correlation coefficient.

*Values significantly different from MSmetrix-long (paired Wilcoxon signed-rank test with $p < 0.05$ significance level).

**Values significantly different from MSmetrix-long (paired Wilcoxon signed-rank test with $p < 0.01$ significance level).

TABLE 3 | Lesion-wise quantitative metrics for measuring the lesion change accuracy of the automatic methods MSmetrix-long, MSmetrix-cross and LST-long with respect to expert lesion segmentations changes on dataset 1.

	New lesions			Enlarging lesions			New and enlarging lesions
	F1 score	LTPR	LFPR	F1 score	LTPR	LFPR	
MSmetrix-long	0.42 (0–0.55)	0.33 (0–0.60)	0 (0–0.38)	0.69 (0.56–0.81)	0.62 (0.53–0.69)	0.16 (0–0.51)	0.77
MSmetrix-cross	0.20 (0.0–0.62)	0.33 (0–0.52)	0.50 (0.31–0.75)	0.68 (0.58–0.80)	0.59 (0.53–0.69)	0.24 (0.15–0.43)	0.63
LST-long	0 (0–0.43)	0 (0–0.29)	0 (0–0)	0.60 (0.51–0.69)	0.50 (0.35–0.60)	0.33 (0.15–0.51)	0.53

Absolute lesion-wise change difference

	PCC	
MSmetrix-long	1 (1–3.5)	0.84
MSmetrix-cross	1.5 (1–3.75)	0.65
LST-long	2 (1–3.5)	0.72

Except PCC, all metrics are reported in median (first quartile–third quartile). PCC, Pearson correlation coefficient. Here, the t-test is not performed, as the sample size is small ($n = 12$).

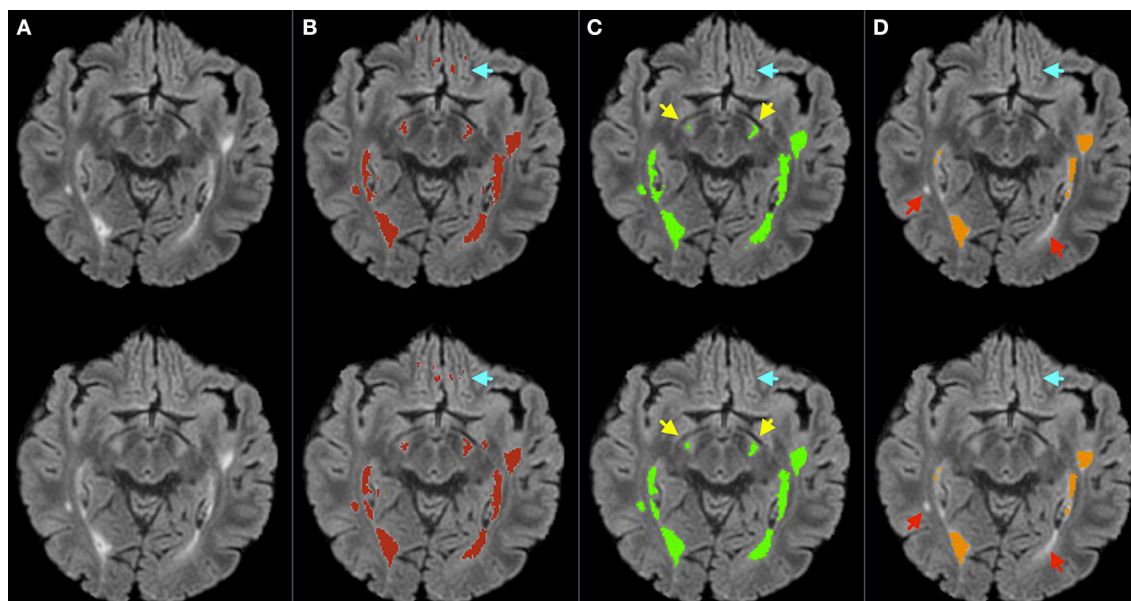


FIGURE 6 | Bias corrected FLAIR image (A) followed by super-imposed lesion segmentation from: (B) MSmetrix-cross (version 1.4), (C) MSmetrix-long, and (D) LST-long. The first row corresponds to the lesion segmentation of time point 1 and the second row corresponds to the lesion segmentation of time point 2. Cyan arrows show some false positives in MSmetrix-cross, which are absent in MSmetrix-long. Yellow arrows specify places where MSmetrix-long has consistently segmented some small lesions and red arrows indicate regions where LST-long misses some potential lesions.

time points but has been segmented in only one of the time point, then the joint lesion segmentation facilitates the recovery of that lesion at the other time point. Moreover, brain atrophy has also minimal impact on the performance of MSmetrix-long because (1) atrophy is generally small and global in nature (2) it occurs near the CSF boundary and these transitions *i.e.* (CSF \rightarrow GM and CSF \rightarrow WM) are excluded in the difference image GMM model, (3) we tested global non-rigid registration in addition to affine registration, *i.e.*, non-rigid registration only on a coarse level, to accommodate for the atrophy and we found out that it has a minimal, but potentially negative impact on the final lesion segmentation. Therefore, to gain computational efficiency we excluded this global non-rigid registration from MSmetrix-long pipeline. Furthermore, if the subject has been scanned more than twice, MSmetrix-long can easily handle this by processing consecutive time points in pairs.

Among the methods proposed in the literature for longitudinal lesion segmentation, our approach has some similarities to Elliott et al. (2013) and Ait-Ali et al. (2005), which are also based on EM frameworks. In contrast with Elliott et al. (2013), our method is unsupervised and can segment new, enlarging, disappearing and shrinking lesions. As opposed to Ait-Ali et al. (2005), our joint EM model takes cross-sectional lesion segmentation as prior information on the lesion class in both time points and processes each time point in its own space to avoid bias in the lesion segmentation.

In order to evaluate the effect of the pruning step, we also calculated the cross-sectional accuracy (Dice, LTPR and LFPR) of MSmetrix-long after the joint lesion segmentation step. The Dice, LTPR, and LFPR (reported in median (first quartile–third quartile)) after the joint lesion segmentation step are 0.60 (0.45–0.65), 0.64 (0.54–0.69), and 0.81 (0.72–0.87) respectively. Comparing these results with the voxel-by-voxel accuracy of MSmetrix-long after the pruning step (see **Table 2**), we observe that the pruning step increases the overall Dice score by decreasing the false positive rate at the expense of a decrease in true positive rate.

In order to investigate the cause of low LTPR for cross-sectional accuracy of MSmetrix-long compared to MSmetrix-cross (see **Table 2**), we calculated the average LTPR for small (0.003–0.01 ml), medium (0.01–0.05 ml) and large (>0.05 ml) lesion volumes. The average LTPR for MSmetrix-long and MSmetrix-cross for small lesions is 0.13 and 0.27 respectively, followed by medium lesions 0.30 and 0.37 and large lesions 0.75 and 0.81. It can be seen that MSmetrix-long misses more small and medium size lesions. The primary cause of missing these lesions is that they are either iso-intense with GM intensities (thus missed by intensity threshold mask used in the pruning step) or they are removed by the binary false positive mask (used in the pruning step). However, it is important to note that both intensity threshold mask and binary false positive mask play a key role in reducing the false positives as described in the previous paragraph.

One important aspect of MSmetrix-long is that its performance is dependent on the cross-sectional lesion segmentation. This suggests that if MSmetrix-cross has either consistently missed a lesion, or segmented a non-lesion at

both time points, then it will be either missed or retained by MSmetrix-long, respectively. As presented in the result section, MSmetrix-long is more accurate and reproducible than MSmetrix-cross. The increase in cross-sectional accuracy (Dice, F1 score) and lesion change accuracy for new lesions (F1 score) is due to the reduction in LFPR using the lesion segmentation information from the other time point. For enlarging lesions, a marginal increase in the median F1 score is observed for MSmetrix-long due to larger differences in the lesion segmentation boundary between the expert and MSmetrix-long. MSmetrix-long has also slightly better absolute lesion-wise change difference compared to MSmetrix-cross primarily due to a reduction in LFPR. A modest decrease in the absolute volume difference is due to the under-segmentation of lesions by MSmetrix-long (**Figure 5**) and the elimination of a few lesions that are close to the cerebral cortex. Interestingly, a substantial LFPR in MSmetrix-cross suggests that the false lesions compensate toward missed lesions volume resulting in a lower absolute volume difference compared to MSmetrix-long. The significant improvement in reproducibility (Dice, number of new lesions and absolute volume difference) of MSmetrix-long could also be explained by the benefit of using the lesion segmentation of the other time point.

In comparison to LST-long, MSmetrix-long is more accurate (Dice, F1 score) and slightly less reproducible. Cross-sectionally, LST-long has higher absolute volume difference and LFPR; lower LTPR and F1 score on dataset 1. The high absolute volume difference of LST-long could be explained by the over-segmentation of lesion boundaries. A high lesion-wise false positive rate of LST-long could be explained by the segmentation of FLAIR artifacts or cortical foldings as lesions. For the lesion change accuracy, MSmetrix-long has superior performance for all measures compared to LST-long. This could be explained by the fact that LST-long segments the most hyper-intense lesions and is thus very consistent (see **Table 4**), but misses many small less hyper-intense lesions (**Figures 4, 6**).

In conclusion, we have presented MSmetrix-long: an iterative two time point WM lesion segmentation method based on a joint EM framework using two time points. The proposed method is unsupervised and can segment new, enlarging, disappearing, shrinking and static lesions. We first analyse both time points

TABLE 4 | The Dice score, the number (Nr.) of new lesions and the absolute volume difference (Abs. vol. diff.) between both time points for measuring the accuracy of the automatic methods MSmetrix-long, MSmetrix-cross and LST-long on dataset 2.

	Dice	Nr. of new les	Abs. vol. diff. (ml)
MSmetrix-long	0.89 (0.85–0.91)	0 (0–1)	0.11 (0.03–0.32)
MSmetrix-cross	0.69 (0.56–0.73)**	3.5 (1–5)**	0.3 (0.17–0.54)*
LST-long	1 (1–1)**	0 (0–0)**	0 (0–0.01)**

All metrics are reported as median (first quartile–third quartile).

*Values significantly different from MSmetrix-long (paired Wilcoxon signed-rank test with $p < 0.05$ significance level).

**Values significantly different from MSmetrix-long (paired Wilcoxon signed-rank test with $p < 0.01$ significance level).

separately followed by a joint lesion segmentation, which models the lesion evolution as a Gaussian mixture model. The accuracy and reproducibility of MSmetrix-long is compared with MSmetrix-cross and the publicly available lesion segmentation tool LST-long on two datasets that are representative for clinically feasible acquisition protocols. MSmetrix-long has outperformed MSmetrix-cross. Compared to LST-long, MSmetrix-long has better accuracy and similar reproducibility.

AUTHOR CONTRIBUTIONS

SJ, AR, DMS, SV, FM, and DS contributed to the design and analysis of the work; MC, JD, CW, and MB contributed to the data acquisition; SJ, AR and DMS wrote the paper; all authors

revised the manuscript critically for important intellectual content.

FUNDING

This study has been supported by TRANSACT (FP7-PEOPLE-2012-ITN-316679), CENTER-TBI (FP7-COOPERATION-2013-602150) and BRAINPATH (FP7-PEOPLE-2013-IAPP-612360).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2016.00576/full#supplementary-material>

REFERENCES

- Ait-Ali, L. S., Prima, S., Hellier, P., Carsin, B., Edan, G., and Barillot, C. (2005). STREM: a robust multidimensional parametric method to segment MS lesion in MRI. *Med. Image Comput. Comput. Assist. Interv.* 8, 409–416. doi: 10.1007/11566465_51
- Bernardis, E., Pohl, K. M., and Davatzikos, C. (2013). “Extracting evolving pathologies via spectral clustering,” in *Information Processing in Medical Imaging, 23rd International Conference, IPMI* (Asilomar), 680–691.
- Blystad, I., Håkansson, I., Tisell, A., Ernerudh, J., Smedby, Ö., Lundberg, P., et al. (2016). Quantitative MRI for analysis of active multiple sclerosis lesions without gadolinium-based contrast agent. *Am. J. Neuroradiol.* 37, 94–100. doi: 10.3174/ajnr.A4501
- Bosc, M., Heitz, F., Armspach, J. P., Namer, I., Gounot, D., and Rumbach, L. (2003). Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *Neuroimage* 20, 643–656. doi: 10.1016/S1053-8119(03)00406-3
- Cardoso, M. J. (2012). *NiftySeg: Statistical Segmentation and Label Fusion Software Package*. Available online at: <http://niftyseg.sourceforge.net/index.html> [Accessed: 25-04-2016].
- Castleman, K. R. (1995). *Digital Image Processing, 2nd Edn*. Englewood: Prentice Hall.
- Deeks, E. D. (2016). Dimethyl fumarate: a review in relapsing-remitting MS. *Drugs* 76, 243–254. doi: 10.1007/s40265-015-0528-1
- Elliott, C., Arnold, D. L., Collins, D. L., and Arbel, T. (2013). Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. *IEEE Trans. Med. Imaging* 32, 1490–1503. doi: 10.1109/TMI.2013.2258403
- Erbayat Altay, E., Fisher, E., Jones, S. E., Hara-Cleaver, C., Lee, J. C., and Rudick, R. A. (2013). Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. *JAMA Neurol.* 70, 338–344. doi: 10.1001/2013.jamaneurol.211
- García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med. Image Anal.* 17, 1–18. doi: 10.1016/j.media.2012.09.004
- Gerig, G., Welti, D., Guttmann, C. R. G., Colchester, A. C. F., and Székely, G. (2000). Exploring the discrimination power of the time domain for segmentation and characterization of active lesions in serial MR data. *Med. Image Anal.* 4, 31–42. doi: 10.1016/S1361-8415(00)00005-0
- Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., et al. (2015). Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *Neuroimage Clin.* 8, 367–375. doi: 10.1016/j.nicl.2015.05.003
- Lewis, E. B., and Fox, N. C. (2004). Correction of differential intensity inhomogeneity in longitudinal MR images. *Neuroimage* 23, 75–83. doi: 10.1016/j.neuroimage.2004.04.030
- Lladé, X., Ganiér, O., Oliver, A., Martí, R., Freixenet, J., Valls, L., et al. (2012). Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology* 54, 787–807. doi: 10.1007/s00234-011-0992-6
- Metcalfe, D., Kikinis, R., Guttmann, C., Vaina, L., and Jolesz, F. (1992). “4D connected component labelling applied to quantitative analysis of MS lesion temporal development,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Paris), 945–946.
- Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., et al. (2010). Fast free-form deformation using graphics processing units. *Comput. Methods Prog. Biomed.* 98, 278–284. doi: 10.1016/j.cmpb.2009.09.002
- Prima, S., Ayache, N., Janke, A., Francis, S. J., Arnold, D. L., and Collins, D. L. (2002). Statistical analysis of longitudinal MRI data: applications for detection of disease activity in MS. *Med. Image Comput. Comput. Assist. Interv.* 2488, 363–371. doi: 10.1007/3-540-45786-0_45
- Rey, D., Subsol, G., Delingette, H., and Ayache, N. (2002). Automatic detection and segmentation of evolving processes in 3D medical images: application to multiple sclerosis. *Med. Image Anal.* 6, 163–179. doi: 10.1016/S1361-8415(02)00056-7
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förchler, A., Berthele, A., et al. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *Neuroimage* 59, 3774–3783. doi: 10.1016/j.neuroimage.2011.11.032
- Shiee, N., Bazin, P. L., Ozturk, A., Reich, D. S., Calabresi, P. A., Pham, D. L., et al. (2010). A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *Neuroimage* 49, 1524–1535. doi: 10.1016/j.neuroimage.2009.09.005
- Smeets, D., Ribbens, A., Sima, D. M., Cambron, M., Horakova, D., Jain, S., et al. (2016). Reliable measurements of brain atrophy in individual patients with multiple sclerosis. *Brain Behav.* 6:e00518. doi: 10.1002/brb3.518
- Solomon, J., and Sood, A. (2004). “4-D lesion detection using expectation-maximization and hidden markov model,” in *IEEE International Symposium Biomedical Imaging* (Arlington, VA), 125–128.
- Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999). Automated model-based tissue classification of mr images of the brain. *IEEE Transactions of the Medical Imaging* 18, 897–908. doi: 10.1109/42.811270
- Welti, D., Gerig, G., Radü, E.-W., Kappos, L. and Székely, G. (2001). “Spatio-temporal segmentation of active multiple sclerosis lesions in serial MRI data,” in *Information Processing in Medical Imaging, 17th International Conference, IPMI* (Davis, CA), 438–445.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Jain, Ribbens, Sima, Cambron, De Keyser, Wang, Barnett, Van Huffel, Maes and Smeets. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Age and Glaucoma-Related Characteristics in Retinal Nerve Fiber Layer and Choroid: Localized Morphometrics and Visualization Using Functional Shapes Registration

OPEN ACCESS

Edited by:

Xiaoying Tang,
SYSU-CMU Joint Institute of
Engineering, China

Reviewed by:

Delia Cabrera DeBuc,
University of Miami, United States
Martin Reuter,
Harvard Medical
School/Massachusetts General
Hospital, United States

*Correspondence:

Mirza Faisal Beg
mfbegl@sfu.ca

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 22 February 2017

Accepted: 19 June 2017

Published: 12 July 2017

Citation:

Lee S, Heisler ML, Popuri K,
Charon N, Charlier B, Trouvé A,
Mackenzie PJ, Sarunic MV and
Beg MF (2017) Age and
Glaucoma-Related Characteristics in
Retinal Nerve Fiber Layer and Choroid:
Localized Morphometrics and
Visualization Using Functional Shapes
Registration. *Front. Neurosci.* 11:381.
doi: 10.3389/fnins.2017.00381

Sieun Lee¹, Morgan L. Heisler¹, Karteek Popuri¹, Nicolas Charon², Benjamin Charlier³,
Alain Trouvé⁴, Paul J. Mackenzie⁵, Marinko V. Sarunic¹ and Mirza Faisal Beg^{1*}

¹ Faculty of Applied Sciences, School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada, ² Center for Imaging Sciences, Johns Hopkins University, Baltimore, MD, United States, ³ Institut Montpellierain Alexander Grothendieck, CNRS, Université de Montpellier, Montpellier, France, ⁴ CMLA, ENS Cachan, Centre National de la Recherche Scientifique, Université Paris-Saclay, Cachan, France, ⁵ Department of Ophthalmology and Visual Sciences, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Optical coherence tomography provides high-resolution 3D imaging of the posterior segment of the eye. However, quantitative morphological analysis, particularly relevant in retinal degenerative diseases such as glaucoma, has been confined to simple sectorization and averaging with limited spatial sensitivity for detection of clinical markers. In this paper, we present point-wise analysis and visualization of the retinal nerve fiber layer and choroid from cross-sectional data using functional shapes (fshape) registration. The fshape framework matches two retinas, or generates a mean of multiple retinas, by jointly optimizing the surface geometry and functional signals mapped on the surface. We generated group-wise mean retinal nerve fiber layer and choroidal surfaces with the respective layer thickness mapping and showed the difference by age (normal, younger vs. older) and by disease (age-matched older, normal vs. glaucomatous) in the two layers, along with a more conventional sector-based analysis for comparison. The fshape results visualized the detailed spatial patterns of the differences between the age-matched normal and glaucomatous retinal nerve fiber layers, with the glaucomatous layers most significantly thinner in the inferior region close to Bruch's membrane opening. Between the young and older normal cases, choroid was shown to be significantly thinner in the older subjects across all regions, but particularly in the nasal and inferior regions. The results demonstrate a comprehensive and detailed analysis with visualization of morphometric patterns by multiple factors.

Keywords: optical coherence tomography, computational anatomy, Bayesian estimation, retina, glaucoma, aging

INTRODUCTION

Volumetric optical coherence tomography (OCT) has emerged as a preferred diagnostic tool in ophthalmology for noninvasive, *in-vivo*, micrometer-resolution imaging of the eye. Recent progress in OCT imaging has allowed the acquisition of highly detailed 3D images from which morphometric measurements can be derived. Optic nerve head measurements and the thickness of the peri-papillary retinal layers have been used clinically to detect and monitor glaucoma progression (Leung, 2014). While these measurements are useful individually, a lack of clear spatial references for biomarkers limits the spatial and anatomical correspondence across multiple images. For example, in the common sectoral layer thickness analysis of OCT scans, the peri-papillary area is split into quadrants (superior, inferior, nasal, and temporal) which are then further subdivided into circumferential areas. This analysis relies on averaging over the local sectors to reduce noise, and to mitigate potential inconsistency in sectoral placement across individuals. Such a sectoral averaging approach is limited by a minimum size of sectors to achieve comparisons in the same vicinity in the different individuals. Hence, the sectorization approach reduces the spatial sensitivity of the measurements due to averaging over larger areas and could potentially impact clinical assessment. This motivates the need to develop tools that can generate measurements on a point-to-point basis, eliminating the need to average data in local regions.

Most previous studies involving registration of OCT images have averaged multiple serially acquired images for noise reduction and motion correction (Jørgensen et al., 2007; Young et al., 2011), or utilized rigid alignment of time-course images (Niemeijer et al., 2009). Relatively little attention has been given to registration of cross-sectional OCT data. Chen et al. (2014) performed intensity-based non-rigid registration of macular OCT scans by a combination of rigid alignment of foveae, and A-scan-wise affine and non-rigid registration using radial basis functions for refined alignment of the retinal layers (Chen et al., 2014). A more recent work (Anthony et al., 2016) by the same group applies this registration technique to perform voxel based morphometry in macular OCT of healthy controls and multiple sclerosis patients. Our group's approach has focused on retinal surface-based registrations and atlas generation. In (Gibson et al., 2010), 3D optic cup surfaces were registered to a single template surface, first by rigid and nonrigid intensity-based volumetric registration, followed by spherical mapping and spherical demons registration of the surfaces. This work was further expanded upon in Lee et al. (2015) which represented the retinal surfaces utilizing the framework of mathematical currents. Two surfaces were brought into close proximity by minimizing a functional of reproducing kernel Hilbert space norm-based energy and a dissimilarity term, then registered by spherical demons to establish homology. More recently, we introduced the functional shape (fshape), framework (Charlier et al., 2015; Lee et al., 2017). In this framework, the retinal surface (shape or geometry) and any value mapped on the surface, for example, retinal layer thickness (function or signal), are considered together as a single mathematical object called *fshape*. One *fshape*

can be matched to another or the mean of multiple *fshapes* can be computed by joint optimization of the energies associated to varifold-based dissimilarity measures of geometry and function. For group analyses, the algorithm can generate population atlases and establish homology across the database, facilitating comparison of morphometric measurements in localized regions. Moreover, *fshape* mean computation or matching can be performed with flexible constructions of *fshapes* that can include multiple geometric shapes and function parameters; this allows the building of specific sets of geometry and function features to compare across multiple groups.

In this paper, we demonstrate the use of this algorithm for investigating the effect of age and glaucoma on retinal nerve fiber layer (RNFL) and choroid. Loss of RNFL in the optic nerve head (ONH) region is a well-known hallmark of glaucoma that leads to irreversible vision loss (Medeiros et al., 2005). Currently, the RNFL thickness profile along a circular scan centered at the ONH and the sectoral average thickness maps are used in clinics to assess the disease progression. Studies have shown regional patterns in glaucomatous RNFL thinning, with most significant changes in the inferior peripapillary region (Leung et al., 2010; Mwanza et al., 2011). Aging has been also associated with RNFL loss (Budenz et al., 2007; Parikh et al., 2007). The effect of glaucoma in the choroid has been more debated, with some studies reporting glaucoma-related thickness changes (Song et al., 2016; Li et al., 2017), and others reporting no changes (Ehrlich et al., 2011; Maul et al., 2011). Recent works on OCT angiography (Lee E. J. et al., 2016; Mammo et al., 2016) suggest vascular impairment in glaucoma, and this motivates simultaneous, localized analysis of the two layers to investigate possible connection in the structural changes due to glaucoma.

In this work, we aim to (i) examine the spatial RNFL thickness patterns by age and by presence of glaucoma by comparing the reference group of older healthy eyes with younger healthy eyes and with age-matched glaucoma eyes, and (ii) study whether there is spatial relationship between age-related changes and glaucoma-related changes.

MATERIALS AND METHODS

Participants and Image Acquisition

Thirty-eight eyes from five young healthy participants, five older healthy participants, and twelve older glaucoma patients were included in the study. The participant demographics are listed in **Table 1**. Before being included in the study, each participant was subject to optic nerve head OCT imaging, dilated stereoscopic examination of the optic nerve, stereo

TABLE 1 | Participant demographics.

Group	N (Subjects)	N (Eyes)	Female/male	Mean age
Young healthy	5	10	3/2	29.8 ± 3.6
Older healthy	5	10	2/3	57.0 ± 4.4
Older glaucoma	12	18	6/6	61.7 ± 7.9

disc photography, intraocular pressure (IOP) measurement, and reproducible Humphrey perimetry at the Eye Care Center at Vancouver General Hospital. Eyes with retinal disease other than primary open-angle glaucoma, uveitis, IOP lower than 10 mmHg or greater than 20 mmHg, or optic neuropathy from causes other than glaucoma were excluded. The mean glaucoma duration at the time of imaging was 3.69 ± 3.80 years. A custom 1,060-nm swept-source OCT system by the Biomedical Optics Research Group at SFU was used for imaging. Each image consisted of 400 B-scans, with 400 A-scans per B-scan and 1,024 pixels per A-scan. The axial voxel resolution was $2.7 \mu\text{m}$, the axial coherence length was $\sim 6 \mu\text{m}$, and the lateral pixel resolution ranged from 11.9 to $14.5 \mu\text{m}$ depending on the eye's axial length. The A-scan rate of 100 kHz resulted in ~ 1.6 s of acquisition time per volume.

Preprocessing, Segmentation, and Layer Thickness Measurement

Images with artifacts, such as large lateral motion artifact or the ONH not being at the center of the field of view were excluded from this analysis. The image underwent axial motion correction by B-scan cross-correlation and 3D bounded-variation smoothing for reducing the effect of speckles and enhancing the visibility of anatomical structures, with no additional normalization. An example of OCT B-scan before and after processing is shown in **Figures 1a,b**. Retinal nerve fiber layer (RNFL) and choroid were segmented automatically by delineating inner limiting membrane (ILM), RNFL-ganglion cell layer boundary, Bruch's membrane (BM), and the choroid-sclera boundary using a 3D graph-cut based algorithm (Li et al., 2006; Lee et al., 2013a,b). All automated segmentation results were checked by a trained rater, and incorrect segmentation was manually corrected using Amira (version 5, FEI). Bruch's membrane opening (BMO) was manually segmented on 80 radial slices extracted from the image volume. An example of the segmentation is shown in **Figures 1c,d**. A best-fit 3D BMO ellipse was computed using principal component analysis and least square fitting. The segmented RNFL and choroid were cropped at 0.25 mm from the BMO ellipse to account for the ambiguity in the retinal layer boundary close to the optic cup (Lee et al., 2014). Recent studies reported on inconsistencies resulting from using the conventional optic cup as a reference due to its ambiguous definition based on 2D projection fundus images, and showed the BMO was a viable alternative reference as a robust anatomical

structure defined in 3D space (Chauhan and Burgoyne, 2013; Gardiner et al., 2014). The layer thickness was measured at each point as the closest 3D Euclidean distance between the posterior and anterior surfaces of the layer. Prior to the fshape registration step, all corresponding surfaces were rigidly aligned by matching the BMO ellipse centroid.

Fshape Registration

The framework of functional shapes (or fshapes) provides a quantitative measure of inter-subject variability in the RNFL and choroid. In this section, we briefly summarize the algorithm which is detailed in Lee et al. (2017). Let the i th subject's RNFL or choroid thickness be represented by (X^i, f^i) , where X is the layer surface (geometry) and f is the surface-indexed function (thickness here) mapped on X . Let a template exemplar for this database be denoted by (X_*, f_*) , consisting of a template surface geometry X_* and an associated surface-indexed signal (thickness here) mapping f_* . Given (X_*, f_*) as the template fshape, and (X^i, f^i) as the i th target fshape to be registered to the template, the fshape framework will estimate a smooth deformation ϕ^i of the template geometry X_* and a residual ζ^i to be added to the template function f_* , such that after transformation of the template fshape with the mapping ϕ^i to the i th target fshape, the geometry of the transformed template matches the geometry of the i th target $X^i \approx \phi^i(X_*)$ and the transformed template function plus residual matches the function of the i th target i.e., $f^i \approx (f_* + \zeta^i) \circ \phi^{i-1}$. Hence, for each of the target fshapes (X^i, f^i) for all $i = 1..N$ a pair (ϕ^i, ζ^i) consisting of a smooth deformation and a function residual are estimated such that:

$$\begin{aligned} (X^i, f^i) &\approx (\phi^i, \zeta^i) \cdot (X_*, f_*) = (\phi^i(X_*), (f_* + \zeta^i) \circ (\phi^i)^{-1}) \\ &= (\tilde{X}^i, \tilde{f}^i) \end{aligned} \quad (1)$$

The function residual added to the template function estimated by fshape matching effectively becomes the representative of the target function in the geometry of the template. By fixing a template geometry and function, a group of target fshapes can be brought into the coordinates of the template such that the residuals representing different target function values are now indexed on the same template geometry. The choice of the template is an important consideration and the mean of the observations in the database is a standard choice. A database fshape mean is hence estimated, to be used as a template, by

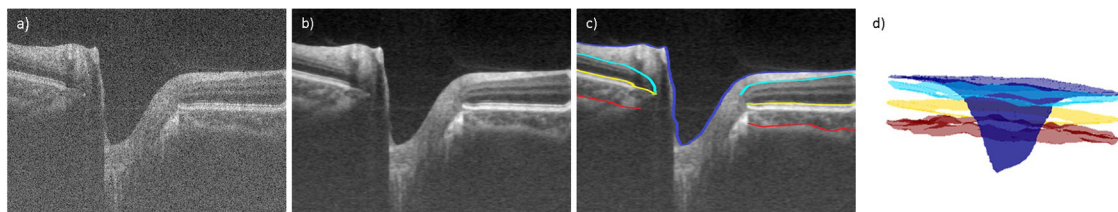


FIGURE 1 | Example OCT B-scan **(a)** without processing, **(b)** after BV smoothing, and **(c)** with the segmentation of inner limiting membrane (purple), posterior boundary of retinal nerve fiber layer (cyan), Bruch's membrane (yellow), Bruch's membrane opening (green), and choroid-sclera boundary (red). 3D surface generation of the retinal layer boundaries is shown in **(d)**.

an adaptive gradient descent algorithm that jointly optimizes a total fshape dissimilarity measure between each target fshape (X^i, f^i) and the transformed template $(\tilde{X}^i, \tilde{f}^i)$ for all i taken together. The process is summarized in **Figure 2**. A similar mean template generation procedure was repeated for the choroid. This maps all the target observations' function values (layer thickness) into a common coordinate system of the mean template where statistical analysis can be applied on the residuals (ζ^i) to compute the point-wise variability of the function values at each point on the template geometry across the database.

Sectorization

In order to compare the fshape analysis with more conventional "intrinsic" analysis where an individual-specific coordinate system is placed on each geometry separately such as using sectoral analysis, the peri-papillary retinal layers were divided into regional sectors (Lee et al., 2014) as shown in **Figure 3**. Unlike standard sectors defined by fixed distances from the center of the optic disk on the enface projection image that do not take into account the individually varying sizes of Bruch's membrane opening (BMO), the sectors in this study were defined in 3D in each eye by the distance from the BMO, which is a more reliable anatomical landmark than the optic disk (Lee et al., 2014). This normalizes the sectors for different retinal tilts and BMO/optic disks sizes. The sectors were first delineated by elliptical annuli at constant distances (0.25, 0.75, 1.25 mm) from the BMO ellipse. These were further divided by superior, nasal, inferior, and temporal sectors, and additionally into superior-nasal (SN), inferior-nasal (IN), inferior-temporal (IT), and superior-temporal (ST) sectors. The first four sectors are 60° wide and the latter are 30° wide. For each sector, the thickness measurements for all points in the sector for that individual eye are taken and averaged to create one scalar number representing the average sectoral thickness value.

Statistical Analysis

Group Analysis

Two analyses were conducted (1) the fshape analysis of computing the mean fshape geometry and function and the residual function for each subject indexed on the common mean template, enabling point-wise comparison across the subjects in the database and (2) sectoral averages within each subject that enable comparison across subjects by the sectors defined over the subject's layer geometry.

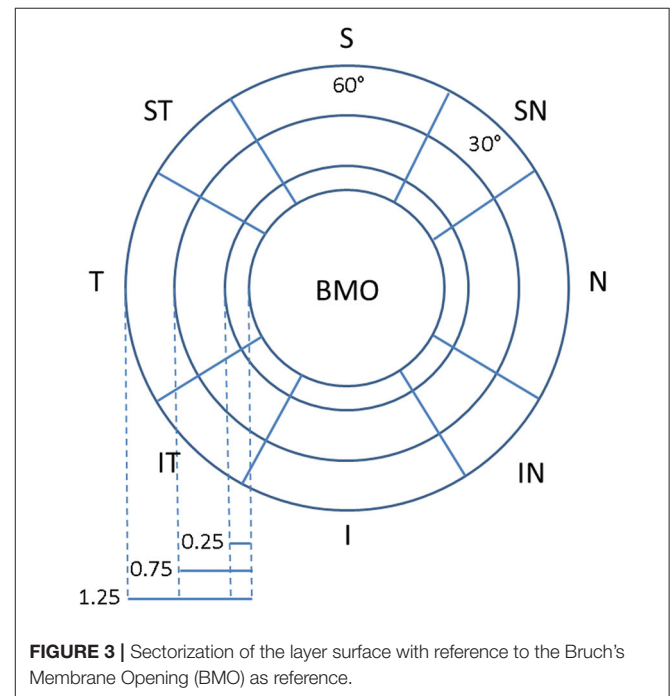


FIGURE 3 | Sectorization of the layer surface with reference to the Bruch's Membrane Opening (BMO) as reference.

Algorithm 1 Computation of the fshape mean template

```

1: Require: An initial template  $(X_o, f_o)$  and  $N$  observations  $(X^i, f^i)$ ,  $1 \leq i \leq N$ 
2: Input: Initial deformation variables  $u = (\phi^o, \zeta^o, \phi^i, \zeta^i)$ ,  $1 \leq i \leq N$ 
3: Begin
4:   repeat
5:     Deformation: mean template  $(X_*, f_*) = (\phi^o, \zeta^o) \cdot (X_o, f_o)$ .
6:     for  $i = 1$  to  $N$  do
7:       Deformation: estimations  $(\tilde{X}^i, \tilde{f}^i) = (\phi^i, \zeta^i) \cdot (X_*, f_*)$ .
8:       Compute energy term  $J_i = g_i + E_i$ , with dissimilarity between estimation
          and observation  $g_i = \|\mu_{(\tilde{X}^i, \tilde{f}^i)} - \mu_{(X^i, f^i)}\|^2$  and deformation energy  $E_i(\phi^i, \zeta^i)$ 
9:       Compute gradient term:  $\nabla_u J_i$ 
10:    end for
11:    Compute deformation energy  $J_o = E_o(\phi^o, \zeta^o)$  and  $\nabla_u J_o$ .
12:    Compute total energy term  $J(u) = J_o + \sum_{i=1}^N J_i$  and gradient term  $\nabla_u J$ 
13:    Update  $u^{new} \leftarrow u + \delta \nabla_u J$ , where  $\delta$  is the step size.
14:    Compute  $J^{new}(u^{new})$ .
15:    If  $J^{new} < J$ ,  $u \leftarrow u^{new}$  and increase  $\delta$ . Else, decrease  $\delta$ .
16:  until Maximum iteration or minimum descent is reached.
17: End
18: Output: Final  $(X_*, f_*)$ ,  $(\tilde{X}^i, \tilde{f}^i)$ , and  $u = (\phi^o, \zeta^o, \phi^i, \zeta^i)$ ,  $1 \leq i \leq N$ .

```

FIGURE 2 | Computation of fshape mean template.

The database consisted of members from three groups: young normal (Group A), older normal (Group B), and older glaucomatous (Group C) individuals. These groups enable analyses of two main questions (1) the effect of age on RNFL and choroid layer thicknesses in healthy young and older individuals (Group A and B comparison) and, (2) the effect of glaucoma between age-matched individuals (Group B and Group C comparison). These two questions were analyzed by point-wise (fshape) and sector-wise group mean thickness difference maps and two-sample *t*-test maps.

Regression Analysis

The effect of age and glaucoma on RNFL and choroidal thicknesses was additionally examined by point-wise and sector-wise linear regression to quantify trend, on Group A and B for the effect of age, and Group B and C for the effect of glaucoma. Each layer thickness (RNFL or choroid) values from Group A and B were fitted to $\text{layer thickness} = a * \text{Age} + b$ to estimate the rate of change (mm/year) in the cross-sectional healthy subjects. To assess change as a function of visual field mean deviation (VFMD), a measure of glaucomatous loss measured in decibels (dB), the layer thickness (RNFL or choroid) values from Group B and C were fitted to $\text{layer thickness} = a * \text{VFMD} + b$ to estimate the rate of change per VFMD (mm/dB).

Point-Wise Visualization of RNFL and Choroid Thickness Pattern

To visualize the measurements across the database highlighting the variability and trends, the fshape measures for each subject on the common template were normalized by using a z-score computed by subtracting the mean of a reference group and dividing by the standard deviation of the measures in the reference group. The z-score was calculated point-wise as $z_k = (x_k - \bar{x}_k^h) / \sigma_k^h$, where x_k is the thickness at k^{th} point for, \bar{x}_k^h is the mean thickness of the reference group at k^{th} point, and σ_k^h is the standard deviation of the reference group at k^{th} point. The reference groups were chosen to be Group A for age comparison and Group B for glaucoma comparison such that the average measures of the young normal (Group A) or older healthy (Group B) subjects are removed to visualize the residual effects of age and glaucoma, respectively. To present a compact visualization, measurement of each subject was unraveled into a column format by subdividing the mean template surface into sectors, and subdividing each sector into smaller sub-sectors, and arranging the z-score values by their sectors in a column format consistent across the database while preserving spatial adjacencies.

RESULTS

This section will present the results of point-wise (using fshape) and sectoral analysis of RNFL and choroid thickness across the three cohorts chosen for this study. All figures are right-eye oriented, with the left side temporal and the right side nasal. Group averages of RNFL thickness are visualized in **Figure 4**. The top row shows the point-wise average computed using fshapes and the bottom row shows the results using sectoral averaging.

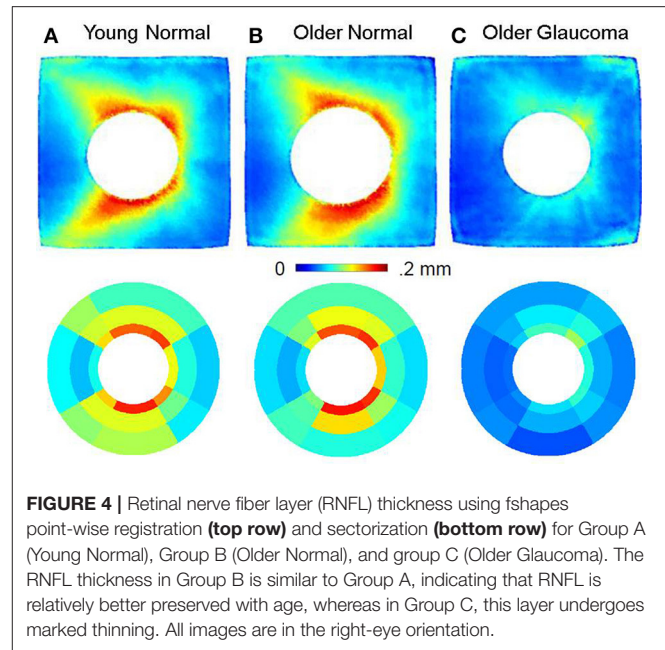
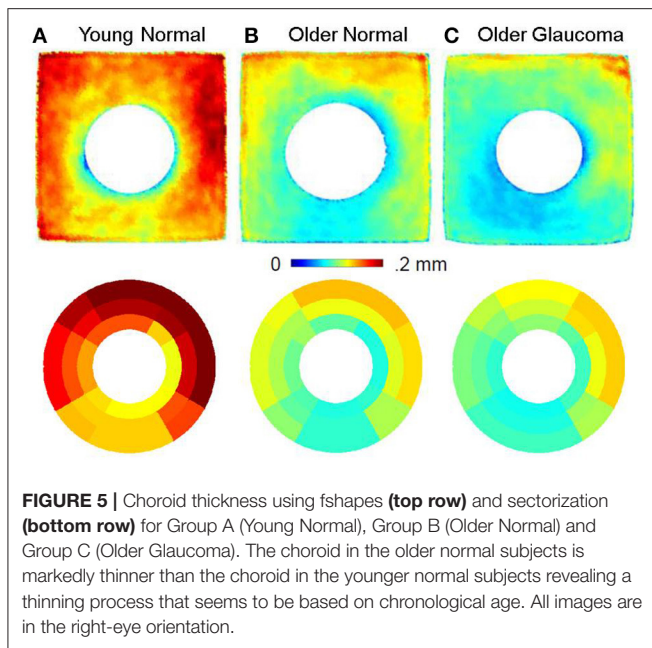


FIGURE 4 | Retinal nerve fiber layer (RNFL) thickness using fshapes point-wise registration (top row) and sectorization (bottom row) for Group A (Young Normal), Group B (Older Normal), and group C (Older Glaucoma). The RNFL thickness in Group B is similar to Group A, indicating that RNFL is relatively better preserved with age, whereas in Group C, this layer undergoes marked thinning. All images are in the right-eye orientation.

The point-wise fshape mean RNFL templates display detailed salient features from multiple, cross-sectional eyes in each group showing the characteristic hourglass pattern of healthy RNFL in both Group A (young normal) and Group B (older normal) whereas Group C (older glaucoma) show marked glaucomatous thinning. There is good correspondence between the fshape mean templates constructed from point-wise registration and sectoral average maps taken from unregistered RNFL thickness averages in each sector in each individual.

The group averages of choroidal thickness by fshape (top row) and sectoral averaging (bottom row) is shown in **Figure 5**. All three groups display thicker choroid in the nasal and superior regions and thinner choroid in the inferior region. The choroid is visibly thinner with age as seen in Group B compared to Group A. As with RNFL thickness, there is overall correspondence between the fshape mean templates and sectoral average maps.

The effect of age and glaucoma in RNFL thickness is shown by difference of group averages in **Figure 6**. The top row shows the effect of age in RNFL thickness in the healthy subjects comparing the mean RNFL thickness in young normal vis-a-vis the older normal individuals. The bottom row shows the effect of glaucoma in RNFL thickness by comparing age-matched older subjects with and without glaucoma. The left panel shows the point- and sector-wise difference of RNFL thickness from Group B by Group A (top, young vs. older) and from Group C by Group B (bottom, healthy vs. glaucoma) at each corresponding points / sectors. The right panel shows point and sector-wise *t*-test results indicating where the group difference is significant. The RNFL thickness is found to not change significantly over age across individuals (top row), whereas the difference due to glaucoma is apparent (bottom row). The loss of RNFL thickness is observed to be higher in regions where normal RNFL thickness is higher and suggests some regional correspondence between the degree of



glaucomatous RNFL loss and the original RNFL thickness. The t -test significance map between the healthy and glaucoma subjects shows the highest statistical significance in the inferior region.

Similar group difference visualization for choroid is shown in **Figure 7**. Compared to RNFL, choroidal thickness is more different by age (top row, Group B – Group A) than by glaucoma (Group C – Group B). Choroidal thickness of Group B is consistently lower than that of Group A across all regions, but in particular in the nasal and superior regions. The locations of statistical significance, as shown in the t -test map, of the age-related group difference is found particularly in the nasal region. The choroidal thickness difference by glaucoma was not as strong as that due to age and although the choroidal thickness in Group B was overall larger than that of group C, the point- and sector-wise t -test show limited statistical significance.

The relationship of RNFL thickness to aging and severity of glaucoma was examined by point- and sector-wise linear regression in **Figure 8**. The estimated slope or the rate of change associated with aging (mm/year) in the cross-sectional healthy subjects was plotted in the top row along with the goodness of fit by r^2 . The estimated slope or the rate of change associated with visual field loss (mm/dB) in the cross-sectional age-matched subjects was plotted in the bottom row along with the goodness of fit by r^2 . Aging did not show consistent, significant trend with RNFL thickness, whereas increasing glaucoma severity (more negative VFMD) was correlated to decreasing RNFL thickness. Similar spatial patterns are observed as seen in the group difference maps shown in **Figure 6**. In the most severely affected regions of superior-temporal and inferior-temporal regions, RNFL thickness change per MD unit decrease exceeded $5 \mu\text{m}$.

The relationship of choroidal thickness to aging and severity of glaucoma was also examined by point- and sector-wise linear regression in **Figure 9**. The estimated slope or the rate of change associated with aging (mm/year) in the cross-sectional healthy

subjects was plotted in the top row along with the goodness of fit by r^2 . The estimated slope or the rate of change associated with visual field loss (mm/dB) in the cross-sectional age-matched subjects was plotted in the bottom row along with the goodness of fit by r^2 . Again, compared to the RNFL thickness, change in the choroidal thickness was associated more with aging than glaucoma. Aging was correlated to globally decreasing choroidal thickness, most significantly in the nasal and inferior regions. In the most severely affected regions, the average choroidal thickness change per year was $\sim 3\text{--}4 \mu\text{m}$.

The point-wise nature of fshape metrics can be utilized by simultaneous visualization of all data points from multiple subjects. **Figure 9** displays the RNFL fshape thickness in the order of VFMD, which measures the glaucomatous functional loss, for the age-matched normal and glaucomatous eyes of Group B and C. In the top panel, each column represents an eye's point-wise RNFL fshape thickness, which approximates the true RNFL thickness as the sum of the RNFL fshape mean template thickness and the residual ($t_i \approx f_* + \zeta_i$) at each point. Horizontally, the eyes are ordered by VFMD, plotted below the thickness plot in grayscale. Vertically, the points are ordered by sectors from Nasal (N) and counter-clockwise to Inferior Nasal (IN). Within each sector, the points are ordered by the distance from BMO center, from the closest to the farthest. This visualization allows one to compare all eyes at each spatial location. There is observed an overall group-wise difference between Healthy older subjects, Early Glaucoma, and Moderate to Severe Glaucoma subjects. Comparing vertically from the top to the bottom, the healthy eyes show the thickness pattern that follows the mean template for Group B in **Figure 4** where the superior and inferior regions are the thickest, and within each region, RNFL is thicker toward BMO and thins farther from BMO.

The bottom panel of **Figure 10** visualizes the f-shape thickness in the top panel using z-scores normalized by the group mean and standard deviation of the RNFL thickness of the young healthy group (Group A) at each point, with increasing magnitude indicating greater deviation from the reference group. The differences between the groups are more apparent in z-score plot, with clear regional characteristics. Regionally, inferior RNFL, and to a lesser degree, inferior-temporal RNFL, are consistently thinner in all glaucoma eyes. In other sectors, the magnitude of z-score is greater for the moderate to severe glaucoma group than early glaucoma. The plot also shows glaucomatous RNFL thickness change is greater nearer BMO by the vertical gradations within individual regions in the glaucomatous eyes.

Figure 11 displays the choroidal fshape thickness in the order of age for the normal eyes to show the thinning of the choroid observed with age. In the top panel, each column represents an eye's point-wise choroidal fshape thickness, which approximates the true choroidal thickness at each point. Horizontally, the eyes are ordered by age, plotted below the thickness plot in grayscale. As in **Figure 10**, the points are vertically ordered by sectors from Nasal (N) and counter-clockwise to Inferior Nasal (IN), and within each sector, the points are ordered by the distance from BMO center, from the closest to the farthest. In the bottom panel, the same data is visualized in z-scores

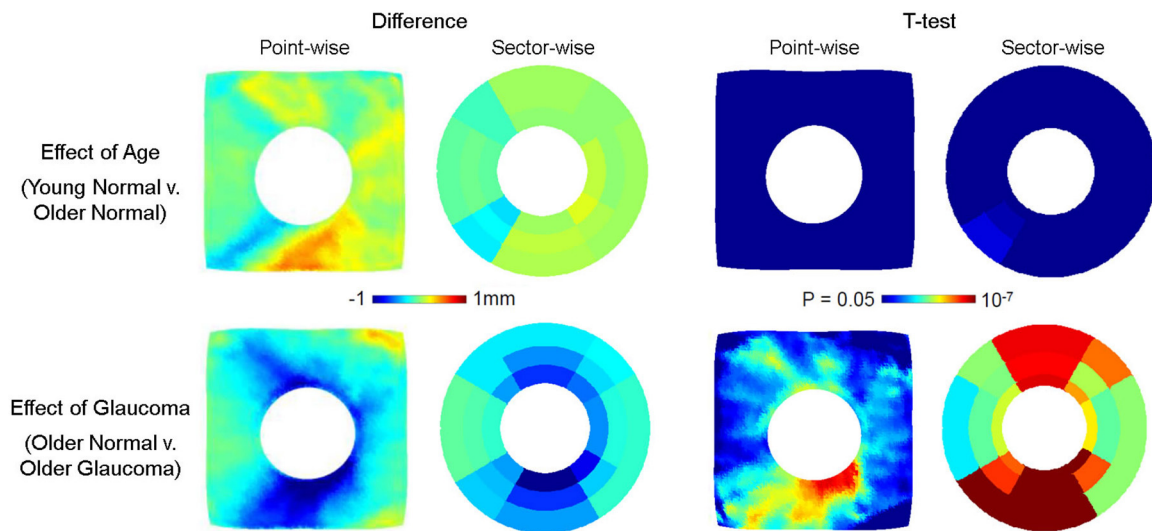


FIGURE 6 | Effect of age (top row) and glaucoma (bottom row) on RNFL thickness. The point- and sector-wise subtraction of mean RNFL thickness between older normal (Group B) and young normal (Group A) is shown in the top left panel, and the p -values from point- and sector-wise t -test is shown in the top right panel. The point- and sector-wise subtraction of RNFL mean between older glaucoma (Group C) and older normal (Group B) is shown in the bottom left panel, and the p -values from point- and sector-wise t -test is shown in the bottom right panel. The RNFL layer is found not to change significantly with age, whereas it changes significantly with glaucoma. The fshapes point-wise comparison shows the pattern of change in greater detail than the sectorization, revealing the localized pattern of glaucomatous RNFL thinning. All images are in the right-eye orientation.

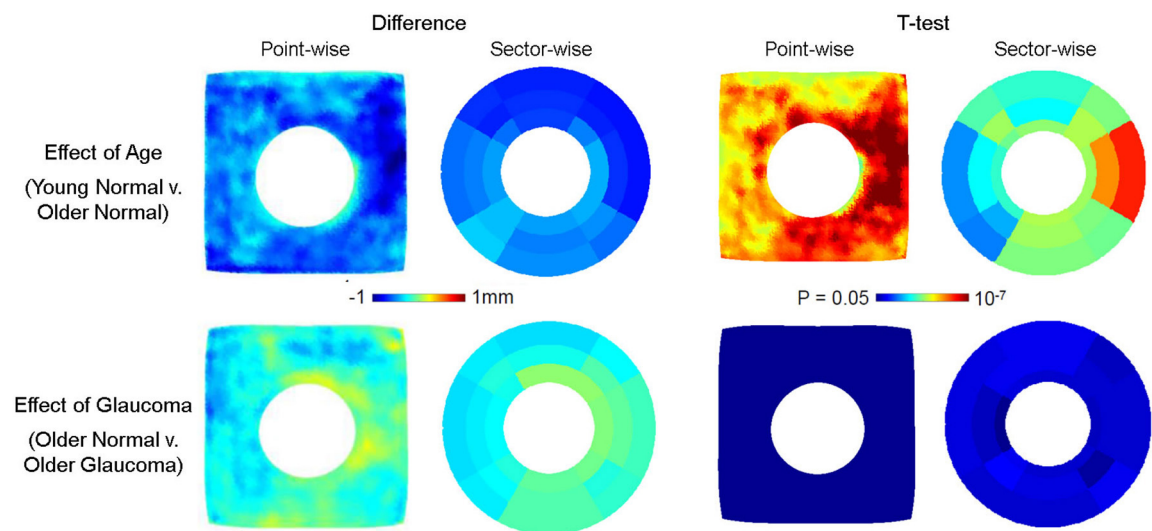


FIGURE 7 | Effect of age (top row) and glaucoma (bottom row) on choroidal thickness. The point- and sector-wise subtraction of mean choroidal thickness across older normal (Group B) and young normal (Group A) is shown in the top left panel, and the p -values from point- and sector-wise t -test is shown in the top right panel. The point- and sector-wise subtraction of mean choroidal thickness between older glaucoma (Group C) and older normal (Group B) is shown in the bottom left panel, and the p -values from point- and sector-wise t -test is shown in the bottom right panel. The choroid is found to thin significantly with age, whereas changes are relatively less with glaucoma. The fshapes point-wise comparison shows the pattern of change in greater detail, revealing the localized pattern of age-related choroidal thinning. All images are in the right-eye orientation.

normalized by the group mean and standard deviation of the choroidal thickness of the young healthy group. As shown in **Figures 5, 7**, there is a marked difference between the young healthy and older healthy eyes, and the choroid appears generally thicker in the superior half than the inferior half. Within

the young eyes, the z -score is generally lower for the older eyes. As seen in **Figure 7**, the nasal region of the older eyes shows the highest magnitudes of z -score, suggesting the age-related choroidal thinning may be the most significant in the region.

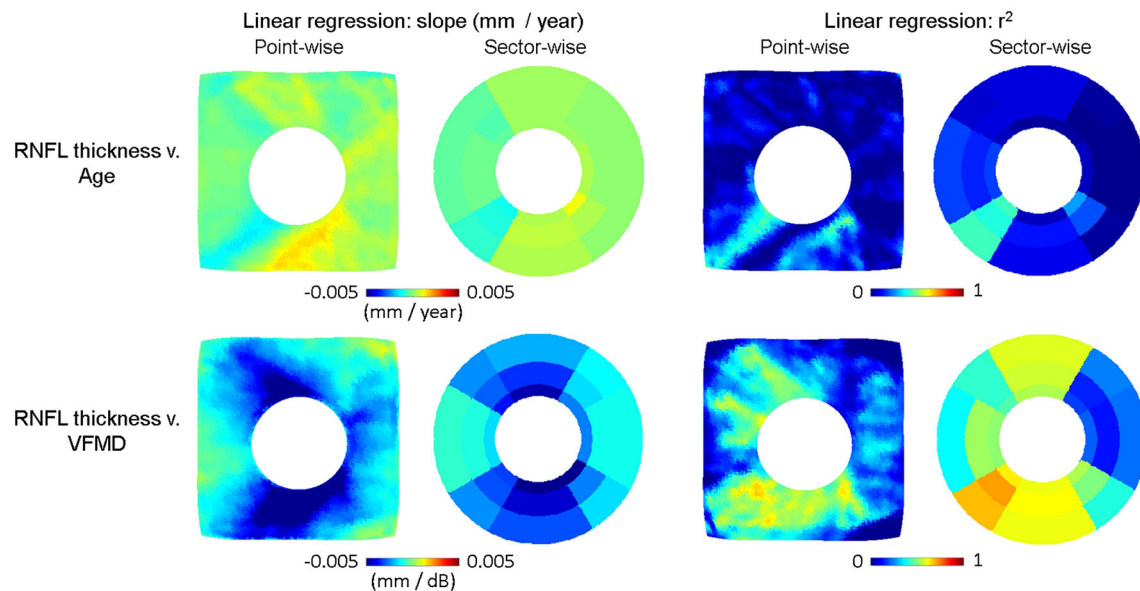


FIGURE 8 | Linear regression plots of RNFL thickness v. age (top row) and v. visual field mean deviation (VFMD). The slope in RNFL thickness = $a \cdot \text{Age} + b$ is plotted point- and sector-wise in the top left panel, along with the goodness of fit (r^2) in the top right panel. The slope of RNFL thickness = $a \cdot \text{VFMD} + b$ is plotted point- and sector-wise in the bottom left panel, along with the goodness of fit (r^2) on the bottom right panel. As with the group difference, RNFL thickness is negatively correlated with VFMD. The point-wise maps reveal that the rate of change has distinct spatial pattern with greater thinning along the regions where RNFL is generally thicker in healthy subjects. All images are in the right-eye orientation.

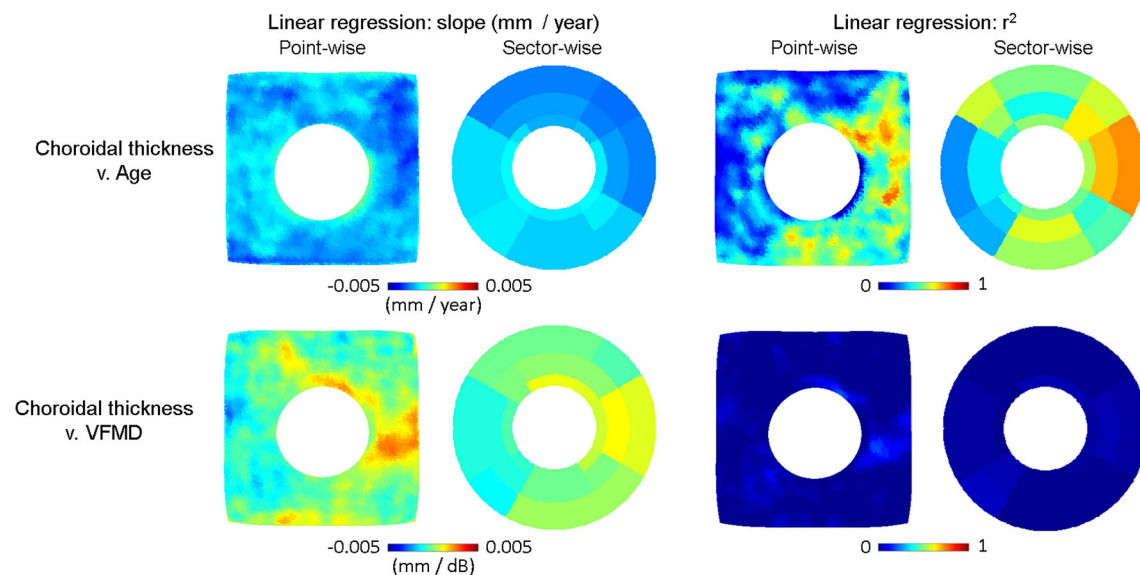
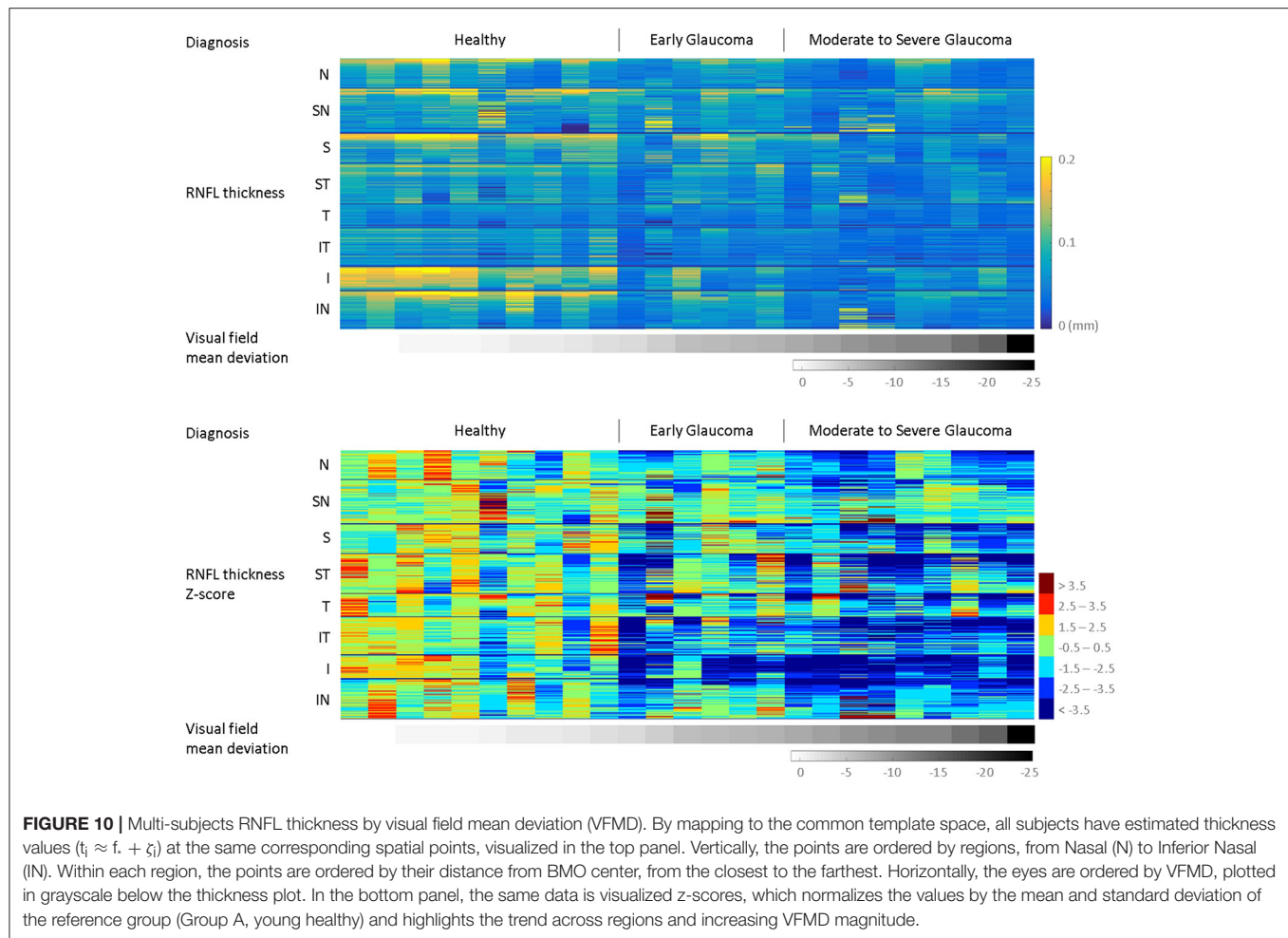


FIGURE 9 | Linear regression plots of choroidal thickness v. age (top row) and v. visual field mean deviation (VFMD). The slope in Choroidal thickness = $a \cdot \text{Age} + b$ is plotted point- and sector-wise in the top left panel, along with the goodness of fit (r^2) in the top right panel. The slope of Choroidal thickness = $a \cdot \text{VFMD} + b$ is plotted point- and sector-wise in the bottom left panel, along with the goodness of fit (r^2) on the bottom right panel. As with the group difference, choroidal thickness is negatively correlated with age. The point-wise maps show the steepest change in the nasal and inferior regions. All images are in the right-eye orientation.

DISCUSSION

The 3D OCT images reveal the structure of the ocular posterior segment in great detail so as to enable visualization and

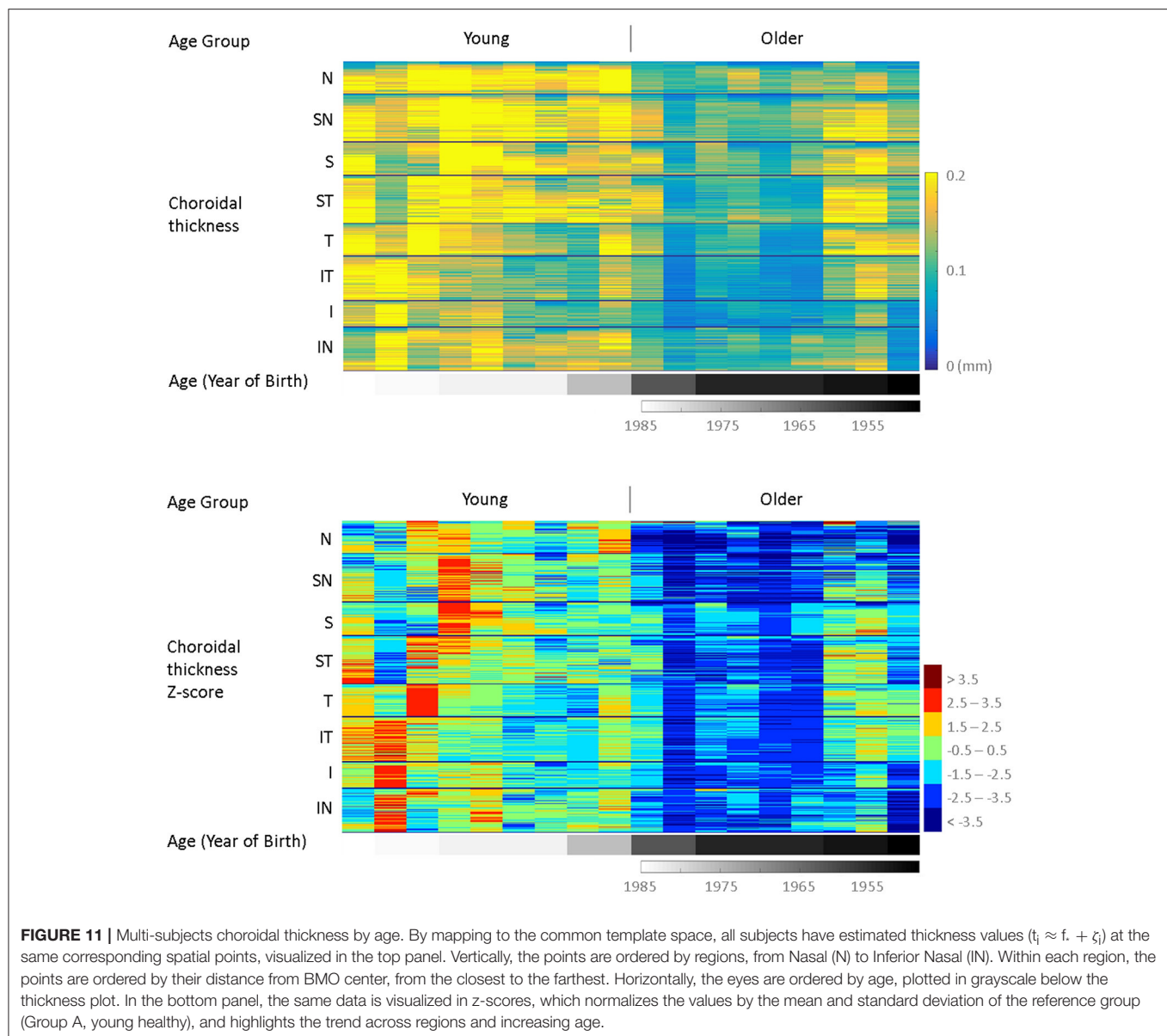
quantification of the retinal layer morphometry. Individual measurements of layer thicknesses can be pooled into population-wide assessments of normative layer thicknesses and any changes that may occur as a function of age and



disease. These allow insights into whether age and disease have a stereotypical pattern of influence in the retina, with common shape features and localizations, as well as variability across individuals and deviation of a particular subject from a reference population. In this paper, we presented the effect of age and glaucoma on retina nerve fiber layer (RNFL) and choroid using our novel f-shapes approach, which enables a point-wise assessment of retinal morphometrics across individuals via a registration approach. The fshape registration estimates a residual function that is added to the template thickness such that the template geometry after transformation matches the subject geometry, and the template thickness plus the residual function after transformation matches the subject thickness. This maps an individual's layer thicknesses onto the common coordinate system of the template geometry via the specific residual function estimated for that individual. At each location on the template surface, subsequent statistical analysis across the database can reveal trends and features that are common across individuals as a function of age and disease. A more conventional approach utilizes sectorization of the layer surface by calculating the average of all thickness measurements within each sector for each individual eye. Assuming that the sectors are

defined with consistent anatomical and spatial correspondence across individuals, within-sector average provides a single scalar summary measure for a given sector that can be statistically analyzed for cross-sectional data taken from the same sector across the individual eyes. However, such approach is limited in spatial sensitivity due to averaging of values in a region. We examined the effect of age and glaucoma in RNFL and choroidal thickness in both of these approaches with four quantitative visualizations: (i) group ranges (Figures 4, 5), (ii) group-wise difference and *t*-test maps (Figures 6, 7), (iii) linear regression with age and visual field mean deviation (VFMD) as predictors (Figures 8, 9), and (iv) multi-subject fshape thickness and z-scores plots (Figures 10, 11). The computation time including automated segmentation and mean template generation was ~40 min on a high-performance GPU cluster.

With age, RNFL showed relatively little difference between the young and older healthy subjects, with regression estimating no strong relationship between age and RNFL thickness. In previous studies using OCT measurements, RNFL thickness has been negatively associated with age (Budenz et al., 2007; Parikh et al., 2007; Bendschneider et al., 2010; Sarunic et al., 2010; More et al., 2011). In this study, the mean age of the young healthy subjects



was thirty, and that of the older healthy subjects was fifty-seven. The age difference between the two groups may be too small for any marked difference, especially with the small sample size in the study. Older age was, however, associated with markedly thinner choroid, and the point-wise registration showed the difference was more significant in the nasal and inferior regions. With the recent advancement in OCT devices enabling the posterior boundary of the choroid to be imaged, age-related choroid thinning has been reported by multiple groups (Manjunath et al., 2010; Maul et al., 2011; Barteselli et al., 2012). Our result suggests overall thinning of peripapillary choroid with age, but with regional differences. That the older healthy subjects had comparable RNFL but thinner choroid compared to the young healthy subjects may indicate the age-related choroidal thinning does not directly and concurrently impact RNFL thickness.

Recent work using speckle-variance OCT (SV-OCT) (Mammo et al., 2016) has shown degradation of RNFL microcapillaries in glaucoma, and it has been suggested the glaucomatous tissue loss may be driven by changes to the microvasculature. Although the choroid is a vascular layer, it mainly supplies the outer layers of the retina that are unaffected in glaucoma, and may therefore be separate from the factors that drive the RNFL tissue and capillary loss in glaucoma.

Glaucoma, as expected, was observed to be associated with significant thinning of RNFL. The pattern of loss visualized in the point-wise maps forms the hourglass crescent-shaped ridge, particularly in the inferior arm, highest toward the middle of the ridge, decreasing further away from the ridge center. The same hourglass-like pattern is observed in young normal subjects. The results of our study are consistent with known patterns of

RNFL loss in glaucoma, but more importantly, it shows that the glaucomatous RNFL loss occur in a specific, uneven pattern that follows the initial RNFL thickness, suggesting that the time of onset and significance of the RNFL loss may be proportional to the initial RNFL thickness in the region. The loss significance was also higher in the region closer to BMO. The fshape RNFL loss map elucidates the results of previous studies that found the highest diagnostic ability of the RNFL loss in the inferior and temporal-inferior sectors (Sehi et al., 2009; Mwanza et al., 2011). Although the choroidal thickness in the older glaucoma subjects seemed to be somewhat less than in the age-matched healthy subjects, it was still statistically comparable, indicating that glaucoma pathology may not have significant, direct impact on the choroid as it does on the RNFL. The role of choroid in glaucoma has been debated, and it is likely complex and multifaceted. Disturbed autoregulation of the choroid has been suggested as part of the disease pathology (Hayreh, 1969; Ulrich et al., 1996). Multiple studies using OCT images (Ehrlich et al., 2011; Maul et al., 2011; Li et al., 2013) consistently reported no changes in the peripapillary choroidal thickness in primary open angle glaucoma (POAG); however, Li et al. (2017) reported thicker temporal peripapillary choroidal area in POAG patients using enhanced depth imaging OCT, and Song et al. (2016) reported global and all 12 clock-hour peripapillary choroidal thickness thinner in OAG patients using swept-source OCT.

Figures 10, 11 presented a large-data visualization with each subject's point-wise thickness values color mapped in a single column, and multiple subjects' data columns displayed concurrently, arranged in the order of visual field mean deviation (VFMD, **Figure 10**) and age (**Figure 11**). This method allows for presentation of all data points from multiple subjects in a way that highlights the trend and discrepancies in the data. Normalizing the data by the mean and standard deviation of the young, healthy group as the reference removes the baseline in the data and further brings out the differences with respect to controls.

The patterns of change shown in the point-wise and sector-wise approaches were overall consistent. The point-wise registration was able to show localized features in higher resolution compared to the sectorization, revealing detailed regional patterns and potentially furthering our understanding of the disease mechanism. This approach may be useful for characterizing the focal localized patterns that are often seen in glaucoma, both in RNFL and in other subsurface structures

such as lamina cribrosa and peripapillary tissues. In this work, we also examined multiple factors (age, glaucoma) in multiple layers (RNFL, choroid) concurrently for a more complete picture in understanding our data. The results showed glaucomatous thinning of RNFL, and age-related thinning of choroid and how the spatial patterns of the tissue loss in the two layers were localized and distinct. Limitations of this work include relatively small sample size, inclusion of fellow eyes, and light beam angle-related uncertainty in retinal layer thickness measurement, although given the relatively small size of the field of view in the images in this study, this effect is likely limited. Our future work will include expanding the analyses presented here to other retinal layers and to the macular region to build a comprehensive presentation of the retinal morphometrics, and incorporating retinal vessels and capillary density measures from SV-OCT as metrics along with layer thickness.

ETHICS STATEMENT

The study followed the tenets of the Declaration of Helsinki, and informed and written consents were obtained from the participants. Ethics review for the study was approved by Simon Fraser University (SFU) and University of British Columbia (UBC).

AUTHOR CONTRIBUTIONS

Designed research: SL, KP, NC, BC, AT, PJM, MVS, and MFB. Analyzed and interpreted data: SL and MFB. Wrote manuscript: SL, MLH, MVS, and MFB. Reviewed and approved manuscript: KP, NC, BC, AT, PJM, MVS, and MFB.

FUNDING

This work was supported by Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Michael Smith Foundation for Health Research, Pacific Alzheimer Research Foundation, and Brain Canada.

ACKNOWLEDGMENTS

This work is based on our conference presentation (Lee S. et al., 2016). It contains extended experimental study with comparison with sectorization and regression analyses.

REFERENCES

- Anthony, B. J., Chen, M., Carass, A., Jedynak, B. M., Al-Louzi, O., Solomon, S., et al. (2016). "Voxel based morphometry in optical coherence tomography: validation & core findings," in *Proceedings of SPIE 9788*. San Diego, CA.
- Barteselli, G., Chhablani, J., El-Emam, S., Wang, H., Chuang, J., Kozak, I., et al. (2012). Choroidal volume variations with age, axial length, and sex in healthy subjects: a three-dimensional analysis. *Ophthalmology* 119, 2572–2578. doi: 10.1016/j.ophtha.2012.06.065
- Bendschneider, D., Tornow, R. P., Horn, F. K., Laemmer, R., Roessler, C. W., Juenemann, A. G., et al. (2010). Retinal nerve fiber layer thickness in normals measured by spectral domain. *O. C. T. J. Glaucoma* 19, 475–482. doi: 10.1097/IJG.0b013e3181c4b0c7
- Budenz, D. L., Anders, D. R., Varma, R., Schuman, J., Cantor, L., Savell, J., et al. (2007). Determinants of normal retinal nerve fiber layer thickness measured by stratus. *O. C. T. J. Ophthalmology* 114, 1046–1052. doi: 10.1016/j.ophtha.2006.08.046
- Charlier, B., Charon, N., and Trouvé, A. (2015). The fshape framework for the variability analysis of functional shapes. *Found. Comput. Math.* 17:287. doi: 10.1007/s10208-015-9288-2
- Chauhan, B. C., and Burgoyne, C. F. (2013). From clinical examination of the optic disc to clinical assessment of the optic nerve head: a paradigm change. *Am. J. Ophthalmol.* 156, 218–227. doi: 10.1016/j.ajo.2013.04.016

- Chen, M., Lang, A., Ying, H. S., Calabresi, P. A., Prince, J. L., and Carass, A. (2014). Analysis of macular OCT images using deformable registration. *Biomed. Opt. Express* 5, 2196–2214. doi: 10.1364/BOE.5.002196
- Ehrlich, J. R., Peterson, J., Parltitsis, G., Kay, K. Y., Kiss, S., and Radcliffe, N. M. (2011). Peripapillary choroidal thickness in glaucoma measured with optical coherence tomography. *Exp. Eye Res.* 92, 189–194. doi: 10.1016/j.exer.2011.01.002
- Gardiner, S. K., Ren, R., Yang, H., Fortune, H., Burgoyne, C. F., and Demirel, S. (2014). A method to estimate the amount of neuroretinal rim tissue in glaucoma: comparison with current methods for measuring rim area. *Am. J. Ophthalmol.* 157, 540–549. doi: 10.1016/j.ajo.2013.11.007
- Gibson, E., Young, M., Sarunic, M. V., and Beg, M. F. (2010). Optic nerve head registration via hemispherical surface and volume registration. *IEEE Trans. Biomed. Eng.* 57, 2592–2595. doi: 10.1109/TBME.2010.2060337
- Hayreh, S. S. (1969). Blood supply of the optic nerve head and its role in optic atrophy, glaucoma, and oedema of the optic disc. *Br. J. Ophthalmol.* 53, 721–748. doi: 10.1136/bjo.53.11.721
- Jørgensen, T. M., Thomadsen, J., Christensen, U., Soliman, W., and Sander, B. (2007). Enhancing the signal-to-noise ratio in ophthalmic optical coherence tomography by image registration—method and clinical examples. *J. Biomed. Opt.* 12:041208. doi: 10.1117/1.2772879
- Lee, E. J., Lee, K. M., Lee, S. H., and Kim, T. W. (2016). OCT angiography of the peripapillary retina in primary open-angle glaucoma. *Invest. Ophthalmol. Vis. Sci.* 57, 6265–6270. doi: 10.1167/iov.16-20287
- Lee, S., Beg, M. F., and Sarunic, M. V. (2013a). “Segmentation of the macular choroid in OCT images acquired at 830nm and 1060nm,” in *Proceedings of SPIE 8802*. Munich. doi: 10.1117/12.2032603
- Lee, S., Fallah, N., Forooghian, F., Ko, A., Pakzad-Vaezi, K., Merkur, A. B., et al. (2013b). Comparative analysis of repeatability of manual and automated choroidal thickness measurements in nonneovascular age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* 54, 2864–2871. doi: 10.1167/iov.12-11521
- Lee, S., Charon, N., Charlier, B., Popuri, K., Lebed, E., Sarunic, M. V., et al. (2017). Atlas-based shape analysis and classification of retinal optical coherence tomography images using the functional shape (fshape) framework. *Med. Image Anal.* 35, 570–581. doi: 10.1016/j.media.2016.08.012
- Lee, S., Han, S. X., Young, M., Beg, M. F., Sarunic, M. V., and Mackenzie, P. J. (2014). Optic nerve head and peripapillary morphometrics in myopic glaucoma. *Invest. Ophthalmol. Vis. Sci.* 55, 4378–4393. doi: 10.1167/iov.14-14227
- Lee, S., Lebed, E., Sarunic, M. V., and Beg, M. F. (2015). Exact surface reistration of retinal surfaces from 3D optical coherence tomography images. *IEEE Trans. Biomed. Eng.* 62, 609–617. doi: 10.1109/TBME.2014.2361778
- Lee, S., Popuri, K., Matsubara, J. A., Eadie, B., Merkur, A., Mackenzie, P., et al. (2016). Novel fshape mapping of peripapillary RNFL and choroidal thickness for group-wise comparison of glaucoma and healthy aging. *Invest. Ophthalmol. Vis. Sci.* 57, 3389.
- Leung, C. K. (2014). Diagnosing glaucoma progression with optical coherence tomography. *Curr. Opin. Ophthalmol.* 25, 104–111. doi: 10.1097/ICU.0000000000000024
- Leung, C. K. S., Choi, N., Weinreb, R. N., Liu, S., Ye, C., Liu, L., et al. (2010). Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: pattern of RNFL defects in glaucoma. *Ophthalmology* 117, 2337–2344. doi: 10.1016/j.ophtha.2010.04.002
- Li, K., Wu, X., Chen, D. Z., and Sonka, M. (2006). Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 119–134. doi: 10.1109/TPAMI.2006.19
- Li, L., Bian, A., Zhou, Q., and Mao, J. (2013). Peripapillary choroidal thickness in both eyes of glaucoma patients with unilateral visual field loss. *Am. J. Ophthalmol.* 156, 1277–1284. doi: 10.1016/j.ajo.2013.07.011
- Li, M., Yan, X., Song, Y., Guo, J., and Zhang, H. (2017). Choroidal area assessment in various fundus sectors of patients at different stages of primary open-angle glaucoma by using enhanced depth imaging optical coherence tomography. *Medicine* 96:e6293. doi: 10.1097/MD.00000000000006293
- Mammo, Z., Heisler, M., Balaratnasingam, C., Lee, S., Yu, D.-Y., Mackenzie, P., et al. (2016). Quantitative optical coherence tomography angiography of radial peripapillary capillaries in glaucoma, glaucoma suspect, and normal eyes. *Am. J. Ophthalmol.* 170, 41–49. doi: 10.1016/j.ajo.2016.07.015
- Manjunath, V., Taha, M., Fujimoto, J. G., and Duker, J. S. (2010). Choroidal thickness in normal eyes measured using cirrus HD optical coherence tomography. *Am. J. Ophthalmol.* 150, 325–329. doi: 10.1016/j.ajo.2010.04.018
- Maul, E. A., Friedman, D. S., Chang, D. S., Boland, M. V., Ramulu, P. Y., Jampel, H. D., et al. (2011). Choroidal thickness measured by spectral domain optical coherence tomography: factors affecting thickness in glaucoma patients. *Ophthalmology* 118, 1571–1579. doi: 10.1016/j.ophtha.2011.01.016
- Medeiros, F. A., Zangwill, L. M., Bowd, C., Vessani, R. M., Susanna, R., and Weinreb, R. N. (2005). Evaluation of retinal nerve fiber layer, optic nerve head, and macular thickness measurements for glaucoma detection using optical coherence tomography. *Am. J. Ophthalmol.* 139, 44–55. doi: 10.1016/j.ajo.2004.08.069
- More, H. L., Chen, J., Gibson, E., Donelan, J. M., and Beg, M. F. (2011). A semi-automated method for identifying and measuring myelinated nerve fibers in scanning electron microscope images. *J. Neurosci. Methods* 201, 149–158. doi: 10.1016/j.jneumeth.2011.07.026
- Mwanza, J.-C., Oakley, J. D., Budenz, D. L., and Anderson, D. R. (2011). Ability of cirrus HD-OCT optic nerve head parameters to discriminate normal from glaucomatous eyes. *Ophthalmology* 118, 241–248. doi: 10.1016/j.ophtha.2010.06.036
- Niemeijer, M., Garvin, M. K., Lee, K., van Ginneken, B., Abramoff, M. D., and Sonka, M. (2009). “Registration of 3D spectral OCT volumes using 3D SIFT feature point matching,” in *Proceedings of SPIE*, vol. 7259. Orlando, FL. doi: 10.1117/12.811906
- Parikh, R. S., Parikh, S. R., Sekhar, G. C., Prabakaran, S., Babu, J. G., and Thomas, R. (2007). Normal age-related decay of retinal nerve fiber layer thickness. *Ophthalmology* 114, 921–926. doi: 10.1016/j.ophtha.2007.01.023
- Sarunic, M. V., Yazdanpanah, A., Gibson, E., Xu, J., Bai, Y., Lee, S., et al. (2010). Longitudinal study of retinal degeneration in a rat using spectral domain optical coherence tomography. *Opt. Exp.* 18, 23435–23441. doi: 10.1364/OE.18.023435
- Sehi, M., Grewal, D. S., Sheets, C. W., and Greenfield, D. S. (2009). Diagnostic ability of fourier-domain vs time-domain optical coherence tomography for glaucoma detection. *Am. J. Ophthalmol.* 148, 597–605. doi: 10.1016/j.ajo.2009.05.030
- Song, Y. J., Kim, Y. K., Jeong, J. W., and Park, K. H. (2016). Assessment of open-angle glaucoma peripapillary and macular choroidal thickness using swept-source optical coherence tomography (SS-OCT). *PLoS ONE* 11:e0157333. doi: 10.1371/journal.pone.0157333
- Ulrich, A., Ulrich, C., Barth, T., and Ulrich, W. (1996). Detection of disturbed autoregulation of the peripapillary choroid in primary open angle glaucoma. *Ophthalmic Surg. Lasers* 27, 746–757.
- Young, M., Lebed, E., Jian, Y., Mackenzie, P. J., Beg, M. F., and Sarunic, M. V. (2011). Real-time high-speed volumetric imaging using compressive sampling optical coherence tomography. *Biomed. Opt. Express* 2, 2690–2697. doi: 10.1364/BOE.2.002690

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Lee, Heisler, Popuri, Charon, Charlier, Trouvé, Mackenzie, Sarunic and Beg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bayesian Tractography Using Geometric Shape Priors

Xiaoming Dong^{1*}, Zhengwu Zhang^{2,3} and Anuj Srivastava¹

¹ Department of Statistics, Florida State University, Tallahassee, FL, United States, ² The Statistical and Applied Mathematical Sciences Institute (SAMSI), Research Triangle Park, Durham, NC, United States, ³ Department of Statistical Science, Duke University, Durham, NC, United States

The problem of estimating neuronal fiber tracts connecting different brain regions is important for various types of brain studies, including understanding brain functionality and diagnosing cognitive impairments. The popular techniques for tractography are mostly sequential—tracts are grown sequentially following principal directions of local water diffusion profiles. Despite several advancements on this basic idea, the solutions easily get stuck in local solutions, and can't incorporate global shape information. We present a global approach where fiber tracts between regions of interest are initialized and updated via deformations based on gradients of a posterior energy. This energy has contributions from diffusion data, global shape models, and roughness penalty. The resulting tracts are relatively immune to issues such as tensor noise and fiber crossings, and achieve more interpretable tractography results. We demonstrate this framework using both simulated and real dMRI and HARDI data.

OPEN ACCESS

Edited by:

Xiaoying Tang,
SYSU-CMU Joint Institute of
Engineering, China

Reviewed by:

Suyash P. Awate,
Indian Institute of Technology Bombay,
India
Chuyang Ye,
Institute of Automation (CAS), China

*Correspondence:

Xiaoming Dong
x.dong@stat.fsu.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 10 March 2017

Accepted: 14 August 2017

Published: 07 September 2017

Citation:

Dong X, Zhang Z and Srivastava A
(2017) Bayesian Tractography Using
Geometric Shape Priors.
Front. Neurosci. 11:483.
doi: 10.3389/fnins.2017.00483

Keywords: tractography, geometric shape analysis, Bayesian estimation, dMRI fiber tracts, active contours

1. INTRODUCTION

This paper considers an important problem of estimating major white matter fiber tracts in human brain using diffusion magnetic resonance imaging (dMRI) images (Mori et al., 2005). The construction of fiber tracts connecting different brain regions is an important first step toward studying brain connectomics and its implications in assessment of brain functionality, including cognitive abilities and general health. Spurred by experimental development of large databases involving human subjects, with samples across different demographic groups, there is an emerging interest in representing and quantifying brain connectivity patterns. Therefore, efficient and reliable fiber tracking algorithms are urgently needed. However, the problem of estimating fiber tracts using dMRI data is far from being solved (Maier-Hein et al., 2016). The current solutions have many limitations, including inefficiency and susceptibility to noisy, corrupt, and low-quality data. The data mostly comes from pre-processed dMRI images, providing at each voxel a measure of diffusivity of water molecule at that location. The representation of this diffusivity is generally a 3×3 symmetric, positive definite matrix (SPDM), also called a *tensor*. In situations where higher resolution data are available, one constructs high angular resolution diffusion imaging (HARDI) data; at each spatial location the orientation diffusion function (ODF, a function on a unit sphere S^2) is estimated (Descoteaux, 2015). Given these local diffusivity measures, one seeks to form fiber tracts, or their collections in the form of fiber bundles, between regions of interest (ROIs), and to further develops structural networks (Cheng et al., 2012; de Reus and van den Heuvel, 2013; Fornito et al., 2013; Durante and Dunson, 2017). This paper focuses on estimation of fiber tracts, also termed *tractography*, using dMRI and HARDI data. For any two regions (voxels) in a brain coordinate system, the goal is to estimate a collection of curves that follow an optimal pattern of fluid flow connecting these locations, while conforming to anatomical reasonings and interpretations.

Due to the importance of tract-based connectivity in brain connectomic analysis, there have been a number of solutions developed for estimating fiber tracts. They can be loosely grouped into two categories: local and global methods. Local methods construct fiber curves sequentially based on the estimated local diffusion directions. Depending on the mechanism for specifying a local propagation direction, one can further classify the local methods into deterministic methods (Mori et al., 1999; Basser et al., 2000) or probabilistic methods (Hagmann et al., 2003). While the deterministic methods mainly follow the local principal directions to grow fiber curves, the probabilistic methods propose a propagation direction from voxelwise probability distribution, e.g., orientation distribution function (ODF), for growing fibers. The first successful deterministic tractography algorithm was dubbed FACT (fiber assignment by continuous tracking), which has been widely studied in the literature (Mori and van Zijl, 2002). But the limitations of FACT and similar methods are obvious. They include sensitivity to initialization, the susceptibility of principal direction estimation to local noise, and lack of connectivity information between regions of the brain. These limitations drive people to use the probabilistic algorithms. One advantage of the probabilistic methods is that they are based on the full, albeit local, distribution of fiber directions, rather than just the principal direction. They can output a connectivity index measure, e.g., the number of fiber curves, between any two regions of interest, indicating the probability with which the regions are connected to one another. However, this creates problems when the local diffusion directions are not well estimated or are overly smooth. On the other hand, the global methods try to reconstruct fiber curves simultaneously by optimizing the configuration that best matches the given data. Finding the fiber curves that best matches the given data is a hard inverse problem. Current solutions are to translate this inverse problem into a forward problem using a Bayesian approach. For example, Reisert et al. (2011) used a Metropolis Hastings sampler to propose small line segments to fit the given dMRI data and use them to further generate long fiber curves. The global methods provide a better stability with respect to the noise and imaging artifacts. However, there are some issues with the current global methods also. The Bayesian methods typically have high computational cost and require huge memory space, to compute and store a whole ensemble of solutions. Also, in an optimization setting, it is difficult to avoid local solutions since no additional structure is imposed on the optimization.

We can summarize the limitations of current methods as follows: (a) The local methods are essentially sequential—they start fibers from one end and grow them over time. This one-boundary solution is not natural for tractography, which is actually a two-boundary problem. (b) The local tractography algorithms are highly susceptible to fiber crossing, noise and imaging artifacts. Incorrect recording or noisy observations of tensors can send algorithms in wrong directions and it is difficult to recover from such misdirections. (c) The global tractography algorithms achieve better stability with respect to noise, but they are very computationally expensive. (d) Both local and global methods tend to produce a large proportion of false positive fibers because of the noise and ambiguity at fiber crossings. **Figure 1**

shows some examples of limitations of a local streamline method, where the blue lines denote ground truth, the red and green lines are tractography results from the classic FACT method. The left panel shows the challenge of fiber crossing, where the sequential approach fails to reach the target region. The right panel shows the effect of having a patch of noisy data in the middle. The fibers from either regions run into this noisy patch and fail to reach the other end. Additional examples of the challenges faced by streamline methods on the real data, are shown later in the experimental results section. A global approach used for estimating fiber tracts, or curves in general image data, is called *active contours*, where one evolves a curve in order to minimize an energy functional (Pichon et al., 2005; Lankton et al., 2008; Melonakos et al., 2008; Eckstein et al., 2009; Mohan et al., 2009; Zach et al., 2009; Li and Hu, 2013). Other global techniques (Faugeras et al., 2004), including a variation of Kalman Filter (Cheng et al., 2015), have also been applied to this problem.

In this paper, we propose a new approach that is essentially a global method but using additional geometry information for ensuring optimal solutions. The proposed method is fast and easy to implement, and robust to the noise in the data. Most importantly, it can incorporate the prior knowledge from anatomical structure and brain connectomics. Rather than growing fiber tracts sequentially, our idea is to initialize fiber tracts between regions of interest as Euclidean curves and then *deform* them iteratively using gradients of a posterior energy. This approach, termed *Bayesian Active Contours* (Joshi and Srivastava, 2009; Bryner et al., 2013), estimates fiber tracts under an energy function that has contributions from three sources: the given data or the likelihood term, the prior knowledge on the geometric shapes of fibers connecting these ROIs, and a roughness penalty. The algorithm uses the gradient of this posterior energy to iteratively update curves into high probability and highly interpretable fiber tracts. The prior on the geometric shapes relies on developing statistical shape models of fiber curves between ROIs, using atlas data, and evaluating expressions for gradient of resulting shape model energy with respect to the shape variable. We use advances in elastic shape analysis of Euclidean curves to develop efficient statistical models for fiber bundles using training (or atlas) data. The training data can be generated using existing local or global tractography algorithms, or can use manual inputs. These models form prior information for future tractography and, in conjunction with diffusion data likelihood, they provide tract estimation results.

In contrast to the probabilistic tractography method (Behrens et al., 2003, 2007), the proposed Bayesian method is a global one. We start with an initial fiber connecting two pre-specified regions and update it under an energy function. The final fiber can best explain the diffusion data under the constraints of prior shape distribution and desired smoothness. Previously, there are some Bayesian tractography methods proposed in the literature (Friman et al., 2006; Cook et al., 2008; Yap et al., 2011). These methods are different from the proposed one: in our method, we assign a prior on the fiber shape space, while in (Friman et al., 2006; Cook et al., 2008; Yap et al., 2011), the prior is imposed on local fiber orientation distribution. Probably, the most similar work to ours is (Christiaens et al., 2014), where an



FIGURE 1 | Two examples of the classic streamline method does not work. The blue lines are ground truth fibers. The red and green lines are the tractography results from the FACT method. Starting from area A, FACT failed to reconstruct the fiber tracts that connect A and B.

atlas-guided global tractography is introduced with a prior on the local tract distribution. However, our work is different in two aspects: Firstly, we have a different energy function. We introduce a novel data term and a smoothness term separately to measure alignment between fibers and diffusion data, and the smoothness of fiber tracts. Secondly, we have a different prior. We incorporate the prior information of fiber shape from the atlas space while (Christiaens et al., 2014) obtains the prior information of local tract distribution from the atlas space.

The rest of this paper is organized as follows. We describe the three components of the posterior energy—data likelihood, shape prior and roughness penalty—and their gradients in Section 2. The resulting tractography algorithm is laid out in Section 3, and experimental results using both simulated and real data, the extension to HARDI data are presented in Section 4. We close the paper with a short discussion in Section 5.

2. MATHEMATICAL FRAMEWORK FOR BAYESIAN TRACTOGRAPHY

Although the framework can be easily generalized to 3D data, we will restrict to 2D data in this paper for simplicity. The theory is general enough to be applicable to 3D data directly.

First, we develop a mathematical framework for estimation of fiber tracts using tensor data and prior shape models. Let \mathcal{P} be the set of 2×2 symmetric, positive definite matrices (or tensors). For the domain, $D = [0, 1]^2$, let $M : D \rightarrow \mathcal{P}$ denote a continuous vector field of tensors defined on this domain. Let $\beta : [0, 1] \rightarrow D$ be an absolutely continuous curve contained in this domain, and let \mathcal{B} be the set of all such curves. Our goal is to find a β with certain boundary constraints that optimizes a chosen objective function that comes from a Bayesian formulation. Thus, we pose the problem of tractography as a MAP estimation. In this formulation we seek parameterized curve $\hat{\beta}$ that minimizes an energy functional according to: $\hat{\beta} = \operatorname{argmin}_{\beta \in \mathcal{B}} E_{\text{total}}(\beta)$, where

$$E_{\text{total}}(\beta) = \lambda_1 E_{\text{data}}(\beta) + \lambda_2 E_{\text{prior}}(\beta) + \lambda_3 E_{\text{smooth}}(\beta). \quad (1)$$

This total energy functional has contributions from three different criteria that are weighted by the coefficients $\lambda_1, \lambda_2, \lambda_3 > 0$. The data energy E_{data} is defined solely from the diffusion data in the image, E_{prior} is the prior shape energy defined from a statistical model on shapes of the fiber β , and the smoothing energy E_{smooth} is a penalty that ensures a certain amount of smoothness in the estimated fiber. In order to minimize E_{total} we use a gradient descent procedure that updates the curve according to $\beta \mapsto \beta - \delta \nabla_{\beta} E$, where

$$\nabla_{\beta} E = \lambda_1 \nabla E_{\text{data}}(\beta) + \lambda_2 \nabla E_{\text{prior}}(\beta) + \lambda_3 \nabla E_{\text{smooth}}(\beta). \quad (2)$$

That is, we search for a local minimization of Equation (1) via gradient descent. The weights λ_i will certainly affect curve evolution, i.e., a large penalty on the smoothness term favors shorter fibers and so on. Through trial and error, one can adjust the λ 's depending on the data and problem context. In the next three sections, we summarize the formulation of each of the three energy terms.

2.1. Data-Likelihood Term

The data term is designed to quantify the agreement of the fiber directions with the diffusion tensor at that location. Let M be a given tensor field and β be a curve lying in the domain D , as shown in the left panel of **Figure 2**. The data energy term is then given by:

$$E_{\text{data}}[\beta] = \int_0^1 n_{\beta}(t)^T M_{\beta(t)}^{-1} n_{\beta}(t) dt, \text{ where } n_{\beta}(t) = \frac{\dot{\beta}(t)}{|\dot{\beta}(t)|}. \quad (3)$$

Here $n_{\beta}(t)$ denotes the unit vector tangent to β at $\beta(t)$ and $M_{\beta(t)}$ is the tensor at location $\beta(t) \in D$. The integrand is lower at the locations where the fiber tract is aligned with the tensor field and vice-versa.

We motivate the choice of this expression by focusing on some Riemannian frameworks used in tractography:

- **Maximal Curves Matching the Given Tensor Field:** One generally wants to find curves such that their velocity vectors

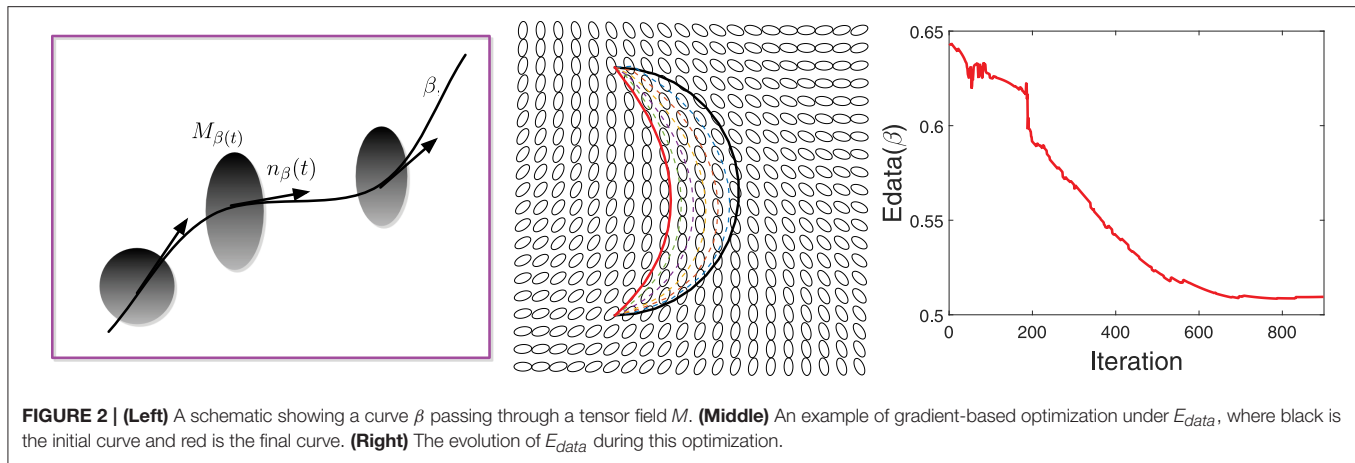


FIGURE 2 | (Left) A schematic showing a curve β passing through a tensor field M . **(Middle)** An example of gradient-based optimization under E_{data} , where black is the initial curve and red is the final curve. **(Right)** The evolution of E_{data} during this optimization.

maximally match the given diffusion tensors. Therefore, one may consider maximizing the term:

$$L_M[\beta] = \int_0^1 \sqrt{\langle \dot{\beta}(t)^T M_{\beta(t)} \dot{\beta}(t) \rangle} dt = \int_0^1 |\dot{\beta}(t)|_{M_{\beta(t)}} dt.$$

This quantity is nothing but the length of a curve β in D under a Riemannian metric defined by the tensor field M . The maximizers of L_M are the longest paths between given points in D . However, the problem with this is that there is no upper bound on the length of the curve, and one can place arbitrarily long curves in D irrespective of M .

- **Geodesics Under Inverse Tensor Field:** A better idea is to use the inverse of the given tensor field at each point and then construct geodesic paths under that Riemannian metric (O'Donnell et al., 2002; Duncan et al., 2004; Melonakos, 2009), according to:

$$\begin{aligned} \beta^* &= \operatorname{argmin}_{\beta} \left(\int_0^1 \sqrt{\langle \dot{\beta}(t)^T M_{\beta(t)}^{-1} \dot{\beta}(t) \rangle} dt \right. \\ &= \left. \int_0^1 |\dot{\beta}(t)|_{M_{\beta(t)}^{-1}} dt \right). \end{aligned}$$

One can solve the optimization problem by minimizing an energy, without the square-root in the integrand, as follows:

$$\beta^* = \operatorname{argmin}_{\beta} \left(\int_0^1 \dot{\beta}(t)^T M_{\beta(t)}^{-1} \dot{\beta}(t) dt \right).$$

This way one gets shortest curves such that their velocities agree with the dominant directions of the original tensor field. This formulation also agrees with a probabilistic approach where one uses the tensor field to define a Gaussian distribution at each point (Lenglet et al., 2004), and seeks maximum likelihood estimates. Although this method favors fiber directions similar to the dominant eigen vectors of the given tensor field, it additionally penalizes the lengths of the such fibers. Similar to the previous bullet, it may be possible to find shorter paths that do not agree with the tensor field. Some other papers (Fuster et al., 2014). Hao et al. (2014)

have expressed this exact issue in different terms, citing the inability of this method to handle high curvature regions. They proposed a solution based on modifying the Riemannian metric by a curvature-based scalar field and then constructing geodesic paths (Hao et al., 2014). The real issue in these ideas is that there is no independent way to control the lengths of estimated fibers.

- **Scale-Invariant Optimal Paths:** We take a different approach where the length of the fibers is separated from the agreement of fiber directions with the given tensor directions. We weight these two quantities differently and are able to better control the length of the fibers. For the domain D , and a given tensor field $M : D \rightarrow \mathcal{P}$, we define an energy term given by

$$E_{data}[\beta] = \int_0^1 n_{\beta}(t)^T M_{\beta(t)}^{-1} n_{\beta}(t) dt, \quad (4)$$

where $n_{\beta}(t) = \dot{\beta}(t)/|\dot{\beta}(t)|$. Note that if we scale the speed of traversal along β by a constant, the energy function remains unchanged. In other words, the integrand only depends on the agreement of the direction $n_{\beta}(t)$ with the dominant eigen vectors of $M_{\beta(t)}$, and not on the speed of traversal at $\beta(t)$. However, this energy function is not invariant to a re-parameterization of β . Let $\gamma : [0, 1] \rightarrow [0, 1]$ be a positive diffeomorphism, the $\beta \circ \gamma$ represents a re-parameterization of β . It can be seen that, in general, $E_{data}[\beta] \neq E_{data}[\beta \circ \gamma]$. If that invariance is desired, one can achieve it by changing the measure of integration from dt to $|\dot{\beta}(t)| dt$ in Equation (4).

The next step is to derive the gradient of E_{data} with respect to β for use in gradient-based optimization. To specify the gradient of E_{data} , we need some additional notation. Note that for any location $x = (x_1, x_2) \in D$, the gradient of $M : D \rightarrow \mathcal{P}$ has two components, $\nabla_{x_1} M_x, \nabla_{x_2} M_x \in T_{M_x}(\mathcal{P})$. Thus, the gradient vector $\nabla_x M_x$ is a higher-order tensor of the size $2 \times 2 \times 2$. For any such tensor $A \in \mathbb{R}^{2 \times 2 \times 2}$ and a vector $x \in \mathbb{R}^2$, we will use the notation: $\langle \langle A, x \rangle \rangle$ to imply $x_1 A(:, :, 1) + x_2 A(:, :, 2) \in T_{M(x)}(\mathcal{P})$. Therefore, $\langle \langle \langle \langle A, x \rangle \rangle \rangle x \rangle$ denotes a 2-vector given by $x_1 A(:, :, 1)x +$

$x_2 A(\cdot, \cdot, 2)x \in \mathbb{R}^2$. With this notation, we can express the gradient of E_{data} as follows.

LEMMA 1. *The gradient of E_{data} with respect to β , under the \mathbb{L}^2 norm, is given by:*

$$\begin{aligned} & -2 \left\{ \frac{1}{|\dot{\beta}(t)|} \left(M_{\beta(t)}^{-1} \dot{\beta}(t) + \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta(t)} \right) \right. \\ & - \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} M_{\beta(t)}^{-1} n_{\beta(t)} - \frac{1}{|\dot{\beta}(t)|} \left(\dot{n}_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t) \right. \\ & + n_{\beta}(t) n_{\beta}^T(t) \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta}(t) + 2 n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} \dot{n}_{\beta}(t) \left. \right) \\ & \left. + \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t) \right\} + \left\langle \text{tran}(\nabla_x M_{\beta(t)}^{-1}), n_{\beta}(t) \right\rangle n_{\beta}(t). \end{aligned} \quad (5)$$

where $\text{tran}(\nabla_x M_{\beta(t)}^{-1})$ is transpose of $\nabla_x M_{\beta(t)}^{-1}$.

A derivation of this expression is presented in the **Appendix**. Having an analytical expression for $\nabla_{\beta} E_{data}$ makes the optimization problem more efficient, as compared to purely numerical solutions.

Figure 2 shows an example of the gradient-based minimization of E_{data} in the middle panel. It shows a tensor field M and an initial curve β (in black). We update β iteratively using $-\nabla_{\beta} E_{data}$ and the result is drawn as a red curve. The corresponding evolution of E_{data} is plotted in the right panel.

2.2. Smoothness or Fiber Length Term

For regulating smoothness of the estimated curve, we follow a common approach from geometric active contours that is motivated in part by Euclidean heat flow. Define the smoothing energy function as $E_{smooth}(\beta) = \int_0^1 |\dot{\beta}(t)| dt$, which is equal to the length of the curve and is naturally invariant to any re-parameterization. It is shown in Kichenassamy et al. (1995) that the gradient of E_{smooth} is given by the Euclidean heat flow equation $\nabla E_{smooth}(\beta) = \kappa_{\beta} \mathbf{n}_{\beta}$, where κ_{β} is the curvature at each point of β and \mathbf{n}_{β} is the unit normal field along the curve. It is well known that this particular penalty on a curve's length leads to simultaneous smoothing and shrinking of a curve. If we rescale the curve to keep the original length, the main effect is that of smoothing. An example of this idea is illustrated in **Figure 3** that shows a curve evolving according to $-\nabla E_{smooth}$. The left panel shows the initial curve (in black), and its updates using the negative gradient of E_{smooth} . The corresponding decrease in E_{smooth} is plotted on the right.

2.3. Atlas-Based Shape Prior

The next term to consider is E_{prior} that forces the shapes of estimated fiber tracts to be similar to certain desired shapes. This term encodes the prior shape information about fibers connecting two ROIs, and is based on a statistical model that is learnt from the training or atlas data (generated by current local or global methods). In a brain connectome study framework, the brain is generally pre-segmented into small anatomical regions using software such as Freesurfer and ANTs (Avants et al., 2011), and fibers connecting two ROIs are extracted. However, due to

differences in sizes, orientations, and coordinate systems, these fibers connecting the same ROIs across subjects can not be directly used as prior for future fiber tractography. Removing these nuisance variable requires a formal definition of shape and shape space, and then one needs to develop a statistical model on this mathematical representation. Here we use elastic shape analysis developed in Srivastava and Klassen (2016) to represent and model fiber shapes. Specifically, we define \mathcal{S} , the shape space of all curves in D and impose a truncated wrapped normal distribution on this space to reach a statistical shape model. The parameters of this model are estimated *a priori* from the training or atlas data. We present a brief summary of the elastic shape analysis here and refer the reader to the textbook (Srivastava and Klassen, 2016) for more details. For a curve $\beta: [0, 1] \rightarrow D$, define $q(t) = \dot{\beta}(t)/\sqrt{|\dot{\beta}(t)|}$ be the *square-root velocity function* (SRVF) of β . This SRVF representation has an important property that a re-parameterization invariant Riemannian metric on the space of curves becomes the simple \mathbb{L}^2 metric under transformation. As a corollary, for any $q_1, q_2 \in \mathbb{L}^2$, we have $\|(q_1, \gamma) - (q_2, \gamma)\| = \|q_1 - q_2\|$, for any $\gamma \in \Gamma$, where Γ is the set of all orientation preserving diffeomorphisms of $[0, 1]$. Here (q, γ) stands for $(q \circ \gamma)\sqrt{\gamma'}$, representing the SRVF of the re-parameterized curve $\beta \circ \gamma$. If we rotate β by $O \in SO(2)$, we get $O^* \beta$, and the corresponding SRVF is given by $O^* q$.

Let β be a rescaled fiber curve such that it has unit length and let q be its SRVF. We define an orbit in the SRVF space as $[q] = \{O(q \circ \gamma)\sqrt{\gamma'} | O \in SO(2), \gamma \in \Gamma\}$, which denotes an equivalence class representing a shape. Let \mathcal{S} be the set of all such equivalence classes; \mathcal{S} is called the shape space. The term E_{prior} in the active contour model is a function of β , but our statistical models are built on \mathcal{S} such that E_{prior} can effectively encode the shape information and be invariant to the different sizes and coordinate systems of different brains. However, \mathcal{S} is a nonlinear manifold space. To build a statistical model on \mathcal{S} , we need some elementary tools such as efficient methods to calculate the mean and covariance matrix for a given set of data. Here we employ Karcher mean to calculate the mean shape of given fiber curves and the covariance matrix is calculated on the tangent space of \mathcal{S} at the estimated Karcher mean denoted by $T_{[\mu]}(\mathcal{S})$. The reader can refer to Srivastava et al. (2011) for the explicit procedures to calculate the Karcher mean and the covariance matrix.

Given a set of prior training shapes $\{[q_i], i = 1, \dots, n\}$ in \mathcal{S} , let us assume that we have computed their Karcher mean $[\mu]$ and covariance K . We define the prior shape model using a *truncated wrapped-normal density*, which is estimated from the data as follows. First, obtain the singular value decomposition of K as $[U, S, V] = \text{svd}(K)$, and let U_m be the m -dimensional principal subspace of $T_{[\mu]}(\mathcal{S})$ spanned by the first m columns of U . The shape prior distribution is defined as a wrapping of the truncated normal distribution mapped from U_m to \mathcal{S} using the exponential map. The truncated normal density on U_m is:

$$v \sim \frac{1}{Z} e^{-\frac{1}{2} (v_{\parallel}^T S_m^{-1} v_{\parallel} + \|v_{\perp}\|^2 / \delta^2)} \mathbf{1}_{\|v\| < \pi}, \quad (6)$$

where $v = \exp_{[\mu]}^{-1}([q])$, $v_{\parallel} = U_m^T v$ is the projection of v into U_m , $v_{\perp} = v - U_m v_{\parallel}$, S_m is the diagonal matrix containing the first m singular values, and Z is the normalizing constant. The scalar

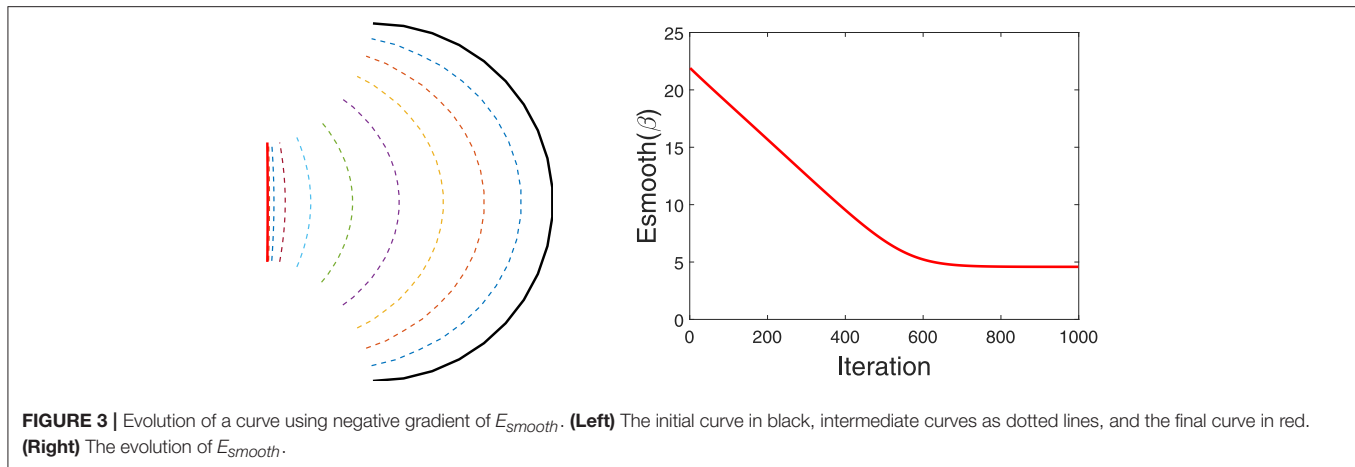


FIGURE 3 | Evolution of a curve using negative gradient of E_{smooth} . **(Left)** The initial curve in black, intermediate curves as dotted lines, and the final curve in red. **(Right)** The evolution of E_{smooth} .

value δ is chosen to be less than the smallest singular value in S_m . Suppose now that we have a test shape $[q]$ that represents a fiber tract during optimization process, and $v = \exp_{[\mu]}^{-1}([q])$ be the shooting vector from the mean $[\mu]$ to $[q]$. Now define $E_{prior}(q)$ to be the negative of the exponent in the shape prior given by Equation (6). That is, define $E_{prior}(q) = \frac{1}{2}v^T(U_m S_m^{-1} U_m^T)v + \frac{1}{2\delta^2}\|v - U_m U_m^T v\|^2$. Minimizing this functional is, therefore, equivalent to maximizing the likelihood of q under the chosen shape model. The gradient of E_{prior} with respect to v is equal to $w = Av$, where A is the matrix $A = U_m S_m^{-1} U_m^T + (I - U_m U_m^T)/\delta^2$. Notice that w is defined on the tangent space at μ rather than at q , so the final step is to parallel translate w from μ to q . Denote this parallel translation as $\tilde{w} = \nabla_q E_{prior}(q)$. An evolution of q along the negative gradient direction will result in an energy minimization precisely at the mean μ . The translated shooting vector \tilde{w} now represent the gradient of E_{prior} with respect to q . As the last step, this gradient is converted to $\nabla_\beta E_{prior}(\beta)$ using a numerical approximation.

Figure 4 shows a simple example of evolving a curve according to E_{prior} . The left panel shows the initial curve, and its updates using the negative gradient of E_{prior} . The corresponding decrease in E_{prior} is plotted on the right.

3. BAYESIAN TRACTOGRAPHY ALGORITHM

When we put together the three components of the energy, the shape of β is controlled by gradients of E_{data} , E_{prior} and E_{smooth} , the smoothness is controlled by E_{prior} and E_{smooth} , and the nuisance variables (placement, scale, and rotation) are controlled only by E_{data} . Now we summarize the overall algorithm for Bayesian tractography using the tensor field (Algorithm 1).

The advantage of the proposed framework is that it uses a global optimization to overcome issues such as fiber crossing and spatial noise. The final tracking result depends not only on the diffusion data, but also on prior shape information. The inclusion of shape prior distinguishes our method from other energy minimization based fiber-tracking algorithms, and is essential for

Algorithm 1: Bayesian Tractography Using Geometric Shape Priors

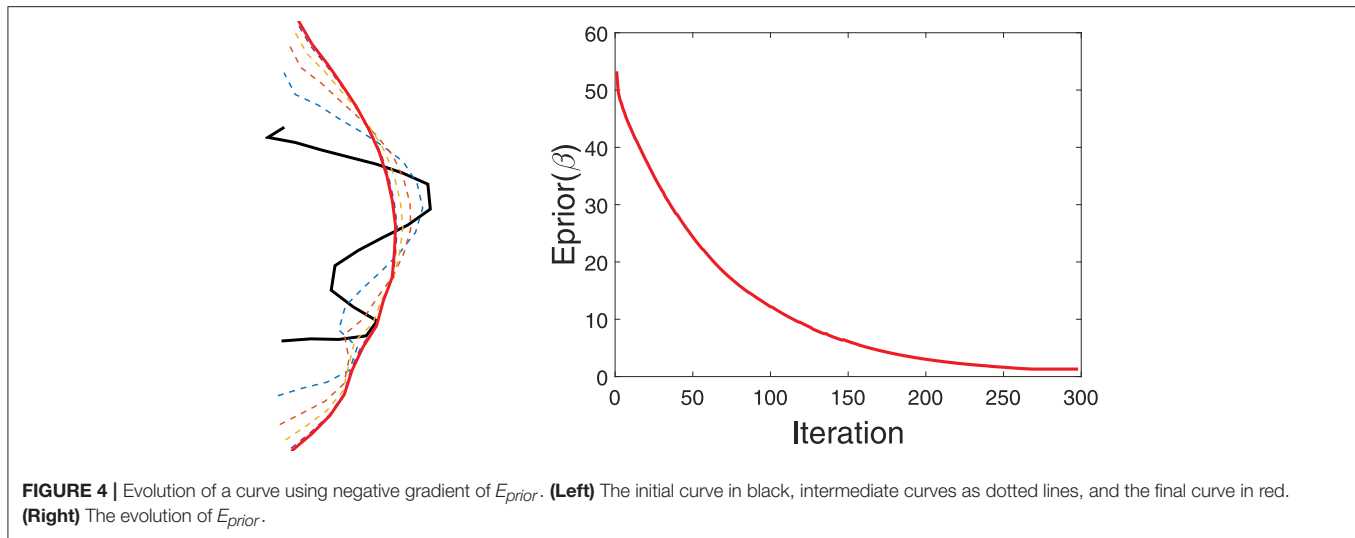
Data: Training fiber tracts connecting a pair of ROIs and the dMRI data

Result: Fiber tract β connecting the given two ROIs
Initialization: Calculate normalized mean shape μ and covariance K from training fiber tracts, perform SVD $[U, S, V] = \text{svd}(K)$. Use an existing method (e.g., probabilistic method) to obtain an initialization of β , denoted as β_1 .

for $i \leftarrow 1$ **to** $iter$ **do**

1. Calculate and save the length and the centroid of the current curve β_i ;
2. Convert β_i to SRVF representation q_i and normalize it $q_i = \frac{q_i}{\|q_i\|}$;
3. Calculate $A = U_m S_m^{-1} U_m^T + (I - U_m U_m^T)/\delta^2$, where U_m be the first m columns of U and $\delta \leq \lambda_m$, where λ_m is the m -th eigenvalue of K ;
4. Calculate shooting vector from μ to q_i , $v_i = \exp_{\mu}^{-1}(q_i)$;
5. Parallel transport Av_i from μ to q_i , $\tilde{w}_i = (Av_i)_{\mu \rightarrow q_i}$;
6. Travel a short distance ϵ from q_i along the geodesic defined by the shooting vector \tilde{w}_i , $q_i^{new} = \exp_{q_i}(-\epsilon \tilde{w}_i)$;
7. Convert q_i^{new} to its curve representation $\tilde{\beta}_i^{new} = \int_0^t q_i^{new} |q_i^{new}| du$ and scale and center $\tilde{\beta}_i^{new}$ to obtain β_i^{new} with the same length and centroid as β_i ;
8. Set $\nabla E_{prior}(\beta_i) = \frac{\beta_i - \beta_i^{new}}{\epsilon}$.
9. Evaluate $\nabla E_{data}(\beta_i)$ using Equation (8).
10. Evaluate $\nabla E_{smooth}(\beta) = \kappa_\beta \mathbf{n}$, where κ_β is the curvature at each point of β ;
11. Update the curves:
 $\beta_{i+1} = \beta_i - \lambda_1 \nabla E_{data}(\beta_i) - \lambda_2 \nabla E_{smooth}(\beta_i) - \lambda_3 \nabla E_{prior}(\beta_i)$.

end



the optimization procedure to come out of local solutions and reach a global solution. Most importantly, in our framework, the brain is parcellated into small regions, and the shapes of fibers connecting any pair of regions are found to be consistent. The proposed truncated wrapped-normal distribution can effectively capture the variation of shapes for each connection in the training data. In addition, since we reconstruct the whole fiber simultaneously by minimizing an energy function, the issue of fiber crossing has almost no detrimental effect of our fiber tracking algorithm.

As stated earlier, this Bayesian approach requires either a the training data or an atlas of fiber tracts between regions of interest, to estimate shape model and develop E_{prior} . We can construct such data using existing tractography algorithms with maybe human inspection for quality control. However, since such a construction is needed only once, it can be performed offline.

4. EXPERIMENTAL RESULTS

In this section we present some results using both simulated and real data to illustrate the performance of the proposed method.

4.1. Simulated 2-D tensor data

We first study our proposed tracking algorithm in the simulated settings. Let domain $D = [0, 1]^2$ for all our simulation examples. The tensor field on D , denoted by $M : D \rightarrow \mathcal{P}$, is generated using certain fibers that play the role of ground truth. We discretize the domain D into a 20×20 grid, and the tensor within each grid is decided by the tangent directions of the line segments within this grid. In addition, a 2D Gaussian smoothing is applied to smooth the tensor field before applying our algorithm.

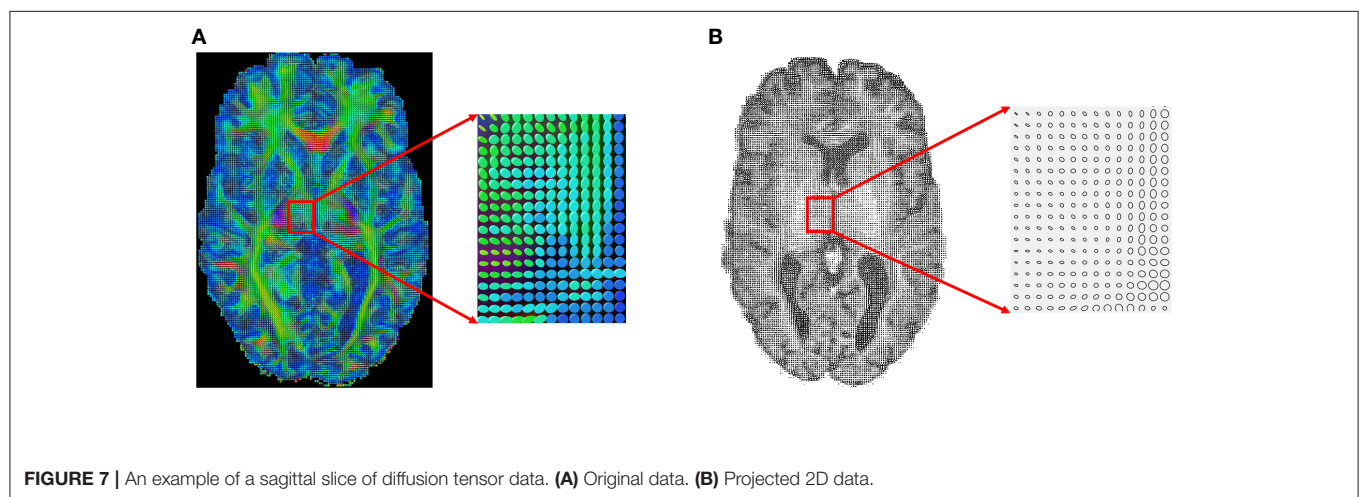
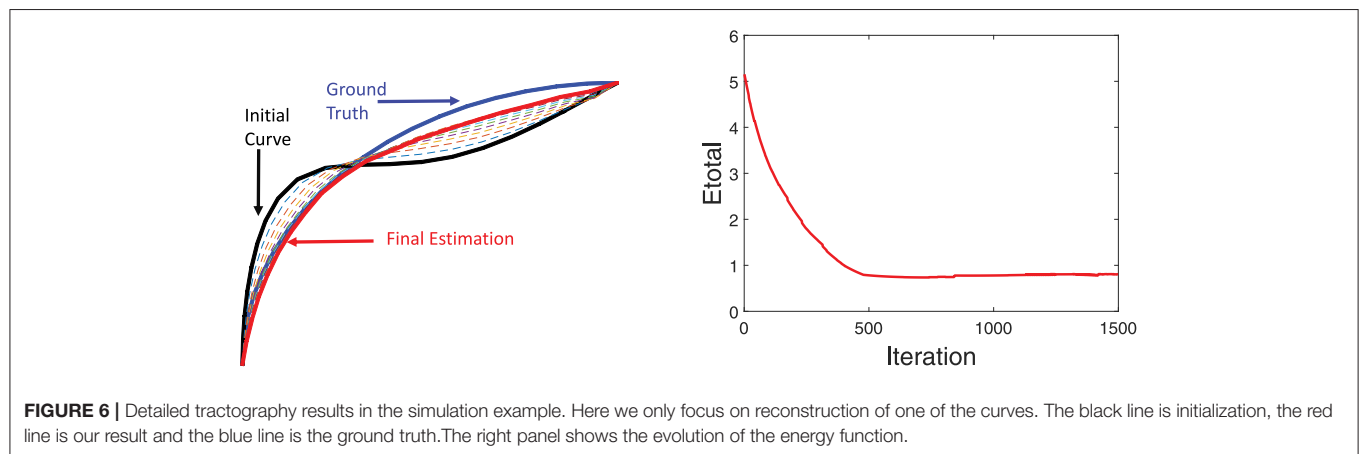
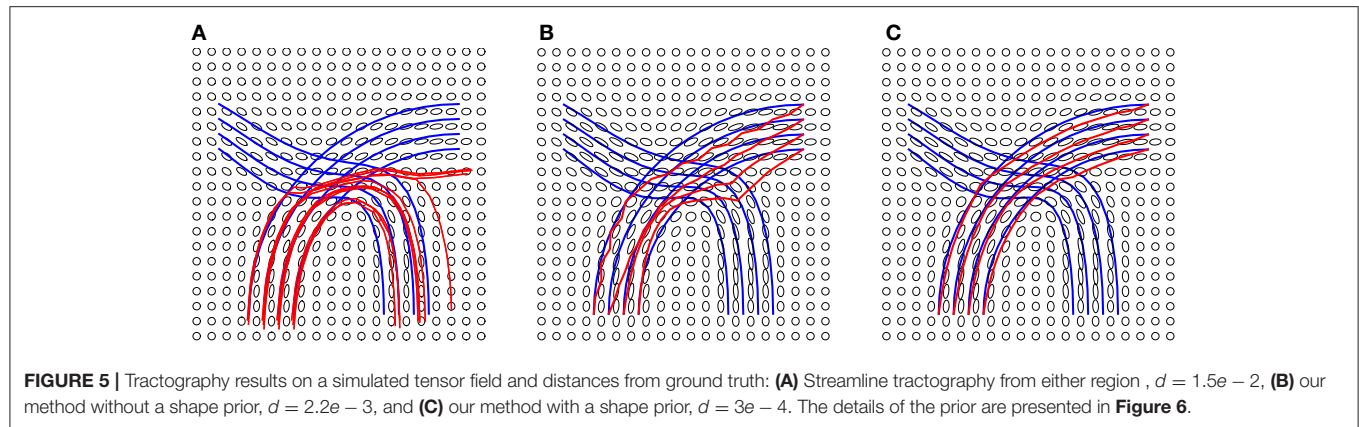
In the experiment presented in **Figure 5**, we use the blue lines as ground truth fiber tracts and generate a tensor field as shown in these panels. Then, using this tensor data, we estimate the fiber tracts using our and other methods, and the results are

shown in red lines. On the left side we show results from standard streamline tractography, using starting points on one end. Due to a crossing of fibers in the middle, these tracts get diverted and sent to wrong directions. In the middle panel, we show results from our method but without using the shape prior term. This time the end points of the tracts are correct (by initialization) but some of the fibers don't quite reach the desired shape. Finally, we optimize fiber tracts using the full energy functional, including the shape prior, and display these results in the right panel. By including all the three terms, we overcame issues caused by fiber crossing and local noise, and reached correct global structures. To better evaluate the tractography results, we calculate the distance between reconstructed fibers and ground truth using the L_2 norm. We first calculate the distance of each fiber from the ground truth and then use the mean of all distances to quantify the difference between reconstructed fiber bundle and ground truth fiber bundle. The distances for each method are given in **Figure 5**.

Additional details of this simulation experiment are presented in **Figure 6**, which shows evolution of a single fiber under E_{total} . The left panel shows the initial curve (black), the final curve (red), and the ground truth curve (blue). The right panel shows the evolution E_{total} during this iteration. In this experiment, we used the weights $\lambda_1 = 0.8$, $\lambda_2 = 0.1$, and $\lambda_3 = 0.1$.

4.2. Experiments Using Real Data

Next, we apply our method to some real datasets—dMRI images downloaded from the Human Connectome Project (HCP) (Van Essen et al., 2012). HCP contains about 900 subjects with diffusion MRI, but here we have used only 30 subjects for our experiments. The dMRI images in HCP has an isotropic resolution of 1.25 mm. To estimate a diffusion tensor at each voxel, we use the open source software Dipy (Garyfallidis et al., 2014). **Figure 7A** shows one slice of the 3×3 diffusion tensors estimated from a randomly selected dMRI image in HCP; a zoom-in of a small part of the image is shown on its right. Since in this paper we restrict to a 2D domain to illustrate our idea,



we convert 3×3 diffusion tensors in the original data to 2×2 tensors by removing the diffusion directions perpendicular to the 2D slice plane. **Figure 7B** shows an example of this projection and shows the 2D tensors in form of their level sets or ellipses at each pixel location.

In the results presented here, we focus on estimating a set of fiber curves connecting the left and right superior frontal gyri.

In order to generate a prior shape model, we use tracts extracted for 30 subjects between these regions as the training dataset. These tracts were manually identified with the help of Freesurfer Destrieux Atlas (Destrieux et al., 2010) and the fiber curves built using the FACT method. These fibers are displayed on the left side of **Figure 8**. The Karcher mean μ of these fibers in the shape space \mathcal{S} is shown in the middle panel and the five dominant

principal components of the Karcher covariance are displayed in the right panel. These dominant directions are computed by projecting the given shapes $[q_i]$ in the tangent space $T_{[\mu]}(\mathcal{S})$ using the inverse exponential map, i.e., $v_i = \exp_{[\mu]}^{-1}([q_i])$, and the computing principal components of the set $\{v_i\}$ in the vector

space $T_{[\mu]}(\mathcal{S})$. These principal directions, which as straight lines in $T_{[\mu]}(\mathcal{S})$ passing through $[\mu]$ in the middle, are then wrapped back on \mathcal{S} using the exponential maps. Each row of the right panel in **Figure 8** shows plots one such direction, going from the largest variability to smallest from top to bottom.

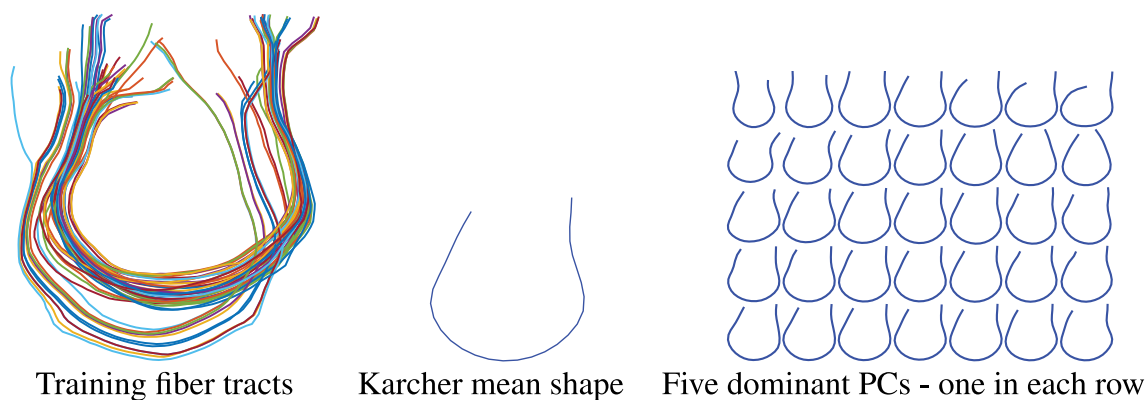


FIGURE 8 | Thirty training samples of fiber tracts, their Karcher mean and principal directions of shape variation. The rightmost panel from top to bottom represents the first 5 principal directions of variation in the training data.

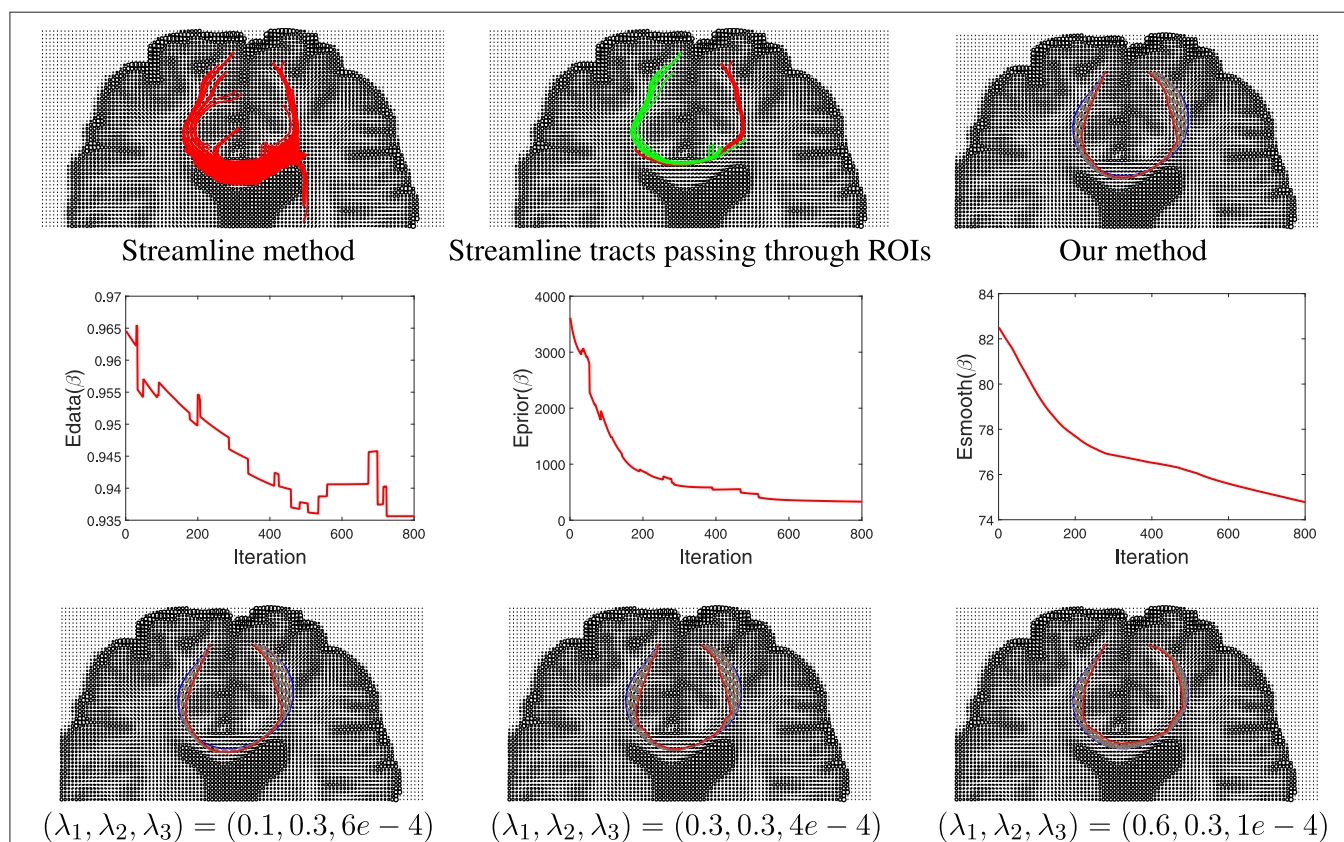


FIGURE 9 | Results comparison between streamline method and our method. In the **top** row, the **left** panel shows the results using a streamline method, the **middle** panel shows some selected curves from that set that reach the two ROIs (different colors represent curves passing different regions), and the **right** panel shows tractography result using our Bayesian method. Here the blue line shows the initialization and red line is final result. The **middle** row shows the evolution of the three energy components in this estimation. The **bottom** row shows our tractography results under different weights of the energy components.

Having developed a prior model for fiber shapes from the training data, we then apply our Bayesian method to the tensor data, especially focusing on the areas where the streamline method fails, and the results are presented in **Figure 9**. We first show the results of the streamline method, using seeds from either ROI, in the first two panels. While the left panel in the top row gives an appearance that we have some fibers connecting the two ROIs, a closer look shows that this is actually not the case. In the middle panel we color the curves differently depending on which ROI is the seed located in. One can see that the set of curves—red and green—do not reach the other ROI. They start from the ROI containing the seeds and diverge in

the middle. This is in contradiction to the anatomical knowledge that the two regions are indeed connected through white matter fiber tracts. Using the proposed Bayesian technique, we obtained result shown in the rightmost panel of the top row. This picture shows an arbitrarily initialized curve drawn in blue, and the final estimated curve drawn in red color. The corresponding evolutions of the three energy terms— E_{data} , E_{prior} , and E_{smooth} —are shown in the middle row of this figure. Each one of these terms show a substantial decrease in their values during the iteration process.

In order to study the impact of the weights λ_1 , λ_2 , and λ_3 on the final result, we generated estimates for a few different

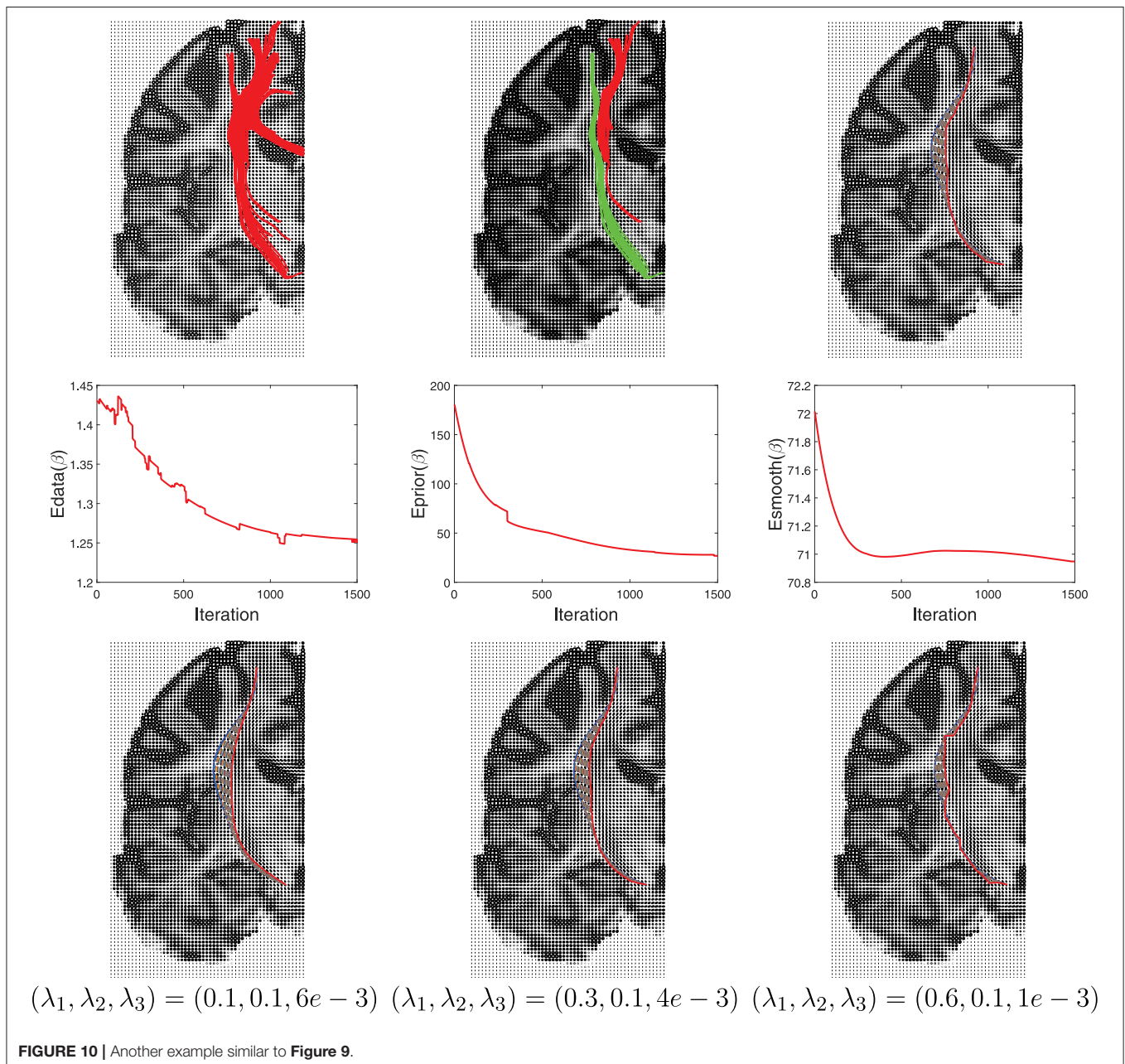


FIGURE 10 | Another example similar to **Figure 9**.

combinations of these weights. The results are shown in the last row of this figure. In case where the weight for shape prior is high, the final result is close to the prior mean. In contrast, when the weight for the data term is high, there is a better agreement between the curve and the tensor field.

Another example of this Bayesian estimation is presented in **Figure 10** with similar settings. In this case the ROIs used are right hippocampus and right pericalcaral.

4.3. Extension to Tractography Using HARDI Data

The proposed framework can be extended to HARDI data, where an ODF is used to better represent the underlying diffusion profile. The data term is now defined as:

$$E_{data}[\beta] = \int_0^1 -f_{\beta(t)}(n_{\beta(t)}) dt, \text{ where } n_{\beta(t)} = \frac{\dot{\beta}(t)}{|\dot{\beta}(t)|}. \quad (7)$$

Here $n_{\beta(t)}$ denotes the unit vector tangent to β at $\beta(t)$ and f_p is the ODF at $p \in D$. The integrand is low at a location where the fiber tract is aligned with the ODF field and vice-versa. The next step is to derive the gradient of E_{data} with respect to β for use in gradient-based optimization. We can express the gradient of E_{data} as follows.

LEMMA 2. The gradient of E_{data} with respect to β , under the \mathbb{L}^2 norm, is given by:

$$-\frac{\dot{\beta}^T(t)\ddot{\beta}(t)}{|\dot{\beta}(t)|^3} \left(I - n_{\beta(t)}n_{\beta(t)}^T \right) \nabla_{n_{\beta(t)}}^T f_{\beta(t)}(n_{\beta(t)}) - \frac{2}{|\dot{\beta}(t)|} \dot{n}_{\beta(t)}n_{\beta(t)}^T(t) \nabla_{n_{\beta(t)}}^T f_{\beta(t)}(n_{\beta(t)}) + \frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta(t)}n_{\beta(t)}^T \right) \nabla_{n_{\beta(t)}}^2 f_{\beta(t)}(n_{\beta(t)}) \dot{n}_{\beta(t)}. \quad (8)$$

A derivation of this expression is presented in the **Appendix**. We also show an experiment result on an ODF data in **Figure 11**. We use the blue lines as ground truth fiber tracts and generate ODF data as shown in **Figure 11A**. Under this ODF field, we estimate the fiber tracts using our method. The final reconstructed tracts are shown in red lines. In the middle panel, we show an evolution of a single fiber under E_{total} . In the right panel, we show the evolution E_{total} of each iteration.

5. CONCLUSION AND DISCUSSION

This paper introduces a Bayesian approach for estimating fiber tracts, between given pairs of points in a human brain, using dMRI and HARDI data. The basic idea is to define a composite energy functional, using a linear combinations of terms that relate to data, curve smoothness, and a prior shape model, and then use the gradient of this energy to iteratively optimize a contour. There are several novelties in this setup: (1) the data term is locally scale-invariant and measures only the agreement of the fiber direction with the given diffusion tensor field, (2) the length of the fiber is kept as a separate term, in order to have an additional control over fiber size, and (3) an external

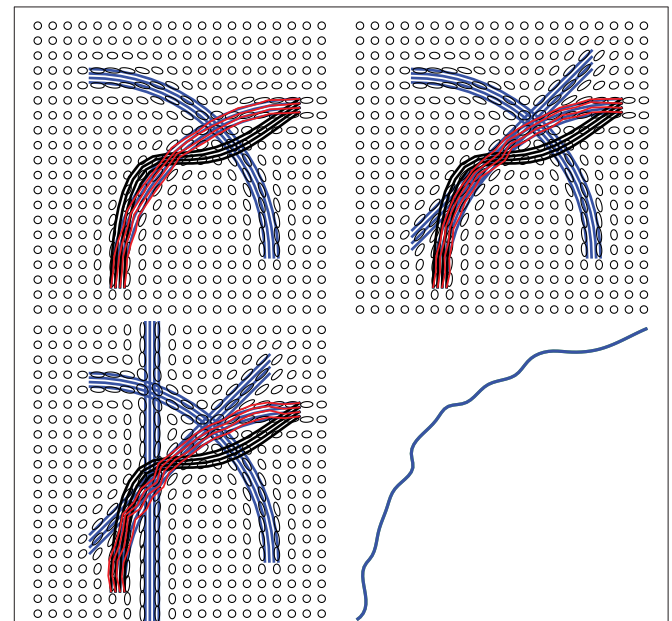


FIGURE 12 | Examples showing that the proposed method can handle crossing and kissing fibers. Red lines are our tractography results, blue lines are ground truth and black lines are initializations. From upper left panel to bottom left panel, more and more crossing bundles are added into the simulation. The bottom right panel shows the shape prior used in our model.

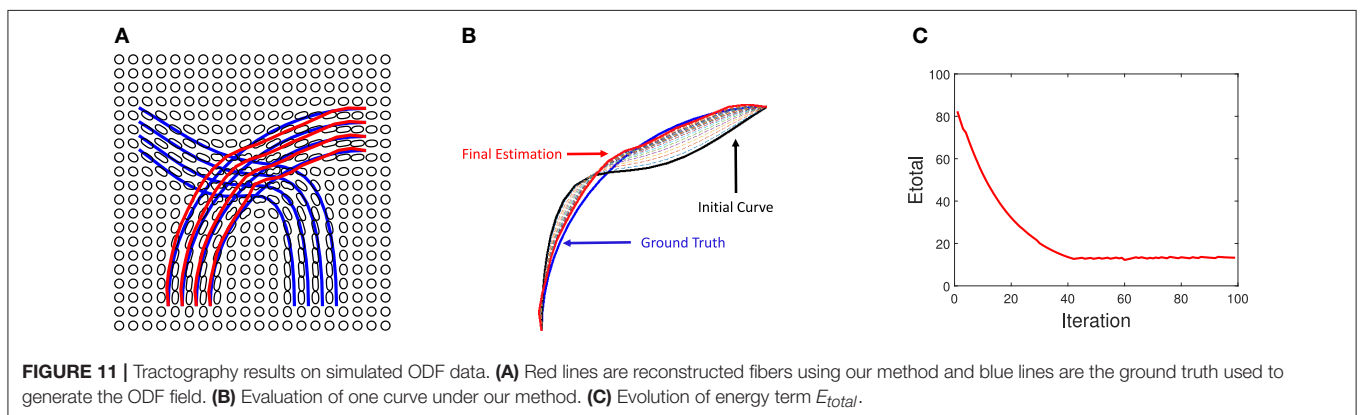


FIGURE 11 | Tractography results on simulated ODF data. **(A)** Red lines are reconstructed fibers using our method and blue lines are the ground truth used to generate the ODF field. **(B)** Evaluation of one curve under our method. **(C)** Evolution of energy term E_{total} .

term involving statistical shape models, of fibers tracts connecting given regions, is used to improve optimization and interpretability. These shape models can come from training data developed using manual interventions or population atlases established from previous studies. The gradients of all the terms have analytical forms, making the gradient-based optimization very efficient. This framework is demonstrated successfully using simulated 2D tensor fields and 2D slices of volume dMRI data.

One advantage of our method is that it can naturally handle crossing bundles since we construct the streamline as a whole object. Relying on the prior shape information, we can reconstruct a fiber curve that have similar geometry to the prior knowledge. **Figure 12** illustrates one example that the proposed method is not sensitive to local fiber crossing. The blue lines are ground truth to generate the tensor field. From upper left to bottom left, more fibers were added to a region, which complicates the underlying tensor field. For the two selected regions, we initialize some black lines to connect them and the red lines are the final tractography results using our method. Those results indicates that our method can successfully reconstruct the fiber bundles in this challenge situation. The bottom right panel shows the shape prior that being used in our implementation.

However, the proposed Bayesian method needs to specify the starting and ending points for each extracted tract. To ensure that there is a tract between two ROIs, we currently rely on the

atlas data. This procedure may end up with false positives, e.g., identifying a tract that does not exist. A future pruning procedure can be added as a post processing step, relying perhaps on the minimum energy as the reviewer has suggested. As another criterion, the diffusion profile along a tract can possibly be used as a feature to determine whether a tract is a false or a true positive.

As a future work, this framework can be naturally implemented using 3D dMRI data, and resulted tractography can be compared with some state of the art techniques.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication. XD, ZZ, and AS have contributed in development of theory and computer implementation.

ACKNOWLEDGMENTS

This research was supported in part by NSF grants DMS 1621787 and CCF 1617397 to AS. ZZ was partially supported by NSF grant DMS-1127914 to SAMSI. Data were provided in part by the HCP, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657).

REFERENCES

- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025
- Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). *In vivo fiber tractography using DT-MRI data. Magn. Reson. Med.* 44, 625–632. doi: 10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O
- Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155. doi: 10.1016/j.neuroimage.2006.09.018
- Behrens, T. E., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted mr imaging. *Magn. Reson. Med.* 50, 1077–1088. doi: 10.1002/mrm.10609
- Bryner, D., Srivastava, A., and Huynh, Q. (2013). Elastic shape models for improving segmentation of object boundaries in synthetic aperture sonar images. *Comput. Vis. Image Understand.* 117, 1695–1710. doi: 10.1016/j.cviu.2013.07.001
- Cheng, G., Salehian, H., Forder, J., and Vemuri, B. C. (2015). Tractography from HARDI using an intrinsic unscented kalman filter. *IEEE Trans. Med. Imaging* 34, 298–305. doi: 10.1109/TMI.2014.2355138
- Cheng, H., Wang, Y., Sheng, J., Kronenberger, W. G., Mathews, V. P., Hummer, T. A., et al. (2012). Characteristics and variability of structural networks derived from diffusion tensor imaging. *Neuroimage* 61, 1153–1164. doi: 10.1016/j.neuroimage.2012.03.036
- Christiaens, D., Reisert, M., Dhollander, T., Maes, F., Sunaert, S., and Suetens, P. (2014). “Atlas-guided global tractography: imposing a prior on the local track orientation,” in *Computational Diffusion MRI* (Cambridge, MA: Springer), 115–123.
- Cook, P. A., Zhang, H., Awate, S. P., and Gee, J. C. (2008). “Atlas-guided probabilistic diffusion-tensor fiber tractography,” in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on* (Paris: IEEE), 951–954. doi: 10.1109/ISBI.2008.4541155
- de Reus, M. A., and van den Heuvel, M. P. (2013). The parcellation-based connectome: limitations and extensions. *Neuroimage* 80, 397–404. doi: 10.1016/j.neuroimage.2013.03.053
- Descoteaux, M. (2015). “High Angular Resolution Diffusion Imaging (HARDI),” in *Wiley Encyclopedia of Electrical and Electronics Engineering*, ed J. G. Webster (John Wiley & Sons, Inc.), 1–25. doi: 10.1002/047134608X.W8258
- Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010
- Duncan, J. S., Papademetris, X., Yang, J., Jackowski, M., Zeng, X., and Staib, L. H. (2004). Geometric strategies for neuroanatomic analysis from MRI. *Neuroimage* 23, S34–S45. doi: 10.1016/j.neuroimage.2004.07.027
- Durante, D., and Dunson, D. B. (2017). Bayesian inference and testing of group differences in brain networks. *Bayesian Anal.* doi: 10.1214/16-BA1030. [Epub ahead of print].
- Eckstein, I., Shattuck, D. W., Stein, J. L., McMahon, K. L., de Zubicaray, G., Wright, M. J., et al. (2009). Active fibers: matching deformable tract templates to diffusion tensor images. *Neuroimage* 47, T82–T89. doi: 10.1016/j.neuroimage.2009.01.065
- Faugeras, O., Adde, G., Charpiat, G., Chef d’Hotel, C., Clerc, M., Deneux, T., et al. (2004). Variational, geometric, and statistical methods for modeling brain anatomy and function. *NeuroImage* 23, S46–S55. doi: 10.1016/j.neuroimage.2004.07.015
- Fornito, A., Zalesky, A., and Breakspear, M. (2013). Graph analysis of the human connectome: promise, progress, and pitfalls. *Neuroimage* 80, 426–444. doi: 10.1016/j.neuroimage.2013.04.087
- Friman, O., Farneback, G., and Westin, C.-F. (2006). A bayesian approach for stochastic white matter tractography. *IEEE Trans. Med. Imaging* 25, 965–978. doi: 10.1109/TMI.2006.877093
- Fuster, A., Tristan-Vega, A., Haije, T. D., Westin, C.-F., and Florack, L. (2014). “A novel Riemannian metric for geodesic tractography in DTI,” in *Computational Diffusion MRI and Brain Connectivity* (Nagoya: Springer), 97–104. doi: 10.1007/978-3-319-02475-2_9

- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., et al. (2014). Dipy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* 8:8. doi: 10.3389/fninf.2014.00008
- Hagmann, P., Thiran, J. P., Jonasson, L., Vanderghenst, P., Clarke, S., Maeder, P., et al. (2003). DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *Neuroimage* 19, 545–554. doi: 10.1016/S1053-8119(03)00142-3
- Hao, X., Zygmunt, K., Whitaker, R. T., and Fletcher, P. T. (2014). Improved segmentation of white matter tracts with adaptive Riemannian metrics. *Med. Image Anal.* 18, 161–175. doi: 10.1016/j.media.2013.10.007
- Joshi, S. H., and Srivastava, A. (2009). Intrinsic Bayesian active contours for extraction of object boundaries in images. *Int. J. Comput. Vis.* 81, 331–355. doi: 10.1007/s11263-008-0179-8
- Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., and Yezzi, A. (1995). “Gradient flows and geometric active contour models,” in *Fifth ICCV* (Cambridge, MA), 810–815. doi: 10.1109/ICCV.1995.466855
- Lankton, S., Melonakos, J., Malcolm, J., Dambreville, S., and Tannenbaum, A. (2008). “Localized statistics for DW-MRI fiber bundle segmentation,” in *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on* (Anchorage, AK: IEEE), 1–8.
- Lenglet, C., Deriche, R., and Faugeras, O. (2004). “Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping,” in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 127–140. doi: 10.1007/978-3-540-24673-2_11
- Li, W., and Hu, X. (2013). Robust tract skeleton extraction of cingulum based on active contour model from diffusion tensor MR imaging. *PLoS ONE* 8:e56113. doi: 10.1371/journal.pone.0056113
- Maier-Hein, K., Neher, P., Houde, J.-C., Cote, M.-A., Garyfallidis, E., Zhong, J., et al. (2016). Tractography-based connectomes are dominated by false-positive connections. *bioRxiv*.
- Melonakos, J. (2009). *Geodesic Tractography Segmentation for Directional Medical Image Analysis*. Atlanta, GA: ProQuest.
- Melonakos, J., Pichon, E., Angenent, S., and Tannenbaum, A. (2008). Finsler active contours. *IEEE Trans. Patt. Anal. Mach. Intell.* 30, 412–423. doi: 10.1109/TPAMI.2007.70713
- Mohan, V., Melonakos, J., Niethammer, M., Kubicki, M., and Tannenbaum, A. R. (2009). *Finsler Level Set Segmentation for Imagery in Oriented Domains*. Georgia Institute of Technology.
- Mori, S., Crain, B. J., Chacko, V. P., and van Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45, 265–269. doi: 10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3
- Mori, S., and van Zijl, P. C. (2002). Fiber tracking: principles and strategies - a technical review. *NMR Biomed.* 15, 468–480. doi: 10.1002/nbm.781
- Mori, S., Wakana, S., Nagae-Poetscher, L., and van Zijl, P. (2005). *MRI Atlas of Human White Matter, 1st Edn*. Baltimore, MD: Elsevier.
- O’Donnell, L., Haker, S., and Westin, C. (2002). “New approaches to estimation of white matter connectivity in diffusion tensor MRI: elliptic PDEs and geodesics in a tensor-warped space,” in *Proceedings of Medical Imaging, Computing and Computer Assisted Intervention, Lecture Notes in Computer Science, Vol. 2488* (Berlin: Springer), 459–466.
- Pichon, E., Westin, C.-F., and Tannenbaum, A. R. (2005). “A hamilton-jacobi-bellman approach to high angular resolution diffusion tractography,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Berlin; Heidelberg: Springer), 180–187.
- Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., and Kiselev, V. (2011). Global fiber reconstruction becomes practical. *Neuroimage* 54, 955–962. doi: 10.1016/j.neuroimage.2010.09.016
- Srivastava, A., and Klassen, E. (2016). *Functional and Shape Data Analysis*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4939-4020-2
- Srivastava, A., Klassen, E., Joshi, S., and Jermyn, I. (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Trans. Patt. Anal. Mach. Intell.* 33, 1415–1428. doi: 10.1109/TPAMI.2010.184
- Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E., Bucholz, R., et al. (2012). The human connectome project: a data acquisition perspective. *NeuroImage* 62, 2222–2231. doi: 10.1016/j.neuroimage.2012.02.018
- Yap, P.-T., Gilmore, J. H., Lin, W., and Shen, D. (2011). “Longitudinal tractography with application to neuronal fiber trajectory reconstruction in neonates,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Berlin; Heidelberg: Springer), 66–73.
- Zach, C., Shan, L., and Niethammer, M. (2009). “Globally optimal finsler active contours,” in *Joint Pattern Recognition Symposium* (Jena: Springer), 552–561. doi: 10.1007/978-3-642-03798-6_56

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Dong, Zhang and Srivastava. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Lemma 1

In this section we derive an expression for the gradient of $E_{data}[\beta]$ with respect to β . Let $h \in \mathcal{B}$ be a perturbation to the curve β such that it is zero at the boundaries, i.e., $h: [0, 1] \rightarrow \mathbb{R}^2$ and $h(0) = h(1) = 0$. Since, $E_{data}[\beta + \epsilon h] = \int_0^1 n_{\beta+\epsilon h}^T(t) M_{\beta(t)+\epsilon h(t)}^{-1} n_{\beta+\epsilon h}^T(t) dt$, the directional derivative of E_{data} in the direction of h is given by:

$$\frac{d}{d\epsilon}|_{\epsilon=0} E_{data}[\beta + \epsilon h] = \int_0^1 \left(2n_{\beta}^T(t) M_{\beta(t)}^{-1} u_{\beta,h}(t) + n_{\beta}^T(t) \left\langle \nabla_x M_{\beta(t)}^{-1}, h(t) \right\rangle n_{\beta}(t) \right) dt,$$

where: $u_{\beta,h}(t) = \frac{d}{d\epsilon}|_{\epsilon=0} (n_{\beta+\epsilon h}(t)) = \frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t) n_{\beta}^T(t) \right) \dot{h}(t) \equiv A_{\beta}(t) \dot{h}(t)$. The last equality is used to define $A_{\beta}(t)$. We simplify the two terms one by one:

- **First Term:** Using integration by parts and using the boundary conditions $h(0) = h(1) = 0$, the first term becomes:

$$\begin{aligned} \int_0^1 2n_{\beta}^T(t) M_{\beta(t)}^{-1} u_{\beta,h}(t) dt &= \int_0^1 2n_{\beta}^T(t) M_{\beta(t)}^{-1} A_{\beta(t)} \dot{h}(t) dt \\ &= - \int_0^1 \left\langle 2 \frac{d}{dt} \left(A_{\beta(t)} M_{\beta(t)}^{-1} n_{\beta}(t) \right), h(t) \right\rangle dt \end{aligned}$$

Here

$$\begin{aligned} \frac{d}{dt} \left(A_{\beta(t)} M_{\beta(t)}^{-1} n_{\beta}(t) \right) &= \frac{d}{dt} \left(\frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t) n_{\beta}^T(t) \right) M_{\beta(t)}^{-1} n_{\beta}(t) \right) \\ &= \frac{1}{|\dot{\beta}(t)|} \left(M_{\beta(t)}^{-1} \dot{n}_{\beta}(t) + \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta}(t) \right) - \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} M_{\beta(t)}^{-1} n_{\beta}(t) \\ &\quad - \frac{1}{|\dot{\beta}(t)|} \left(\dot{n}_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t) + n_{\beta}(t) n_{\beta}^T(t) \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta}(t) + 2n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} \dot{n}_{\beta}(t) \right) \\ &\quad + \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t), \text{ where } \dot{n}_{\beta}(t) = \frac{d}{dt} n_{\beta}(t) = \frac{\ddot{\beta}(t)}{|\dot{\beta}(t)|} - \frac{\dot{\beta}(t) \dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3}. \end{aligned}$$

- **Second Term:** The second term can be rearranged as:

$$\int_0^1 \left\langle \left\langle \text{tran}(\nabla_x M_{\beta(t)}^{-1}), n_{\beta}(t) n_{\beta}(t), h(t) \right\rangle \right\rangle dt$$

where $\text{tran}(\nabla_x M_{\beta(t)}^{-1})$ is transpose of $\nabla_x M_{\beta(t)}^{-1}$.

Thus, the full gradient of E_{data} with respect to β is given by:

$$\begin{aligned} &- 2 \left\{ \frac{1}{|\dot{\beta}(t)|} \left(M_{\beta(t)}^{-1} \dot{n}_{\beta}(t) + \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta}(t) \right) - \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} M_{\beta(t)}^{-1} n_{\beta}(t) \right. \\ &- \frac{1}{|\dot{\beta}(t)|} \left(\dot{n}_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t) + n_{\beta}(t) n_{\beta}^T(t) \left\langle \nabla_x M_{\beta(t)}^{-1}, \dot{\beta}(t) \right\rangle n_{\beta}(t) + 2n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} \dot{n}_{\beta}(t) \right) \\ &\left. + \frac{\dot{\beta}^T(t) \ddot{\beta}(t)}{|\dot{\beta}(t)|^3} n_{\beta}(t) n_{\beta}^T(t) M_{\beta(t)}^{-1} n_{\beta}(t) \right\} + \left\langle \left\langle \text{tran}(\nabla_x M_{\beta(t)}^{-1}), n_{\beta}(t) \right\rangle \right\rangle n_{\beta}(t). \end{aligned}$$

Lemma 2

Let's denote f_p as the ODF at $p \in D$ and for simplicity, $f(t)$ will be used to denote $f_{\beta(t)}(n_{\beta}(t))$ in the following derivation. In this section we derive an expression for the gradient of $E_{data}[\beta]$ with respect to β . Let $h \in \mathcal{B}$ be a perturbation to the curve β such that it is zero at the boundaries, i.e., $h: [0, 1] \rightarrow \mathbb{R}^2$ and $h(0) = h(1) = 0$. Since, $E_{data}[\beta + \epsilon h] = \int_0^1 f(n_{\beta+\epsilon h}(t)) dt$, the directional derivative of E_{data} in the direction of h is given by:

$$\frac{d}{d\epsilon}|_{\epsilon=0} E_{data}[\beta + \epsilon h] = \int_0^1 \nabla_{n_{\beta}} f(t) u_{\beta,h}(t) dt,$$

where: $u_{\beta,h}(t) = \frac{d}{d\epsilon}|_{\epsilon=0}(n_{\beta+\epsilon h}(t)) = \frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \dot{h}(t) \equiv A_{\beta}(t)\dot{h}(t)$. Using integration by parts and using the boundary conditions $h(0) = h(1) = 0$, the term becomes:

$$\int_0^1 \nabla_{n_{\beta}} f(t) u_{\beta,h}(t) dt = \int_0^1 \nabla_{n_{\beta}} f(t) A_{\beta}(t) \dot{h}(t) dt = - \int_0^1 \left\langle \frac{d}{dt} \left(A_{\beta}(t) \nabla_{n_{\beta}}^T f(t) \right) h(t) \right\rangle dt$$

Here

$$\begin{aligned} \frac{d}{dt} \left(A_{\beta}(t) \nabla_{n_{\beta}}^T f(t) \right) &= \frac{d}{dt} \left(\frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \nabla_{n_{\beta}}^T f(t) \right) \\ &= - \frac{\dot{\beta}^T(t)\ddot{\beta}(t)}{|\dot{\beta}(t)|^3} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \nabla_{n_{\beta}}^T f(t) - \frac{2}{|\dot{\beta}(t)|} \dot{n}_{\beta}(t)n_{\beta}^T(t) \nabla_{n_{\beta}}^T f(t) \\ &\quad + \frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \nabla_{n_{\beta}}^2 f(t) \dot{n}_{\beta}(t), \end{aligned}$$

$$\text{where } \dot{n}_{\beta}(t) = \frac{d}{dt} n_{\beta}(t) = \frac{\ddot{\beta}(t)}{|\dot{\beta}(t)|} - \frac{\dot{\beta}(t)\dot{\beta}^T(t)\ddot{\beta}(t)}{|\dot{\beta}(t)|^3}.$$

Thus, the full gradient of E_{data} with respect to β is given by:

$$- \frac{\dot{\beta}^T(t)\ddot{\beta}(t)}{|\dot{\beta}(t)|^3} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \nabla_{n_{\beta}}^T f(t) - \frac{2}{|\dot{\beta}(t)|} \dot{n}_{\beta}(t)n_{\beta}^T(t) \nabla_{n_{\beta}}^T f(t) + \frac{1}{|\dot{\beta}(t)|} \left(I - n_{\beta}(t)n_{\beta}^T(t) \right) \nabla_{n_{\beta}}^2 f(t) \dot{n}_{\beta}(t).$$



OPEN ACCESS

Edited by:

Pedro Antonio Valdes-Sosa,
Joint China-Cuba Laboratory for
Frontier Research in Translational
Neurotechnology, China

Reviewed by:

Fabio Grizzi,
Humanitas Clinical and Research
Center, Italy

M. Mallar Chakravarty,
McGill University, Canada

*Correspondence:

Daniel J. Tward
dtward@cis.jhu.edu

[†] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The key portion is "the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Specialty section:

This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 27 April 2017

Accepted: 02 October 2017

Published: 18 October 2017

Citation:

Tward DJ and Miller MI for the Alzheimer's Disease Neuroimaging Initiative (2017) On the Complexity of Human Neuroanatomy at the Millimeter Morphome Scale: Developing Codes and Characterizing Entropy Indexed to Spatial Scale. *Front. Neurosci.* 11:577. doi: 10.3389/fnins.2017.00577

On the Complexity of Human Neuroanatomy at the Millimeter Morphome Scale: Developing Codes and Characterizing Entropy Indexed to Spatial Scale

Daniel J. Tward* and Michael I. Miller for the Alzheimer's Disease Neuroimaging Initiative[†]

Center for Imaging Science, Department of Biomedical Engineering, Kavli Neuroscience Discovery Institute, Johns Hopkins University, Baltimore, MD, United States

In this work we devise a strategy for discrete coding of anatomical form as described by a Bayesian prior model, quantifying the entropy of this representation as a function of code rate (number of bits), and its relationship geometric accuracy at clinically relevant scales. We study the shape of subcortical gray matter structures in the human brain through diffeomorphic transformations that relate them to a template, using data from the Alzheimer's Disease Neuroimaging Initiative to train a multivariate Gaussian prior model. We find that the at 1 mm accuracy all subcortical structures can be described with less than 35 bits, and at 1.5 mm error all structures can be described with less than 12 bits. This work represents a first step towards quantifying the amount of information ordering a neuroimaging study can provide about disease status.

Keywords: computational anatomy, diffeomorphometry, neuroimaging, anatomical prior, entropy, complexity, rate distortion

1. INTRODUCTION

The trend toward a quantitative, task based, understanding of medical images leads to the simple goal of answering "how many bits of information would one expect a medical image to contain about disease status?" Knowing the answer to this question could impact a clinician's decision of whether or not to order an imaging study, particularly in the case where it involves ionizing radiation. This quantity can be studied in terms of mutual information between disease status and anatomical form.

$$MI(\text{disease, anatomy}) = H(\text{anatomy}) - H(\text{anatomy}|\text{disease}) \quad (1)$$

where MI is mutual information, and $H(\cdot)$ is entropy and $H(\cdot|\cdot)$ is conditional entropy.

In general, the higher the complexity of a population of normal anatomy, the less informative is a realization as manifest by an MRI concerning some disease. On the other hand, the simpler the class of anatomy, the more information gained by making an MRI. This is reflected by sensitivity and specificity of statistical tests.

Other information theoretic quantities could have a direct impact on clinical decision making as well. The inverse of the Fisher information puts a lower bound on the variance of any unbiased estimator (the Cramér-Rau inequality). The Kullback-Leibler divergence $D(P_1||P_2)$ between two

probability distributions P_1 and P_2 can be used to quantify bounds on error rates (false positives or false negatives) for any statistical test (Sanov's theorem). More specifically, for a fixed false positive rate, the false negative rate is bounded by $\exp(-nD(P_1\|P_2))$ for sample size n . In the typical setting of "multivariate normal, common covariance Σ , different means μ_1, μ_2 ," this quantity is given by $D(P_1\|P_2) = \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)$, a well known signal to noise ratio related to linear discriminant analysis.

To begin applying the powerful machinery of information theory to the study of anatomical form, we turn our attention to the quantity at the heart of information theory: the entropy. We propose a new method for quantifying the entropy of human anatomy at clinically relevant spatial resolutions, biological organization at the millimeter or *morphome* scale (Hunter and Borg, 2003; Crampin et al., 2004). In this work we focus our attention on developing this method and quantifying entropy for a single population, leaving inferences about specific populations or disease states to future work.

Since Shannon's original characterization of the entropy of natural language in the early 50's, the characterization of the combinatoric complexity of natural patterns such as human shape and form remains open. Human anatomical form, unlike word strings in English, are essentially continuum objects, extending all the way to the mesoscales of variation. Therefore, computing the entropy subject to a resolution, or measurement quantile becomes the natural approach to quantifying the complexity of human anatomy. Rate-distortion therefore plays a natural role. The distortion measure is played by the resolution, and in this paper we introduce the natural resolution metric that any anatomist or pathologist would use in examining tissue which would be the sup-norm distance in defining the boundary of an anatomical structure.

This paper focuses on these issues, calculating what we believe is the first bound on the complexity of human anatomy at the 1 mm scale. 1 mm seems appropriate since so much data is available via high throughput magnetic resonance imaging (MRI) and therefore that scale of data becomes ubiquitously available. Also so many studies of neuroanatomy and psychiatric disorders today are focused on the anatomical phenotype at this scale.

While the entropy of human anatomy seems difficult to define, the theory of Kolmogorov complexity gives us a precise tool for describing arbitrary objects in such a manner. The complexity of any object, which is related to its entropy by an additive constant, can be defined as the length of the shortest computer program that produces it as an output. As discussed in Cover and Thomas (2012), this quantity generally cannot be computed; doing so would be equivalent to solving the halting problem. However, any example of such a program serves as an upper bound on complexity. In what follows we describe our approach, which will serve as one such upper bound.

Our approach is to follow on Kolmogoroff's beautiful theory for calculating complexity of subcortical neuroanatomy by demonstrating codebooks that attain given logarithmic sizes coupled to a computer program which decodes elements of the codebook and attain the distortion measure. We also calculate

various rate-distortion curves showing the trade off in complexity as a function of distortion.

The field of computational anatomy (Miller et al., 2014) has been developing the random orbit model of human anatomy, where a given realization can be generated from a template (a typical example of an anatomical form) acted on by an element of the diffeomorphism group. Such diffeomorphic transformations can be generated from an initial momentum vector (i.e., closed under linear combinations) though geodesic shooting (Miller et al., 2006). Our work has largely focused on brain imaging and neurodegenerative diseases, and we therefore carry out an examination of subcortical gray matter structures. By using a sparse representation of initial momenta supported on anatomical boundaries, and learning Bayesian prior models for initial momenta from large populations (Tward et al., 2016), we can produce an efficient representation of anatomical form.

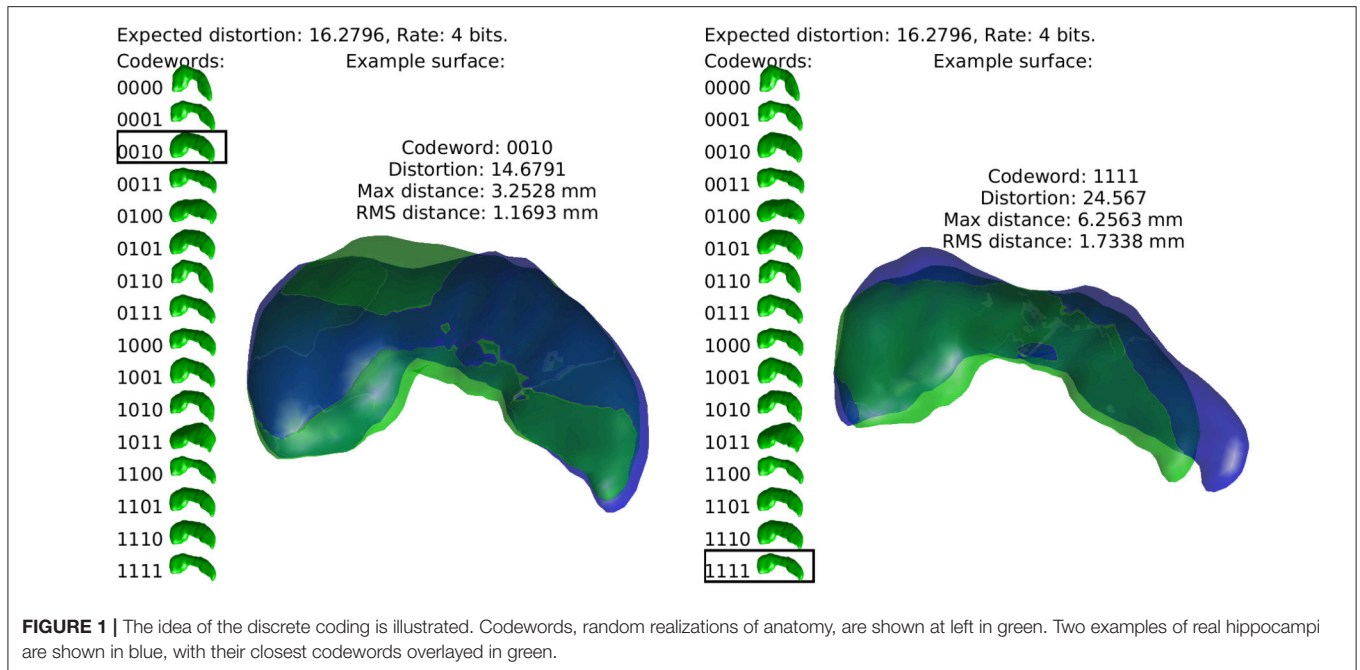
Our approach is to build sets of "codewords," specific examples of anatomical structures, and to encode a newly observed anatomy as one these words. This continuous to discrete process necessarily introduces distortion, and the relationship between the number of codewords required (the rate of our code) and this distortion measure is studied through rate distortion theory. By relating distortion to geometric error, we can establish the code rate required for errors at a certain spatial scale. This idea is illustrated in **Figure 1**, using a simple example of describing the hippocampus with a four bit code. In what follows we describe how this procedure is used to characterize the complexity of human anatomy at clinically relevant scales.

Much of the existing work in computational anatomy has focused on addressing the complexity of human anatomy through data reduction techniques. Foremost, the object of study was moved from high dimensional images to smooth diffeomorphisms via the random orbit model, with a fixed template (Miller et al., 1997) or several templates (Tang et al., 2013). Later, the construction of diffeomorphisms, typically created from a time varying velocity field, was moved to an initial velocity, with dynamics fixed via a conservation of momentum law (Miller et al., 2006). Sparsity was introduced, both optimized for specific data types (Miller et al., 2006), and for ease of interpretation and computational burden (Durrleman et al., 2014). Further, low dimensional models were developed based on empirical distributions such as PCA (Vaillant et al., 2004), or linear discriminant analysis (see Tang et al., 2014 for one example), or other techniques such as locally linear embedding (Yang et al., 2011). Instead of continuing the trend of *dimensionality reduction*, the novelty of this work is to address *discretization*. Our specific contribution is to develop a coding procedure informed by Bayesian priors, opening the study of anatomy through medical imaging to information theoretic techniques, and for the first time estimate the entropy of a population of neuroanatomy.

2. METHODS

2.1. Empirical Priors

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI)



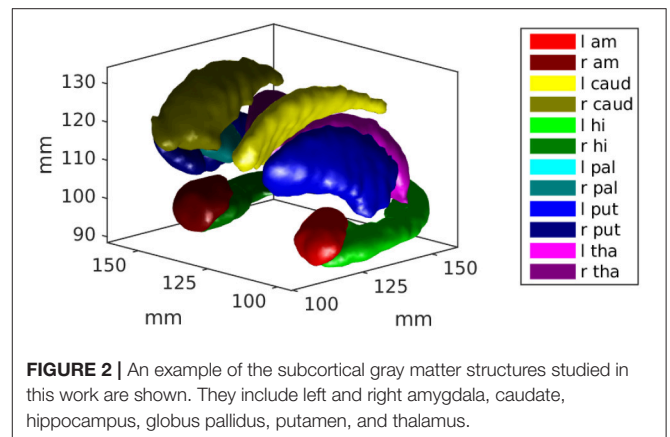
database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

Using 650 brains from the ADNI and the Open Access Series of Imaging Studies (OASIS), we extract 12 subcortical gray matter structures (left and right amygdala, caudate, hippocampus, globus pallidus, putamen, and thalamus) using FreeSurfer (Fischl et al., 2002) and create triangulated surfaces. For each structure, population surface templates were estimated following (Ma et al., 2010), and diffeomorphic mappings from template to each target were computed using current matching (Vaillant and Glaunès, 2005). The subcortical structure surface templates are shown in **Figure 2**.

These datasets were combined to provide a larger and more diverse sample. This is useful for achieving our goal of characterizing a population, as opposed to using more well controlled samples for hypothesis testing between populations.

As described in Miller et al. (2006), these diffeomorphic transformations are parameterized by an initial momentum vector, with three components per triangulated surface vertex at point $x^i \in \mathbb{R}^3$ denoted by p_0^i . This momentum defines a smooth velocity field v which is integrated over time to construct diffeomorphisms φ , as described by the following system of equations.

$$v(x) = \sum_i K(x - x^i) p^i \quad (2)$$



$$\dot{x}^i = v(x^i), \quad x_0 = \text{template} \quad (3)$$

$$\dot{p}^i = -Dv^T(x^i)p^i \quad (4)$$

$$\dot{\varphi} = v(\varphi), \quad \varphi_0 = \text{identity}, \quad (5)$$

where K is a Gaussian kernel of standard deviation 6.5 mm. The space of possible parameterizations is a vector space, in the sense that it is closed under scalar multiplication and addition. This substantial difference from the diffeomorphisms themselves, which are only closed under composition, allows us to study shape using multivariate Gaussian models.

The initial momentum vectors are analyzed using tangent space PCA as proposed in Vaillant et al. (2004), and described for this population in Tward et al. (2013). A low, B dimensional representation is chosen by selecting the largest principal components that account for 95% of the trace of the covariance

matrix. The low dimensional approximation of our initial momentum vector p_0 is written

$$p_0 = b^0 + \sum_{i=1}^B \beta^i b^i$$

where p_0, b^0, b^i are vectors of dimension three times the number of vertices, and β^i are scalar parameters. As described in the references, the basis vectors b^i are chosen to be orthonormal with respect to an inner product in the dual space of smooth functions, $\langle b^i, b^j \rangle = \sum_k b^{ikT} K(x_0^i, x_0^j) b^{jk} = \delta_{ij}$, where T denotes the transpose of a vector in \mathbb{R}^3 , and δ_{ij} is the Kronecker delta (1 if $i = j$ and 0 otherwise).

Our empirical prior model corresponds to choosing the β^i as independent Gaussian random variables with mean 0 and variance σ^{2i} , measured from the population. We create one empirical prior for each of the 12 subcortical structures examined.

2.2. Rate Distortion Theory for Multivariate Gaussians

For readers unfamiliar with rate distortion theory we review some standard terminology and results which will be necessary for our purposes. More details can be found in Cover and Thomas (2012).

Our empirical prior is a continuous distribution and must be discretized to be understood in terms entropy and complexity. This can be achieved through *encoding* our continuous random vectors β^i . That is, through constructing a mapping $e(\beta)$ from $\beta \in \mathbb{R}^B$ to a finite set S . Here S is chosen to be the set of binary strings of fixed length, as shown in the left side of each subfigure in **Figure 1**. Associated to this encoder is a *decoder*, a mapping $e(s)$ from $s \in S$ back to \mathbb{R}^B . Because S is finite, $d(e(\beta))$ can take only a finite number of values in \mathbb{R}^B , which we enumerate as $\hat{\beta}^i$ for positive integers i and refer to as *codewords*. The distribution of $d(e(\beta))$ is therefore a weighted sum of Dirac measures at these specific codewords $\hat{\beta}^i$. Examples of anatomies represented by a set of 16 codewords are shown toward the left side of each subfigure in **Figure 1**.

One can reason that an encoder/decoder pair is good if β is similar to $d(e(\beta))$ on average. The difference between the two is known as *distortion*. Because it admits well characterized solutions, we measure distortion using sum of square error in this work. Distortion can be minimized if we discretize β by mapping it to its closest codeword. In other words, we choose the encoder by

$$e(\beta) = s_i, \text{ the } i^{\text{th}} \text{ string in } S, \\ \text{where } i = \arg \min_j |\beta - \hat{\beta}^j|^2,$$

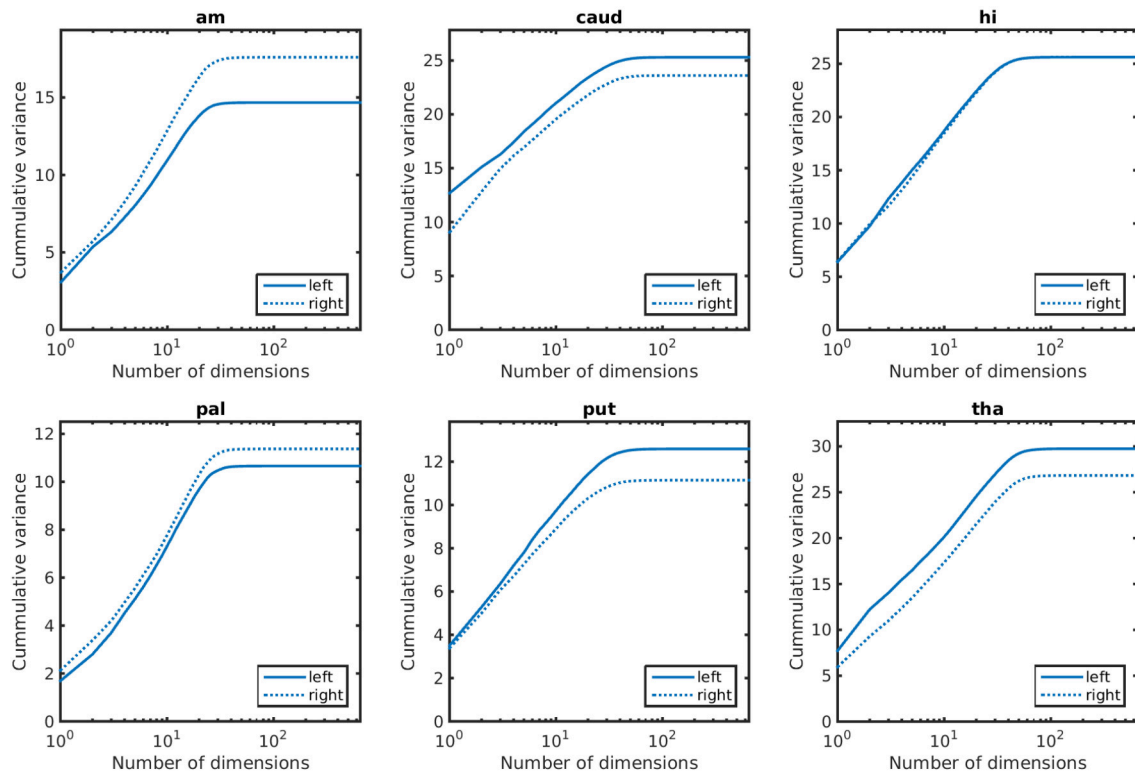


FIGURE 3 | Cumulative variance as a function of dimensions for anatomical priors. In lexicographic order: amygdala, caudate, hippocampus, globus pallidus, putamen, thalamus.

for $\|\cdot\|^2$ the norm squared in \mathbb{R}^B , and the decoder by

$$d(s_i) = \hat{\beta}^i.$$

Furthermore, one notices that lower distortion can be achieved with larger sets S . We refer to the size of S as $|S| = 2^R$ for a code rate R . We note that R is the length of the binary strings in S , so that the examples in **Figure 1** have a rate of $R = 4$ bits.

We aim to identify the minimum number of codewords that are required to achieve a given amount of expected distortion D . The best achievable code is characterized by the *rate distortion curve* (D as a function of R). This can be shown to be equal to the minimum of the mutual information between β and $d(e(\beta))$ while enforcing distortion less than or equal to D (i.e., the shortest code respecting the distortion constraints is the worst one: that with the smallest mutual information with β). This definition, while arcane, can be used to compute rate distortion curves in closed form in several situations. In general this curve can be approached asymptotically, by coding blocks of N structures simultaneously using 2^{NR} codewords, considering the average distortion, and letting $N \rightarrow \infty$.

The details of Gaussian rate distortion curves can be found in Cover and Thomas (2012) chapter 13. For single variate Gaussian

random variables with square error distortion the rate distortion curve can be computed in closed form:

$$R(D) = \begin{cases} \frac{1}{2} \log_2 \frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

Note that if the desired distortion is greater than the variance, we need only 1 codeword, or $R = 0$. If this 1 codeword is the mean, the expected distortion is equal to the variance. Otherwise, we require more codewords in a manner increasing logarithmically with the variance.

We finally specify how our codewords are chosen. This minimal distortion can be achieved for codewords chosen as independent realizations of a Gaussian random variable. We can motivate this as follows. Let the joint distribution of data β and codewords $\hat{\beta}$ be described by drawing β from the distribution $\hat{\beta} \sim \mathcal{N}(0, \sigma^2 - D)$, and $\beta = \hat{\beta} + \text{err}$ with error $\text{err} \sim \mathcal{N}(0, D)$. This coding scheme has square error distortion at most D . The mutual information between β and $\hat{\beta}$ can be calculated as $\frac{1}{2} \log \frac{\sigma^2}{D}$, the value of the rate distortion curve. On the other hand, if the allowable distortion $D > \sigma^2$, we can simply choose $\hat{\beta} = 0$ and achieve $R(D) = 0$.

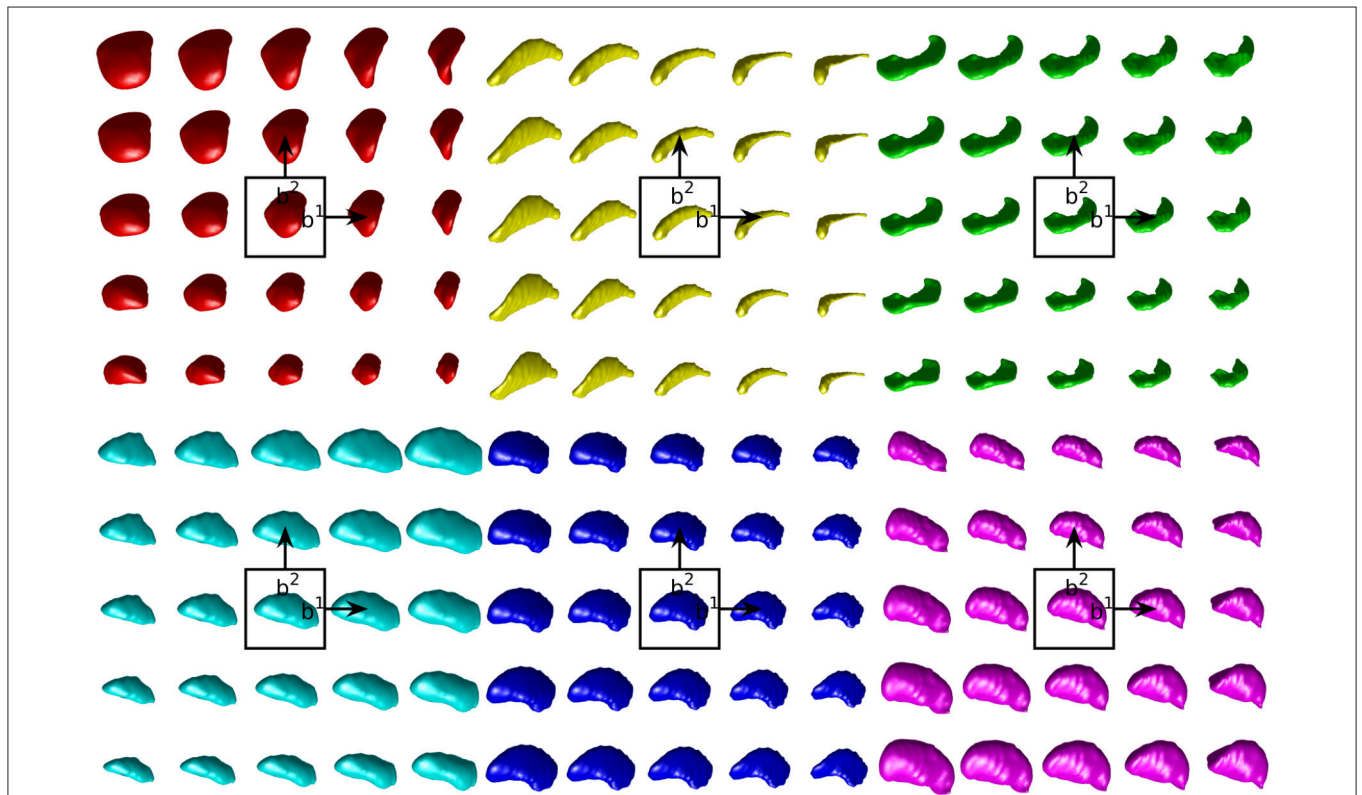


FIGURE 4 | Examples of the first two modes of variability in our empirical prior for left side structures. The mean shape is shown in the center. Each step to the right (top) moves one standard deviation in the direction of the first (second) mode of variation. In lexicographic order: amygdala, caudate, hippocampus, globus pallidus, putamen, thalamus.

This approach can be extended to B independent Gaussians using the *reverse water filling* method.

$$D_i = \begin{cases} \lambda, & \lambda < \sigma_i^2 \\ \sigma_i^2, & \lambda \geq \sigma_i^2 \end{cases}$$

$$\lambda \text{ s.t. } \sum_{i=1}^B D_i = D$$

The optimum corresponds to choosing a fixed amount of distortion per dimension for variables with “large” variance ($\sigma_i^2 > \lambda$), and no additional codewords for those of “small” variance.

This leads to the rate distortion curve

$$R(D) = \sum_{i=1}^B \frac{1}{2} \log \frac{\sigma_i^2}{D_i} \quad (6)$$

which can be asymptotically approached (coding blocks of N anatomies simultaneously, and allowing $N \rightarrow \infty$) with a random code, with the i th component of a codeword generated according to

$$\hat{\beta}^i \sim \begin{cases} \mathcal{N}(0, \sigma_i^2 - \lambda), & \sigma_i^2 \geq \lambda \\ \mathcal{N}(0, 0), & \sigma_i^2 < \lambda \end{cases}$$

The reverse waterfilling method is named by imagining each independent Gaussian to be represented by an object of height σ_i^2 in a room with rising water. As the water rises, those Gaussians with small variance become submerged. Everything below the surface represents distortion, a fixed amount for each of the variables with large variance, and amount equal to its variance for the others. We allow the water to continue to rise until the total distortion is given by D .

For our experiments, from the empirical prior for each subcortical structure a set of codewords is generated for rates from 0 to 32 bits, and for coding $N = 1$ and $N = 2$ examples simultaneously.

2.3. Complexity at Clinically Relevant Spatial Scales

By shooting our template with the initial momentum from a given codeword, we can compute the expected geometric error between an anatomical structure defined by our continuous model and its discretely coded version. Error in units of mm are considered, using Hausdorff distance between surfaces (max error between closest pairs of vertices between realization and codeword). We measure geometric error as a function of rate, fit this curve to a simple model, and compute the code rate required at clinically relevant scales. Owing to the computational

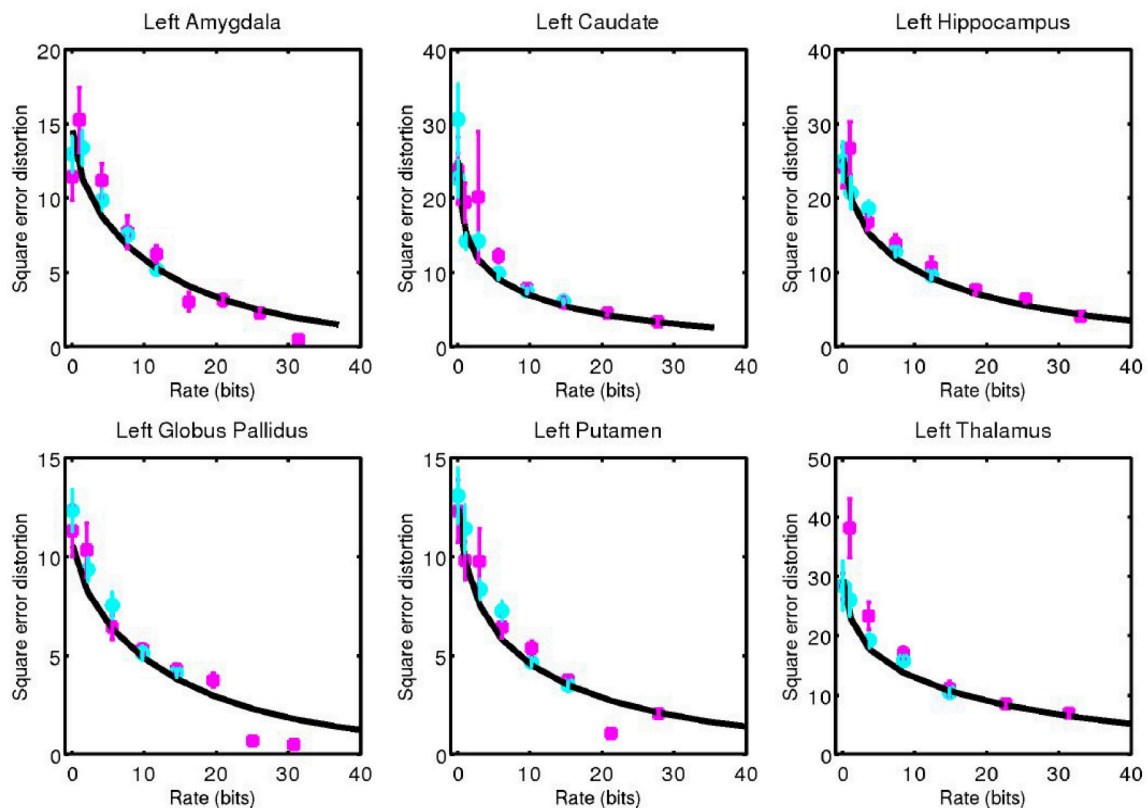


FIGURE 5 | Square error distortion as a function of code rate for left side structures. Coding one structure is shown in magenta, and two structures simultaneously is shown in cyan. The rate distortion curve for a multivariate Gaussian model is shown in black.

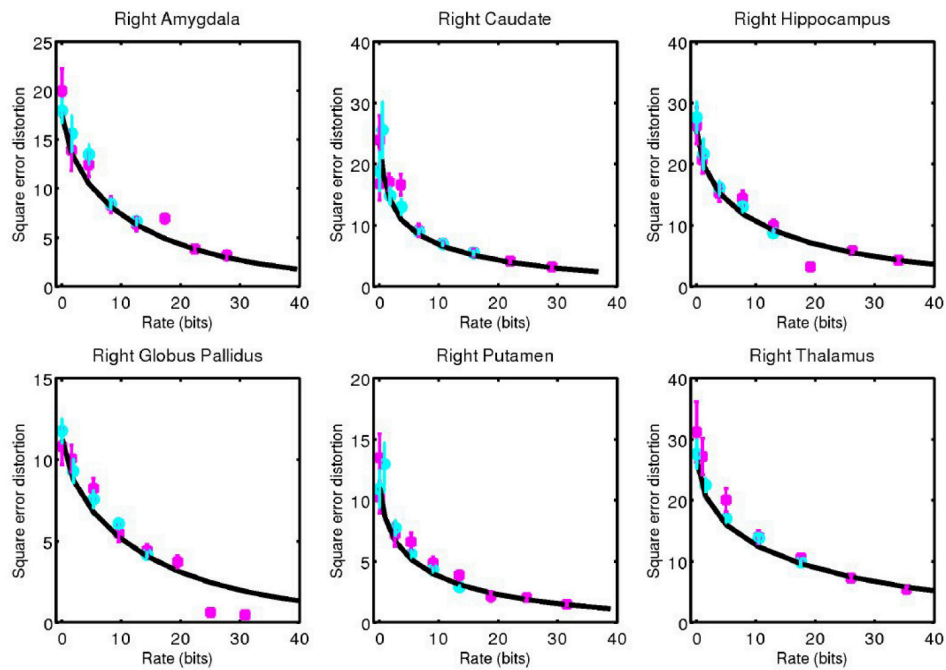


FIGURE 6 | Corresponding data from **Figure 5** for right side structures.

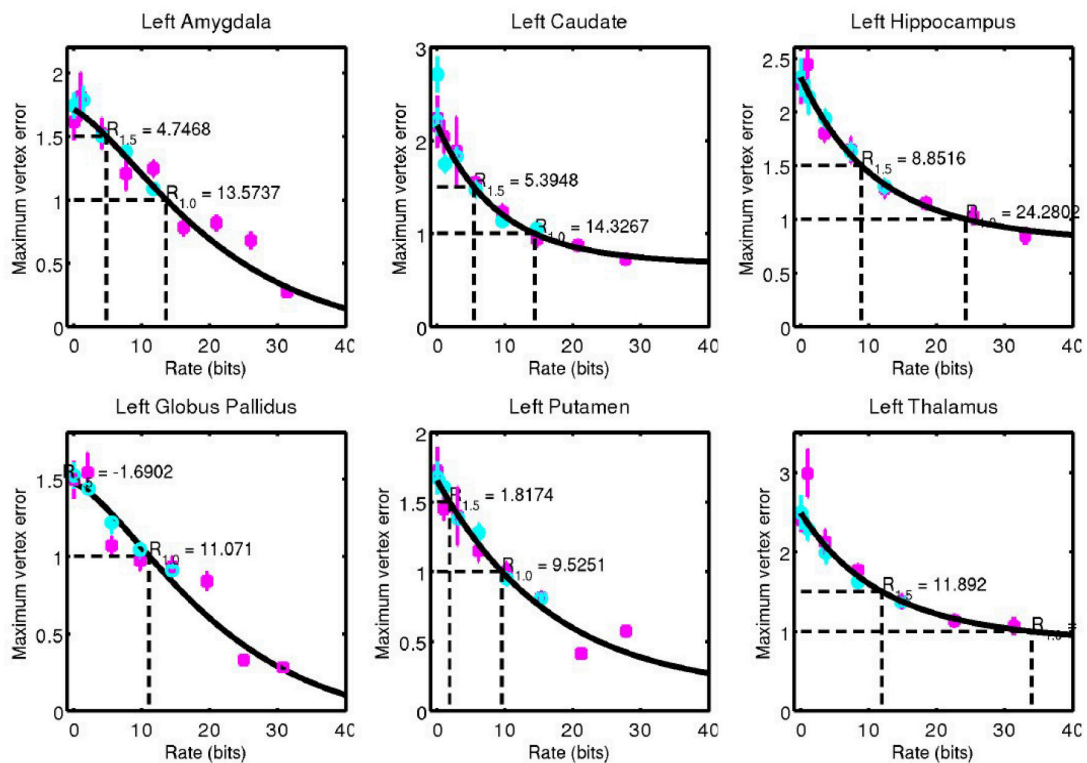


FIGURE 7 | Hausdorff distance between example surfaces and closest codeword. Coding one structure is shown in magenta, and two structures simultaneously is shown in cyan. The black curve is a simple fit through the data (not a model), and is used for estimating code rate at 1 and 1.5 mm geometric error.

complexity of looping through 2^{32} codewords and solving system Equation (2), this procedure is repeated for 10 observations of each subcortical structure.

3. RESULTS

3.1. Empirical Priors

Empirical prior models for the 6 structures examined are quantified in terms of their variance spectra in **Figure 3**. The number of dimensions that captured 95% of the trace of the covariance matrix for each left (right) structure was found to be: amygdala 21 (22), caudate 26 (26), hippocampus 31 (32), globus pallidus 24 (24), putamen 27 (25), thalamus 39 (41). These numbers are quite similar for the left and right hand sides of the same structure. Examples of the first two modes of variability are shown for the left side structures in **Figure 4**.

3.2. Rate Distortion Calculations

For each subcortical structure we calculate square error distortion as a function of code rate. For coding one structure at a time, we use codes with rate from 0 to 32 bits. For coding two structures at a time, we use codes with rate from 0 to 16 bits. The results of these calculations are shown for left side structures in **Figure 5** and for right side structures in **Figure 6**. Mean and standard error for coding one structure is shown in magenta, and that for two structures simultaneously

is shown in cyan. The two results are seen to be similar, indicating that not much is gained by encoding several structures simultaneously, since the coefficients β are already high (as compared to 1) dimensional. For each structure, we calculate the rate distortion curve described by Equation (6) from the corresponding multivariate Gaussian. This represents a lower bound on the expected value of the data shown. That our data is close to these curves serves as an indication that our procedure is valid.

3.3. Complexity at Clinical Scale

For each structure examined, we consider the geometric error between our codeword and the anatomy they represent. We quantified this through the Hausdorff distance between triangulated surfaces. Mean and standard error of this data is shown for left side structures in **Figure 7** and for right side structures in **Figure 8**.

A simple curve was fit through the data and used to estimate the code rate required for 1 and 1.5 mm of maximum error, values that are on the order of 1 voxel in a typical clinical MRI. These rates are shown in **Figure 9**.

4. CONCLUSION

The complexity of the subcortical gray matter structures we have examined range from the order of 5–35 bits for 1.0 mm geometric

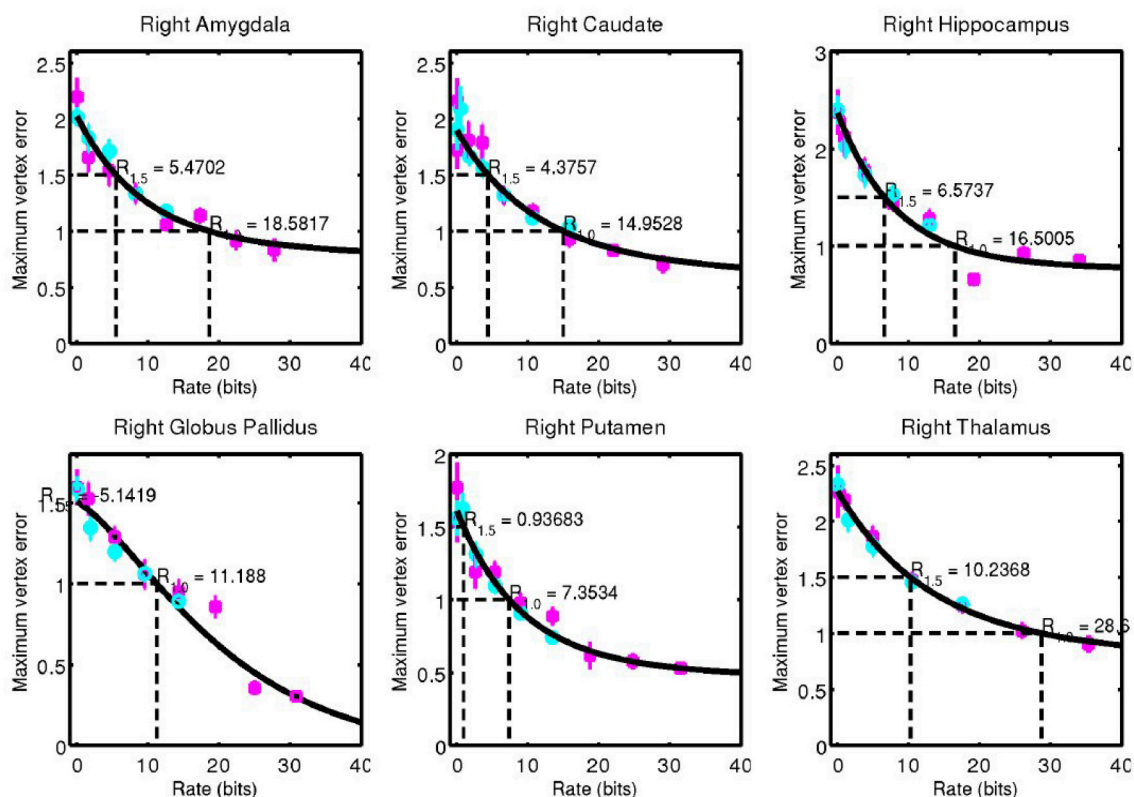
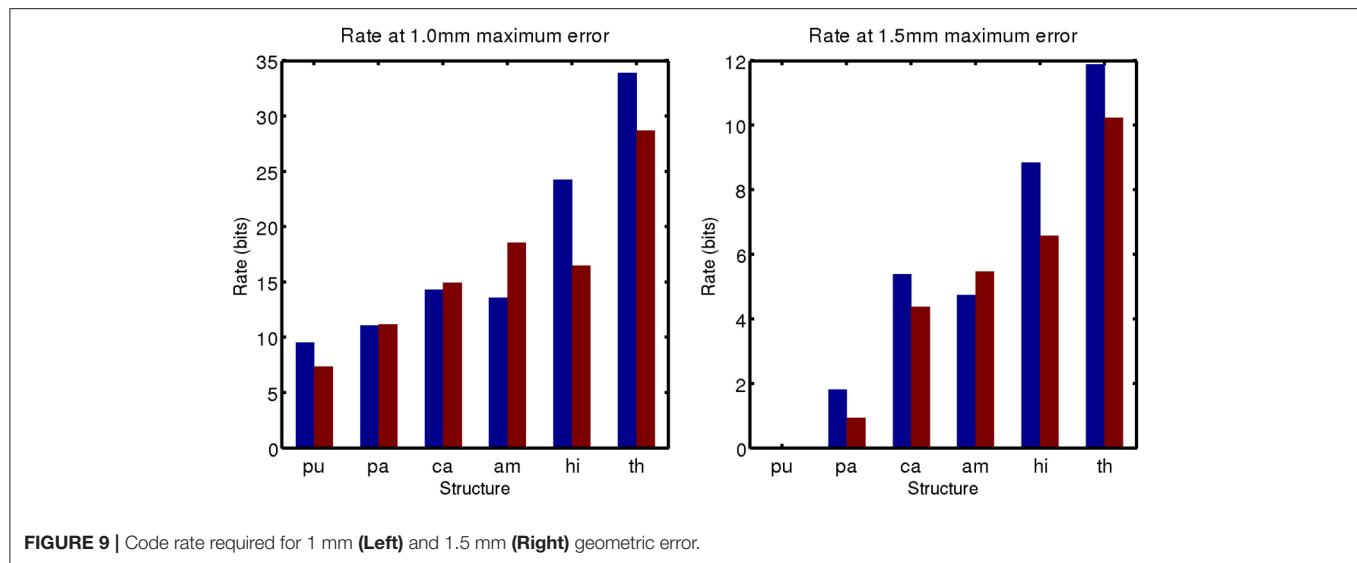


FIGURE 8 | Corresponding data from **Figure 7** for right side structures.



error, and 0–12 bits for 1.5 mm geometric error. Note that at 1.5 mm error, a 0 bit code is sufficient for the putamen. Its low amount of variability means it can be represented by an average template only at this accuracy.

While using up to 2^{32} , or more than 4 billion, codewords may seem excessive, this still represents a huge amount of data compression. Binary segmentation images, contain roughly 100^3 voxels, or the order of one million bits. The triangulated surfaces have roughly 1,000 vertices, each component stored to double precision, which correspond to about 192,000 bits. We have shown that 32 bits, or an amount of data equivalent to one single precision floating point number, is enough to encode the variability of gray matter subcortical structures at clinically relevant spatial scales.

The potential for this work to impact clinical practice stems from the fact that entropy can be used to devise lower bounds on the variance of estimators, and that information can be used as an important figure of merit. When this work is extended to considering mutual information between anatomical form and diagnostic status, it could directly influence clinical decision making and optimization of imaging procedures.

For example, the Image Gently campaign (Goske et al., 2008), a program designed to reduce radiation exposure to pediatric patients, suggests first to “reduce or ‘child-size’ the amount of radiation used” and second to “scan only when necessary” through a discussion of a risk-benefit ratio. Because lower radiation doses can be used at lower resolution, the analysis presented as a function of resolution could lead to appropriately choosing a dose level for a given level of certainty required. Further, a scan could be avoided if it will not reduce entropy about diagnostic status sufficiently.

Turning to imaging optimization, task based analysis of image quality (Sharp et al., 1996) has been used for many years, but figures of merit have been largely designed to reflect the performance of idealized observers on simple detection or estimation tasks (Barrett et al., 1995). Anatomical variability is often described simply as stationary power law noise (see for example Burgess, 1999). Mutual information between observed

anatomy and diagnostic status could be used as a figure of merit for system design that appropriately accounts for anatomical variation and models realistic imaging tasks.

One limitation of this study is that we have encoded only a small number of structures. Due to the computational complexity of searching through each codeword and solving a high dimensional geodesic shooting equation in each case, we limited the number examined. As this work progresses, we will include larger samples. In what follows, we will restrict ourselves to disease specific populations to measure how entropy changes with disease state. This will enable calculation of the mutual information between anatomical phenotype and disease state as shown in Equation (1).

AUTHOR CONTRIBUTIONS

DT and MM developed the approach and planned experiments. DT developed tools for computational analysis.

FUNDING

This work was supported by the Kavli Neuroscience Discovery Institute. This work was supported by the National Institute of Health through grant numbers P41-EB015909, R01-EB020062, and U19-AG033655. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (Towns et al., 2014), which is supported by National Science Foundation grant number ACI-1053575. Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company;

EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org).

REFERENCES

- Barrett, H. H., Denny, J., Wagner, R. F., and Myers, K. J. (1995). Objective assessment of image quality. ii. fisher information, fourier crosstalk, and figures of merit for task performance. *JOSA A* 12, 834–852. doi: 10.1364/JOSAA.12.000834
- Burgess, A. E. (1999). Mammographic structure: data preparation and spatial statistics analysis. *Proc. SPIE*, 3661, 642–653. doi: 10.1117/12.348620
- Cover, T. M., and Thomas, J. A. (2012). *Elements of information theory*. Hoboken, NJ: John Wiley & Sons.
- Crampin, E. J., Halstead, M., Hunter, P., Nielsen, P., Noble, D., Smith, N., et al. (2004). Computational physiology and the physiome project. *Exp. Physiol.* 89, 1–26. doi: 10.1113/expphysiol.2003.026740
- Durrleman, S., Prastawa, M., Charon, N., Korenberg, J. R., Joshi, S., Gerig, G., et al. (2014). Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *Neuroimage* 101, 35–49. doi: 10.1016/j.neuroimage.2014.06.043
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/S0896-6273(02)00569-X
- Goske, M. J., Applegate, K. E., Boylan, J., Butler, P. F., Callahan, M. J., Coley, B. D., et al. (2008). The image gently campaign: increasing ct radiation dose awareness through a national education and awareness program. *Pediat. Radiol.* 38, 265–269. doi: 10.1007/s00247-007-0743-3
- Hunter, P. J., and Borg, T. K. (2003). Integration from proteins to organs: the physiome project. *Nat. Rev. Mol. Cell Biol.* 4:237. doi: 10.1038/nrm1054
- Ma, J., Miller, M. I., and Younes, L. (2010). A bayesian generative model for surface template estimation. *J. Biomed. Imaging* 2010:16. doi: 10.1155/2010/974957
- Miller, M., Banerjee, A., Christensen, G., Joshi, S., Khaneja, N., Grenander, U., et al. (1997). Statistical methods in computational anatomy. *Statist. Methods Med. Res.* 6, 267–299. doi: 10.1177/096228029700600305
- Miller, M. I., Trounev, A., and Younes, L. (2006). Geodesic shooting for computational anatomy. *J. Math. Imag. vis.* 24, 209–228. doi: 10.1007/s10851-005-3624-0
- Miller, M. I., Younes, L., and Trounev, A. (2014). Diffeomorphometry and geodesic positioning systems for human anatomy. *Technology* 2, 36–43. doi: 10.1142/S2339547814500010
- Sharp, P., Metz, C., Wagner, R., Myers, K., and Burgess, A. (1996). “Medical imaging: The assessment of image quality,” in *International Commission on Radiation Units and Measurement* (Bethesda, MD: ICRU Report), 54.
- Tang, X., Holland, D., Dale, A. M., Younes, L., and Miller, M. I. (2014). Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer’s disease: detecting, quantifying, and predicting. *Hum. Brain Mapp.* 35, 3701–3725. doi: 10.1002/hbm.22431
- Tang, X., Oishi, K., Faria, A. V., Hillis, A. E., Albert, M. S., Mori, S., et al. (2013). Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. *PLoS ONE* 8:e65591. doi: 10.1371/journal.pone.0065591
- Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., et al. (2014). Xsede: accelerating scientific discovery. *Comput. Sci. Eng.* 16, 62–74. doi: 10.1109/MCSE.2014.80
- Tward, D., Miller, M., Trounev, A., and Younes, L. (2016). Parametric surface diffeomorphometry for low dimensional embeddings of dense segmentations and imagery. *IEEE Trans. Pattern Anal. Mach. Intel.* 39, 1195–1208. doi: 10.1109/TPAMI.2016.2578317
- Tward, D. J., Ma, J., Miller, M. I., and Younes, L. (2013). Robust diffeomorphic mapping via geodesically controlled active shapes. *J. Biomed. Imag.* 2013:3. doi: 10.1155/2013/205494
- Vaillant, M., and Glaunès, J. (2005). “Surface matching via currents,” in *Biennial International Conference on Information Processing in Medical Imaging* (Glenwood Springs, CO: Springer), 381–392.
- Vaillant, M., Miller, M. I., Younes, L., and Trounev, A. (2004). Statistics on diffeomorphisms via tangent space representations. *Neuroimage* 23, S161–S169. doi: 10.1016/j.neuroimage.2004.07.023
- Yang, X., Goh, A., and Qiu, A. (2011). Locally linear diffeomorphic metric embedding (lldme) for surface-based anatomical shape modeling. *Neuroimage* 56, 149–161. doi: 10.1016/j.neuroimage.2011.01.069

Conflict of Interest Statement: MM reports personal fees from AnatomyWorks, LLC, outside the submitted work and jointly owns AnatomyWorks. This arrangement is being managed by the Johns Hopkins University in accordance with its conflict of interest policies. MM’s relationship with AnatomyWorks is being handled under full disclosure by the Johns Hopkins University.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Tward and Miller for the Alzheimer’s Disease Neuroimaging Initiative. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Brain MRI Pattern Recognition Translated to Clinical Scenarios

Andreia V. Faria^{1*}, Zifei Liang², Michael I. Miller³ and Susumu Mori¹

¹ Department of Radiology, Johns Hopkins University, Baltimore, MD, United States, ² Department of Radiology, New York University, New York, NY, United States, ³ Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Jennifer L. Robinson,
Auburn University, United States

Reviewed by:

Jingyu Liu,
Mind Research Network,
United States
Hyunjin Park,
Sungkyunkwan University,
South Korea

*Correspondence:

Andreia V. Faria
afaria1@jhmi.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 29 June 2017

Accepted: 02 October 2017

Published: 20 October 2017

Citation:

Faria AV, Liang Z, Miller MI and Mori S
(2017) Brain MRI Pattern Recognition
Translated to Clinical Scenarios.
Front. Neurosci. 11:578.
doi: 10.3389/fnins.2017.00578

We explored the performance of structure-based computational analysis in four neurodegenerative conditions [Ataxia (AT, $n = 16$), Huntington's Disease (HD, $n = 52$), Alzheimer's Disease (AD, $n = 66$), and Primary Progressive Aphasia (PPA, $n = 50$)], all characterized by brain atrophy. The independent variables were the volumes of 283 anatomical areas, derived from automated segmentation of T1-high resolution brain MRIs. The segmentation based volumetric quantification reduces image dimensionality from the voxel level [on the order of $\mathcal{O}(10^6)$] to anatomical structures [$\mathcal{O}(10^2)$] for subsequent statistical analysis. We evaluated the effectiveness of this approach on extracting anatomical features, already described by human experience and a priori biological knowledge, in specific scenarios: (1) when pathologies were relatively homogeneous, with evident image alterations (e.g., AT); (2) when the time course was highly correlated with the anatomical changes (e.g., HD), an analogy for prediction; (3) when the pathology embraced heterogeneous phenotypes (e.g., AD) so the classification was less efficient but, in compensation, anatomical and clinical information were less redundant; and (4) when the entity was composed of multiple subgroups that had some degree of anatomical representation (e.g., PPA), showing the potential of this method for the clustering of more homogeneous phenotypes that can be of clinical importance. Using the structure-based quantification and simple linear classifiers (partial least square), we achieve 87.5 and 73% of accuracy on differentiating AT and pre-symptomatic HD patients from controls, respectively. More importantly, the anatomical features automatically revealed by the classifiers agreed with the patterns previously described on these pathologies. The accuracy was lower (68%) on differentiating AD from controls, as AD does not display a clear anatomical phenotype. On the other hand, the method identified PPA clinical phenotypes and their respective anatomical signatures. Although most of the data are presented here as proof of concept in simulated clinical scenarios, structure-based analysis was potentially effective in characterizing phenotypes, retrieving relevant anatomical features, predicting prognosis, and aiding diagnosis, with the advantage of being easily translatable to clinics and understandable biologically.

Keywords: precision medicine, pattern recognition, quantitative MRI, computer aid, automated MRI analysis

INTRODUCTION

A longtime dream of clinicians is to use computational tools for aiding decisions. Like using the spelling and grammar checkers when writing a text or Google for searching, clinical computational tools would neither define purposes nor change goals, but add a higher level of quality and speed to the results. There are three must-haves for computational-aid tools: speed, automation, and, of course, efficacy. The development of such tools for medical records and imaging, in particular, is extremely complex, involving knowledge in multiple domains. Consequently, more than two decades after the initial attempts (for review and perspectives see Orphanoudakis et al., 1996; Akgul et al., 2011; Hwang et al., 2012; Kalpathy-Cramer et al., 2015; Pinho et al., 2017; Spanier et al., 2017), no system is yet adequately suited for practical daily use. The key to translating the computational models to radiological practice is to resolve the so-called semantic gap: “the differences between image similarity on the high level of human perception and the low level of a few numbers” (Depeursinge et al., 2011). Three basic steps are involved: precise quantification, optimal feature selection and combination, and, eventually, meaningful applications and testing.

The first step, image quantification, is straightforward if one is simply interested in the intensity of a given voxel. What is not simple, however, is to extract some biological meaning from the noisy voxel-by-voxel information, which can be of the order of 10^6 , considering only T1-weighted images, one of the multiple MRI contrasts. There are numerous papers on voxel-based analysis (VBA) in which human involvement is eliminated on the assumption that a human being’s ability to detect abnormality is neither sensitive nor reliable. A PubMed search for “VBA,” “brain,” and “MRI” results in more than 2,300 publications in the last 10 years. These studies provide a wealth of descriptive imaging results that are usually not perceptible at an individual level and fail to be translated to clinical practice, which meanwhile, remains supported by human judgment. If we flip this approach 180° by asking: Can a computational approach describe abnormalities that agree with human perception?, we find the number of publications to be surprisingly small. A PubMed search for “structure-based analysis” or “atlas-based analysis,” “brain,” and “MRI” results in fewer than 200 publications in the last 10 years. An old strategy to replicate human perception is to group voxels in regions of interest (ROIs) and label them according to existing anatomical knowledge. For example, all the voxels associated with certain x, y, z coordinates are called “thalamus,” or “frontal lobe,” or “internal capsule,” and so on. This is what radiologists do, increasing the signal-to-noise ratio and adding a biological domain to their subjective analysis. However, objectively quantifying, structurizing, and recording the information for subsequent use is much more complex. In addition, defining ROIs in multiple subjects multidimensionally is just not feasible; precise automated tools are vital.

This structure-based analysis is linked to the second step to solve the semantic gap: the feature selection and combination. Here, two components are essential: the existing knowledge of normal and abnormal patterns and the ability to recognize these

patterns in future patients. For example, when a patient has striatum atrophy and motor disabilities, Huntington’s Disease (HD) is a possible diagnosis because physicians learned that these two features are associated with this disease. In addition to centuries of pathophysiological knowledge, what is hidden behind this apparently simple conclusion is an enormous amount of comprehension about normal variation. In order to conclude that those regions, in an individual of a certain age and gender, are smaller than expected, an analysis of multiple granularity levels (looking to the caudate, or the basal ganglia, or the deep gray matter, or the lobe, or the whole brain), and multiple image domains (volume, intensity, shapes), and finally the combination of features in different fields (clinical and imaging) are necessary. This leads to the amazing capability of pattern recognition that humans have and that machine-learning methods try to replicate.

Finally, even if we are able to quantify structures precisely in different levels and domains, to compare individual cases with large and variable normal and pathological databases, and to extract and combine important features efficiently, we still have to suit the computer-aid tools to the appropriate applications and test them. If the goal is a diagnostic-aid tool, this may be the most challenging step because the gold-standard is the clinical diagnosis, which does not necessarily reflect the actual situation. In addition, the correlation between pathology and anatomy may be weak or indirect. This is usually the case in pathologies in which the anatomical changes are subtle or happen later, or when the time course is unknown, or in those that embrace heterogeneous phenotypes. These cases are challenging and may reduce the efficiency of classification models, but they also offer an opportunity to design tools for binning a given entity into subgroups, for example, that may be of clinical relevance.

Previously, our group and others advanced in the first two steps (quantification and feature extraction). The brain quantification and segmentation accuracy improved drastically in this decade due to the advances in multi-atlas technologies (Warfield et al., 2004; Artaechevarria et al., 2009; Langerak et al., 2010; Lotjonen et al., 2010; Sabuncu et al., 2010; van Rikxoort et al., 2010; Jia et al., 2012; Wang et al., 2013), allowing use of state-of-the-art techniques for quantification and extraction of clinically meaningful image features. We confirmed the accuracy of these techniques in different populations and protocols (Liang et al., 2015). We then tested whether the structured anatomical data extracted actually captured the anatomical features that can be perceived by trained clinicians (Faria et al., 2015). In the present study, we advance to the next step and report progress on feature selection, combination, and classification, showing the potential of structure-based analysis for computer-aided decisions.

This study focused on the brain MRIs of patients with these neurodegenerative conditions: Ataxia (AT), Huntington’s Disease (HD), Alzheimer’s Disease (AD), and Primary Progressive Aphasia (PPA). Briefly, Ataxia, or more specifically, the Spinocerebellar ataxia type 6 (SCA-6) which is considered here, is an autosomal dominant disorder that is characterized by a slowly progressive cerebellar ataxia, dysarthria and nystagmus (Zhuchenko et al., 1997). The cerebellar atrophy, demonstrated

by several prior MRI studies, is a constant (Butteriss et al., 2005) and relates with clinical manifestations (Eichler et al., 2011).

HD is a progressive lethal neurodegenerative disorder characterized by movement disorders and progressive cognitive and psychological manifestations (Huntington, 1872). The anatomical hallmark of HD is striatal atrophy. Although the atrophy may start as early as 15 years before the onset of motor symptoms, and continue through the pre-manifest period (Tabrizi et al., 2009, 2012, 2013; Paulsen et al., 2014a,b), it is mostly undetectable by clinical evaluation of MRIs, at individual level, in pre-symptomatic patients. The early quantitative characterization of the atrophy, both at group and individual level, is an important piece of information for the development of disease-modifying treatments (Faria et al., 2016; Wu et al., 2016).

Alzheimer disease (AD) is a chronic neurodegenerative disease characterized by short-term memory loss in the early disease stages and progressive cognitive and functional deficits as the disease advances. It is actually not a single disease but a clinically, anatomically and biologically heterogeneous disorder encompassing a wide spectrum of cognitive and anatomical profiles (Zhang et al., 2016). Although a classical pattern of atrophy is reported for AD as a group, first noticeable in the medial temporal lobe (including hippocampus and entorhinal cortex), eventually spreading through the remainder of the brain (Apostolova et al., 2007), this pattern is not highly discriminant at individual level (Frisoni et al., 2017). In addition, the atrophy is usually clinically evident long after the cognitive deficits start. The heterogeneity of phenotypes and subtleness of early anatomical changes are extra challenges for the development of therapeutics and prognostic models.

Primary progressive aphasia (PPA) is a clinical syndrome characterized by insidious progressive language impairment that is initially unaccompanied by other cognitive deficits (Mesulam, 1982). It is caused by various neurodegenerative diseases and has a highly variable course. There are three main variants that are distinguished by their key features and supporting brain imaging characteristics, which are generally associated with distinct underlying pathologies (Gorno-Tempini et al., 2011): agrammatic (Av) is supported by left posterior frontal and (Zhuchenko et al., 1997) insular atrophy; semantic (Sv) is associated with left greater than right anterior and inferior temporal atrophy; logopenic (Lv) is associated with posterior temporal and inferior parietal atrophy (Rohrer and Rosen, 2013; Wilson et al., 2016). The identification of the variant provides some clues regarding the subsequent course (Leyton et al., 2016), and would be of great value for prognosis in the initial stages. However, the early classification is particularly challenging because the clinical deficits are common to all three variants and the anatomical changes are still clinically silent. Methods for phenotypically characterization, particularly at early phases, would be of great assistance.

The choice of these clinical entities was due to the fact that the common feature (atrophy) varies in extension and location, providing an appropriate dynamic range of abnormalities. In addition, the atrophy is mostly visible, which enables validation by qualitative human evaluation. The overall goal of this study was to test the performance of structure-based computational analysis on extracting anatomical features, already described by

human experience and a priori biological knowledge, in specific patient populations. The variables in question were the volumes of 283 structures. We showed the potential of the structure-based analysis on characterization and classification (1) when pathologies were relatively homogeneous, with evident image alterations (e.g., Ataxias); (2) when the time course was highly correlated with the anatomical changes (e.g., HD), an analogy for prediction; (3) when the pathology embraced heterogeneous phenotypes (e.g., AD) so the classification was less efficient but, in compensation, anatomical and clinical information were less redundant; and (4) when the entity was composed of multiple subgroups that had some degree of anatomical representation (e.g., Primary Progressive Aphasia), showing the potential of this method for the clustering of more homogeneous phenotypes that can be of clinical importance.

MATERIALS AND METHODS

Database

The overall goal was to test the performance of structure-based computational analysis in extracting anatomical features, previously described by human experience and a priori biological knowledge, in specific patient populations.

The data consisted of high-resolution T1-weighted brain MRIs (MPRAGE), for five groups of individuals: healthy individuals (controls, $n = 208$), AT ($n = 16$), HD ($n = 52$), AD ($n = 66$), and PPA ($n = 50$) (Table 1). The data from healthy individuals (controls) were obtained from three sources: (1) internal datasets from Johns Hopkins University (JHU), (2) International Consortium for Brain Mapping (ICBM, loni.usc.edu/ICBM), and (3) the AD Neuroimaging Initiative (ADNI, adni.loni.usc.edu). The control dataset included more than 10 different protocols (including different machine manufacturers, strength of magnetic field, and resolution), thus replicating the heterogeneity encountered in clinical scenarios. Individuals with AT were from JHU and had spinocerebellar ataxia type 6 (SAC6). Individuals with HD, also from JHU, were grouped into three different stages, according to their CAG-Age Product (CAP) scores (Penney et al., 1997) and clinical symptoms: pre-symptomatic far from onset ($n = 23$), pre-symptomatic close to onset ($n = 16$), and early symptomatic ($n = 13$). Individuals with AD, from JHU and ADNI, were diagnosed according to new clinical guidelines (Albert et al., 2011; Jack et al., 2011; McKhann et al., 2011; Sperling et al., 2011). Individuals with PPA, from JHU, were diagnosed and classified into three variants: logopenic (Lv, $n = 18$), semantic (Sv, $n = 16$), and agrammatic (Av, $n = 16$), based on current clinical guidelines (Mesulam, 1982; Gorno-Tempini et al., 2011). All the data had previously been de-identified, and the participants consented to enrolling by written consent.

Image Processing

In the present study, quantification of regional brain volume was performed on a structural level, which involved the mapping of each brain to 29 templates in which the structures in question had previously been labeled. The brain mapping was performed with large deformation diffeomorphic metric mapping (LDDMM)

TABLE 1 | Demographic and protocol information.

Group	Sample	Age range (Years)	Mean age \pm St.dev. (Years)	Male/Female	Protocols (Manufacturer, Field Strength (T), Voxel Size (mm), Sample)
Controls	208	20–95	57.9 \pm 18.8	98/110	Phillips, 1.5, 1 \times 0.875 \times 0.875, 28 Phillips, 1.5, 1.2 \times 0.94 \times 0.94, 9 Phillips, 3, 1 \times 1 \times 1, 7 Phillips, 3, 1.2 \times 1 \times 1, 26 Phillips, 3, 0.9 \times 0.9 \times 0.9, 52 Phillips, 3, 1.1 \times 0.83 \times 0.83, 16 GE, 1.5, 1.2 \times 0.94 \times 0.94, 6 GE, 3, 1.2 \times 1.02 \times 1.02, 6 Siemens, 1.5, 1.2 \times 1.25 \times 1.25, 7 Siemens, 3, 1 \times 1 \times 1, 28 Siemens, 3, 1.2 \times 1 \times 1, 23
Ataxia	16	48–73	60.8 \pm 6.8	13/3	Phillips, 3, 1.1 \times 0.83 \times 0.83, 16
HD					
Far from onset	23	21–51	36.8 \pm 9.7	10/13	
Near to onset	16	20–55	45.1 \pm 8.6	13/3	Phillips, 3, 0.9 \times 0.9 \times 0.9, 52
Early symptoms	13	30–59	50.8 \pm 7.9	7/6	
AD	66	55–93	74 \pm 10.5	40/26	Siemens, 3, 1.2 \times 1 \times 1, 27 Siemens, 1.5, 1.2 \times 1.25 \times 1.25, 7 Phillips, 1.5, 1.2 \times 0.94 \times 0.94, 7 Phillips, 3, 1.2 \times 1 \times 1, 8 GE, 1.5, 1.2 \times 0.94 \times 0.94, 9 GE, 3, 1.2 \times 1.02 \times 1.02, 8
PPA					
Lv	18	51–79	68.3 \pm 5.4	10/8	Siemens, 3, 1 \times 1 \times 1, 21
Sv	16	57–77	65.5 \pm 6.5	11/5	Phillips, 3, 1.2 \times 1 \times 1, 29
Av	16	48–84	68.2 \pm 10.7	9/7	

(Wang et al., 2007; Ceritoglu et al., 2009; Djamanakova et al., 2013). Inversely, the labels were warped to each subject space and then fused by a likelihood fusion algorithm, which took into account both the location and intensity information of each label (Langerak et al., 2010; Sabuncu et al., 2010; Wang et al., 2013). The details of this method, the atlas creation, and the validation in diverse protocols and anatomical phenotypes are described in our previous publications (Tang et al., 2013; Liang et al., 2015; Ma et al., 2015; Wu et al., 2016).

By this multi-atlas automated brain segmentation tool, the raw images, which consisted of more than 1 million voxels were converted to 286 structural representations, of which the volumes were measured. Based on the hierarchical relationship defined in the atlas, these structures can be combined to create five ontological levels with 8–19–53–125–286 structures respectively (Figure 1). Details of the hierarchical-ontological grouping are found in our previous publications (Djamanakova et al., 2014; Wu et al., 2016). One of the reasons for choosing the structure-based multi-level design is that the physician's analysis does not operate at the voxel level, but at the structural level, migrating

freely along the hierarchy. The choice of level is a trade-off between regional specificity and noise: in higher levels, more structures are defined and spatial specificity increases, yet noise also increases. In hypothesis-driven studies, the choice of the level depends on the interest in a given structure. In data-driven studies, the data can be analyzed using all ontological levels combined, or at each level independently. Our present analyses were performed according to the latter approach.

Statistical Analysis and Outputs

We used partial least square—discriminant analysis (PLS-DA) to classify individuals in three different analyses: (1) AT vs. controls, (2) HD vs. controls, and (3) AD vs. controls. As many different protocols as possible were included for each analysis, yet keeping the individuals paired by age, gender, and image protocol in each group compared. The PLS-DA inputs were the regional volumes of brain structures in the five ontological levels, normalized by the intracranial volume. As the classification accuracy increased with the level of granularity and converged at level 3, the results are reported at this level. Level 3 is a medium level of

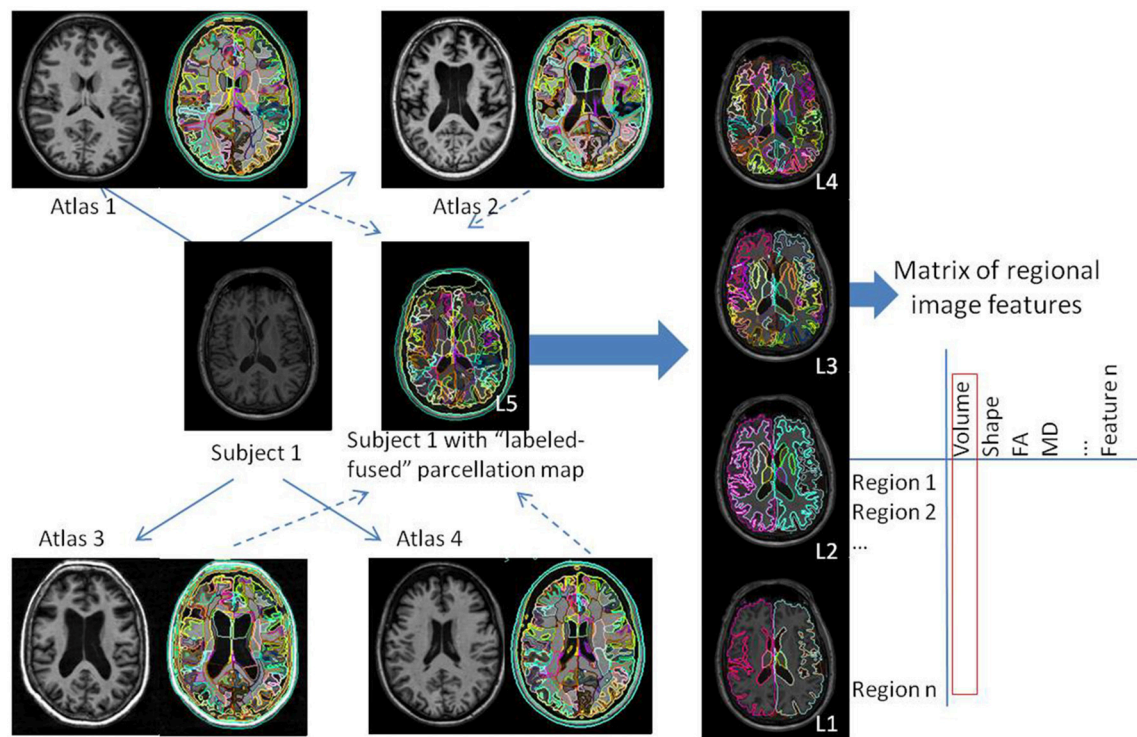


FIGURE 1 | Schematic representation of the automated image parcellation using a multi-atlas likelihood fusion (MALF) algorithm. Each brain image is mapped to each atlas, and the pre-defined labels are correlated with each original brain. The labels can be grouped into five ontological hierarchical levels (L1–L5). By this process, the images are converted to matrices of structures by image features; in the present study, we used the regional volumes.

granularity, where the whole brain is segmented in lobes, deep gray matter, major deep white matter structures, ventricles, and sulci (Djamanakova et al., 2014). It matches well the radiologists reading (Faria et al., 2015), and the segmentation reproducibility is high (Djamanakova et al., 2013; Faria et al., 2015; Liang et al., 2015).

We opted for using simple linear classifiers to reduce the chance of overfitting, increase the potential for generalization of the results, and facilitate the translation to clinical practice, which is our aim, rather than the greatness of the classification. We could have obtained higher classification accuracy using more elaborate classifiers (such as a support vector machine and black-box models). Briefly, PLS is the least restrictive extension of the multiple linear regression models, therefore applicable to situations where the number of predictor variables exceeds the number of observations. As in the principal component analysis (PCA), the scores, or components, are the sets of values of linearly uncorrelated variables and the regression coefficients (loadings or weights) reflect the importance of the predictor variables in the model.

In each analysis, the samples were divided in training set, in which the classifier was built, and test set, in which the accuracy was tested. The validation in an independent test set reduces the impact of overfitting by biased variable selection and results in more realistic classification accuracy. In addition to the classifier accuracy, the outputs of interest were (1) the anatomical features

important for the classification (related to the PLS loading weights) or, in other words, the regional pattern of atrophy that characterizes each group, and (2) the individual's chances of belonging to different groups, which can be of direct importance for clinical guidance. Secondary outputs of interest are (1) the distance among individuals in the principal component space, which can be used for image retrieval of individuals with similar phenotypes, and (2) the individual z-score maps of atrophy.

In the case of PPA, we qualitatively explored a possible natural segregation among the phenotypes with PCA. The inputs were, again, the regional volumes of brain structures, normalized by the intracranial volume. We then assessed the potential of our tools on subdividing groups according to anatomical phenotype, using hierarchical clustering.

RESULTS

Ataxia: Extraction of Homogeneous and Noticeable Image Features

The analysis performed on 16 individuals with ataxia (8 for training, 8 for testing; Supplementary Table), and controls paired by age, gender, and image protocol achieved accuracy of 0.875 in differentiating individuals with AT from controls. **Figure 2** shows the PLS-DA plot (scores vs. loadings) and the two components used by the classifier. Component 1 is mostly responsible for

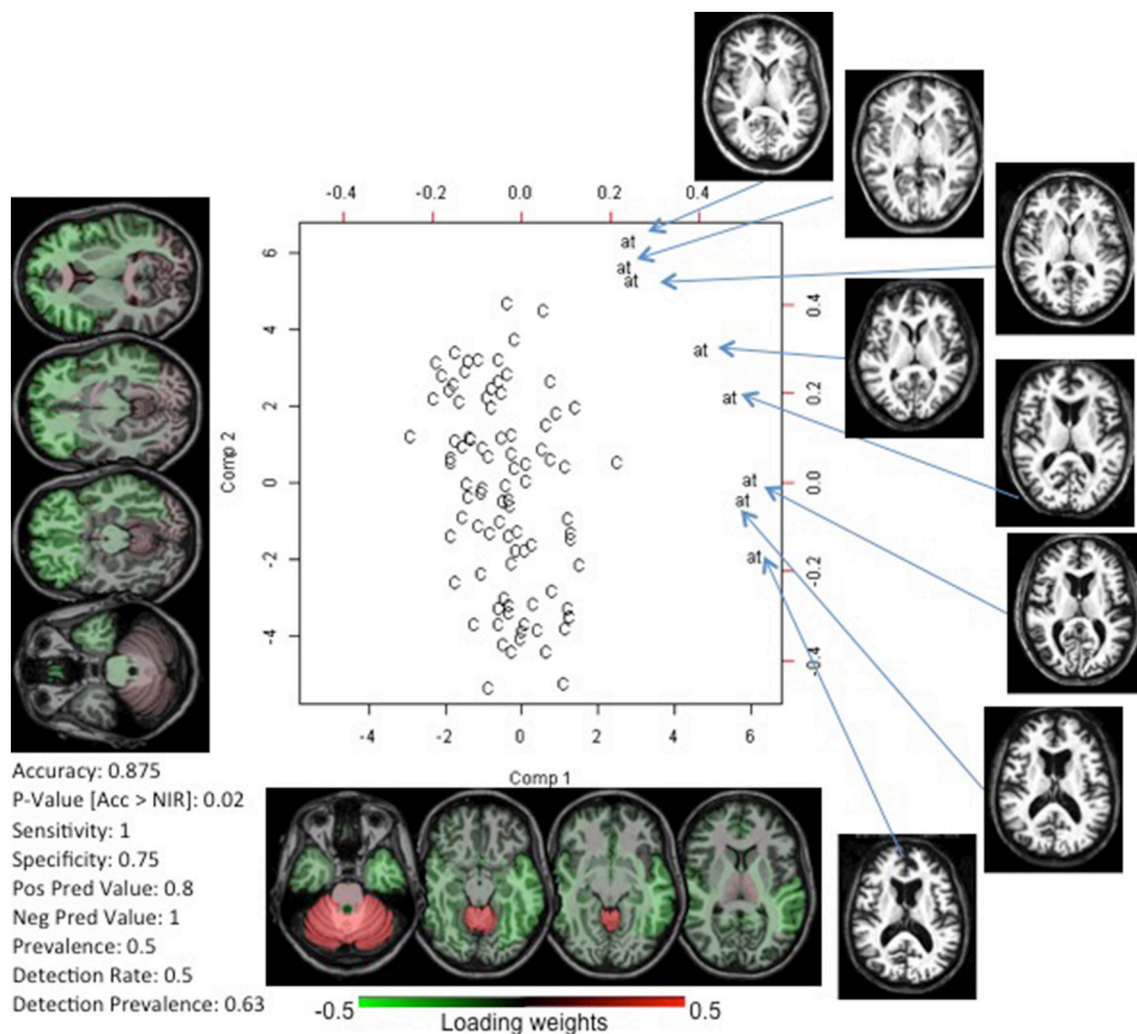


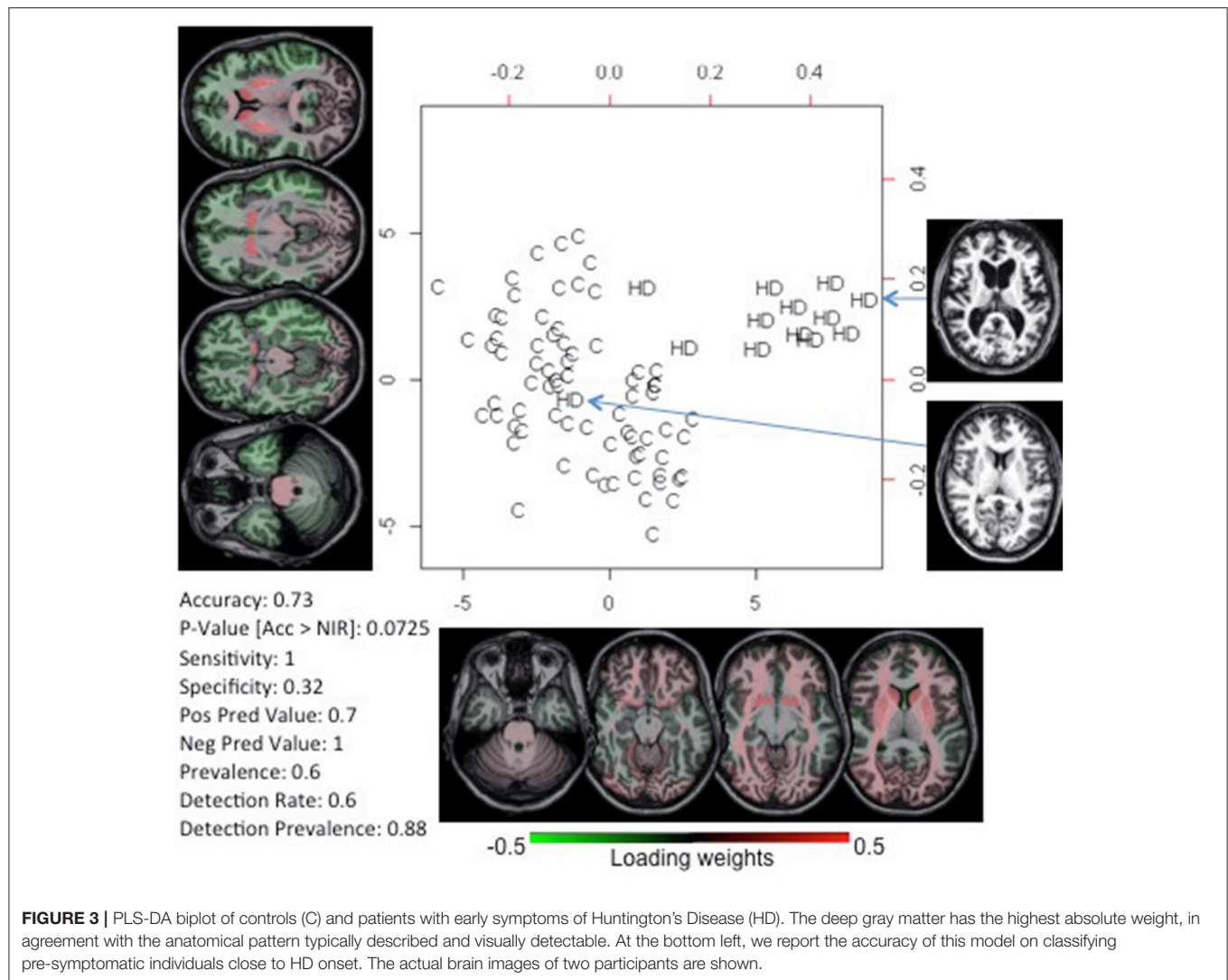
FIGURE 2 | Biplot of scores and loadings from a PLS-DA analysis between controls (C) and patients with ataxia (at). The loading weights of the regional volumes, or the importance of regional atrophy in the classifier, are color-coded on the axial MRIs (radiological view). This gives, at a glance, a snapshot of the important features of the disease. In this case, the volume of the cerebellum (component 1) and brainstem/mesencephalon (component 2) had the highest absolute weights, in agreement with the physiopathology of ataxia. At the bottom left, the classifier accuracy in an external test set is reported. The actual brain images of the patients used in the model are shown.

the segregation between the two groups. The cerebellum had the highest loading, i.e., the cerebellar atrophy played a major role on the classification, in agreement with the well-known and apparent cerebellar atrophy in ataxia. The highest absolute loadings of component 2 are diffusely distributed among the frontal, temporal and parietal lobes; it directly correlated with the degree of atrophy on these lobes, as measured by their volumetric z-score (Pearson rho of 0.72, 0.67, 0.61 for frontal, temporal, and parietal, respectively), and inversely correlated with age ($\rho = -0.77$). Therefore, we infer that component 2 reflects age-related atrophy in individuals with AT.

Huntington's Disease: Prediction

We tested whether we could correctly classify individuals with pre-symptomatic HD using the anatomic features of individuals

with early symptomatic HD. The goal was to use HD as a model to predict conversion to a specific anatomical phenotype rather than to diagnose HD, which can be done precisely by genetic tests. The classifier was built with individuals with early symptoms ($n = 13$) vs. paired controls, and tested in pre-symptomatic individuals close ($n = 16$) and far ($n = 23$) from the onset, vs. paired controls (Supplementary Table). Again, two components were enough to create a model with 73% accuracy in classifying pre-symptomatic individuals near to disease onset (Figure 3). The highest loading weights were in the striatum, as expected, based on the disease physiopathogeny. As described by previous studies, striatum atrophy can barely be determined at the individual level on the pre-symptomatic stage, although it can be detected quantitatively, at the group level, up to 15 years before clinical onset. In addition, the early-symptomatic



HD group is anatomically heterogeneous, with some individuals presenting very clear striatum atrophy and others being very close to normal (**Figure 3**). This indicates that in certain disease types or at certain stages of a disease, the anatomy may not encode enough information to provide diagnosis for all patients. Regardless, we were effective enough in capturing and using this feature for the individual classification. The model did not achieve accuracy significantly higher than the by-chance for classifying pre-symptomatic individuals far from HD onset.

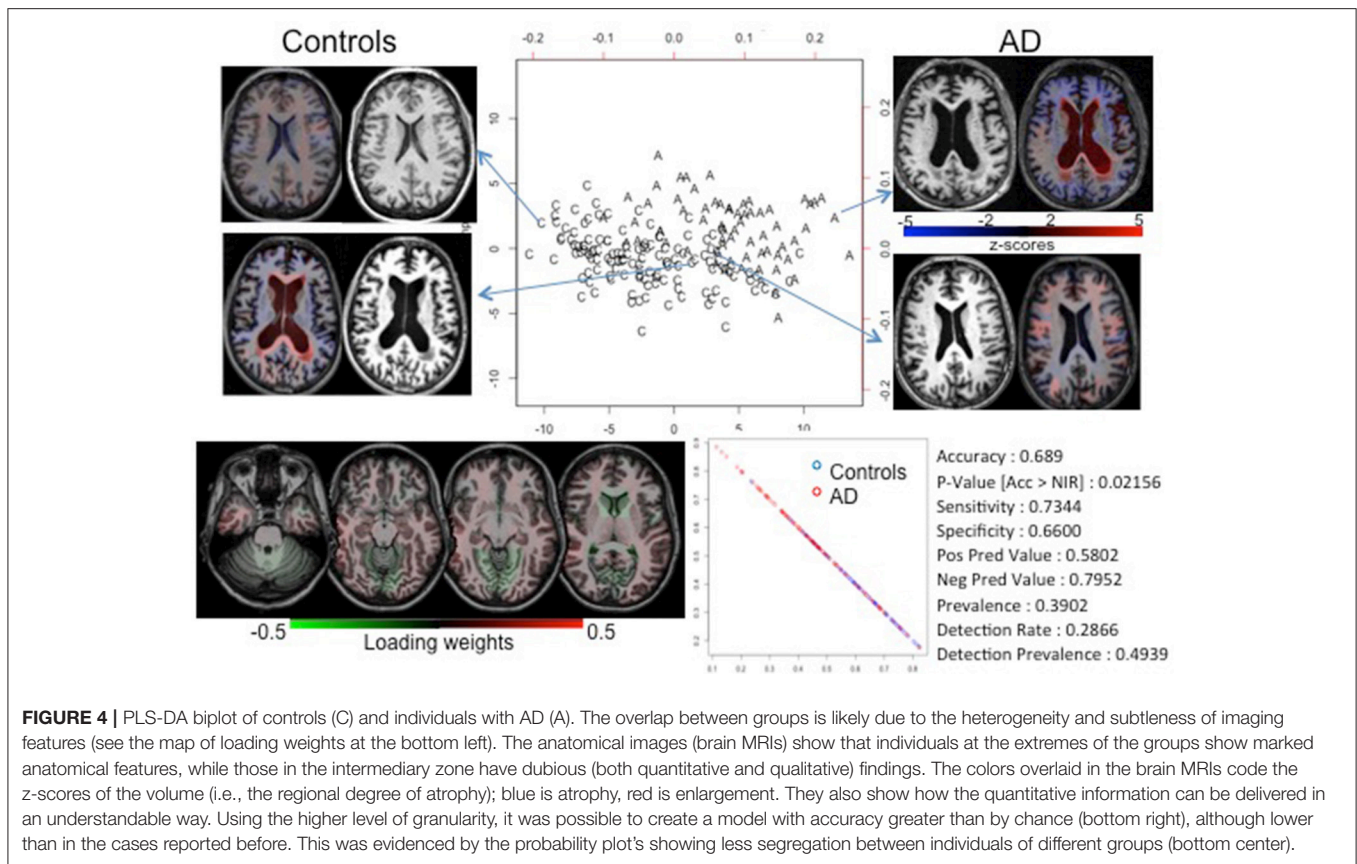
Alzheimer's Disease: Classification of Diseases with Subtle or Heterogeneous Abnormalities

Unlike in ataxia and HD, the atrophy in most of the neurodegenerative diseases is detectable at the late stage of the disease and is regionally heterogeneous. This is the case with AD. We achieved a reasonable accuracy (69%) in diagnosing AD (model built in 33 AD individuals vs. paired controls, and tested in independent 33 AD individuals vs. paired controls;

see Supplementary Table), significantly higher than the by-chance classification. However, there was an enormous overlap among groups, as notable in the PLS-DA plot and in the probability plot that represents the chance of each individual's belonging to each group (**Figure 4**). The loading-weights map showed no distinguishing features; the weights are comparable and widespread, indicating that the anatomy in AD is mildly or heterogeneously affected, which can be confirmed by visual inspection of the brain MRIs.

Primary Progressive Aphasia: Binning by Anatomical Phenotype

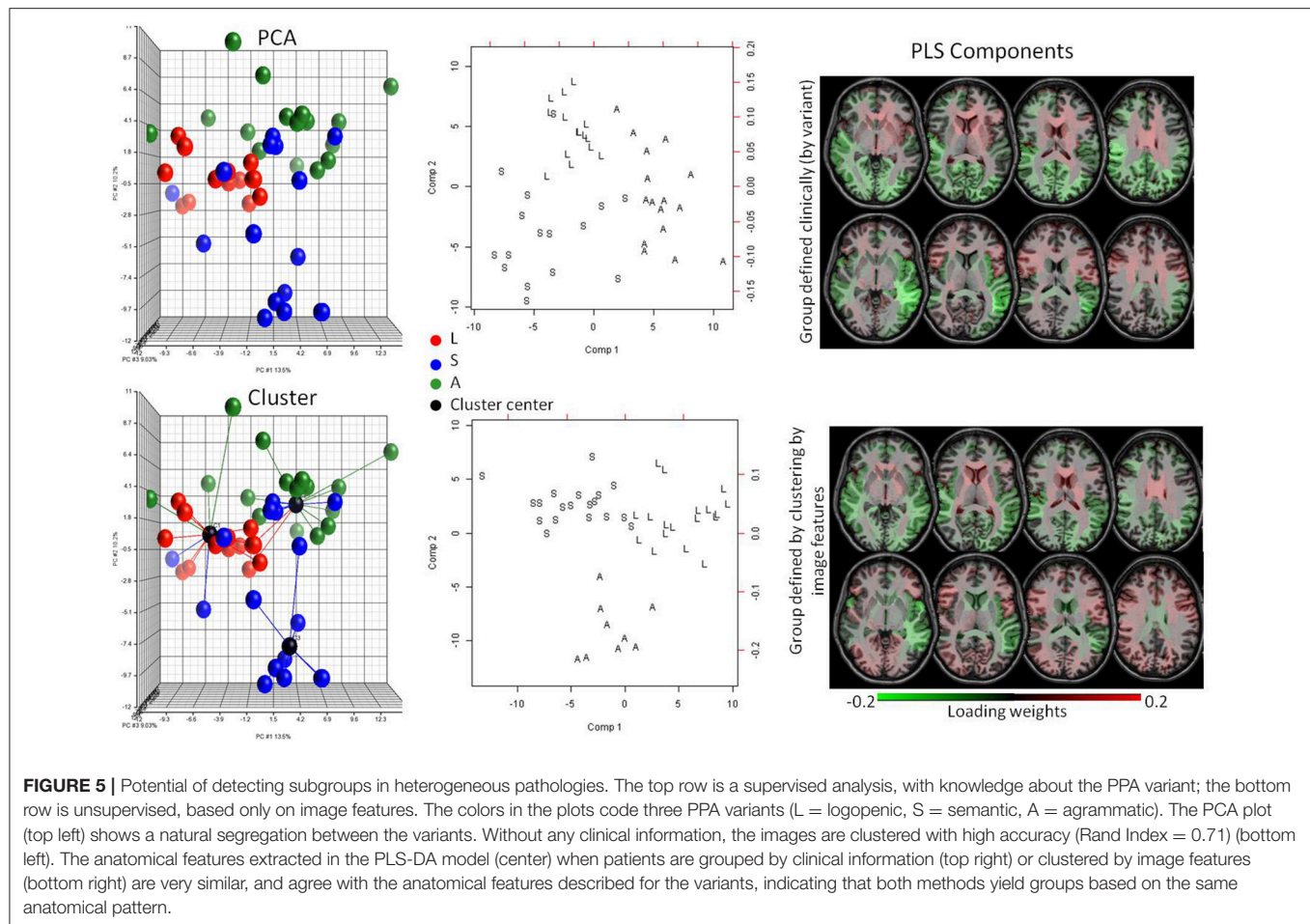
As mentioned in the previous section, increasingly therapies are targeting the early stages of neurodegenerative diseases. However, accurate diagnosis is more difficult because of the lack of clear and/or specific clinical deficits. At this stage, the initial stratification of the heterogeneous patient population is of critical importance. The difficulty arises because potential patient subgroups are degenerate both in the clinical and anatomical



domains. In this case, we are interested less in correlation between present diagnosis and anatomical features (because the diagnosis based on clinical information cannot separate important subgroups) and more in expanding the patient populations using both clinical and anatomical manifestations, potentially identifying a way to define subgroups. Binning a disease into subgroups may facilitate the design of therapies and the creation of predictive models because the subgroups may be related to specific pathological substrates, deficits or prognoses. We used PPA as a model system because of the existence of three well-known clinical variants. The knowledge of their anatomical correlates, albeit loose, could serve as our gold standard. In the PCA of the anatomical features (the regional volumes) there was a natural segregation into three clinically labeled groups (Figure 5). By clustering the data using only the anatomical features, we found groups that accurately agreed with the variant diagnosis (Rand Index = 0.71). Then, by using PLS-DA and extracting the loading weights, we confirmed that the features for automated classification according to clusters agreed with those for the classification according to clinical diagnosis. In addition, these anatomical features agreed with what is clinically defined for the variants, such as predominance of atrophy in the left temporal lobe for the Semantic variant, in the inferior parietal for the Logopenic, and in the inferior frontal lobe and the insula for the Agrammatic (Gorno-Tempini et al., 2004).

DISCUSSION

We evaluated the performance of structure-based computational analysis on extracting anatomical features, previously described by human experience and a priori biological knowledge, in specific patient populations. Previously, we tested the robustness of our automated quantification approach against different image protocols and scanners, using subjects with different patterns and degrees of brain atrophy, and compared our conclusions with those of trained clinicians using visual analysis (Djamanakova et al., 2013; Faria et al., 2015; Liang et al., 2015). In the present study, we tested whether we could classify individuals and anatomically characterize different diseases in simulated clinical scenarios. Our database contains diverse image protocols and scanners. The demographic information taken into account by the linear classifiers include only age and gender, which are always clinically available. Although we could create better classification models by adding other clinical information, homogenizing the dataset, or using classifiers more sophisticated than PLS-DA, this would reduce the potential for generalization and translation to real clinical situations. In summary, rather than the greatness of classification, our aim was to create models robust enough to be translated to clinical practice, and at least in a first step, perform as well as clinicians in terms of extraction of important anatomical features and detection of anatomical patterns, helping to fill the semantic gap.



Detection of Abnormal Imaging Patterns

In a disease with a clear anatomical phenotype (Ataxia), we obtained 87.5% accuracy, using a small sample size of patients in different stages of the disease. More important, the anatomical features extracted agreed with what is previously described as the hallmark of Ataxia (cerebellar and brainstem atrophy) (Klockgether et al., 1998; Schulz et al., 1999; Eichler et al., 2011; Reetz et al., 2013). The maps of the loading weights and the visual inspection of the images (Figure 2) reveal that the first component carries mostly information about the disease's anatomical phenotype, while the second component basically reflects brain atrophy directly related to age. Thus, the components extracted carry biological meaning, i.e., they contain information that can be interpreted in the light of actual medical knowledge because they both (our quantification tool and the medical knowledge) are based at the level of anatomical structures. In consequence, the classification models and the feature extraction machinery can be easily interpreted and translated to clinical practice. Although this result is purely confirmatory, the quantitative and systematic characterization of the anatomical feature in the PLS-DA space may give us an interesting clue about the patient status. For example, if there are ataxia patients who not only have the typical ataxia

feature (component 1), but also are located at an unexpected position in component 2 (i.e., accelerated whole-brain atrophy related to age), this may correlate with poor future outcomes. Thus, a quantitative approach of this type could provide new insight into diagnosis and prognosis, further facilitating research.

To investigate the prognostic value of quantitative anatomical description, we tested the classification performance in diseases where the anatomy clearly correlates with the time course, applying the classifiers in stages where the abnormal features couldn't be detected visually, at the individual level. In other words, we tested the potential for prognostic prediction using the HD population. We achieved 73% accuracy in classifying pre-symptomatic HD individuals, with a model based on features of early symptomatic HD individuals. The feature selection identified the deep gray matter as the most important region for the classification, again agreeing with the physiopathology of HD (Figure 3) (Aylward et al., 2000; Nopoulos et al., 2010; Paulsen et al., 2010, 2014a,b; Guo et al., 2012; Delmaire et al., 2013; Georgiou-Karistianis et al., 2013; Faria et al., 2016). HD is a genetic disease where the product of genetic load and age correlates very well with the time to onset (Ross et al., 2014). Therefore, one can reasonably argue that predictive models

based on imaging features are useless. The same applies to Ataxia to some extent. However, our aim was not to diagnosis HD or Ataxia. These diseases were taken as models for proof of concept because the gold standard (clinical diagnosis) is well-established. The aim was to evaluate the structure-based automated quantification approach, in terms of feature selection and robustness against heterogeneous datasets, and its potential to detect features that go beyond the artifactual noise. Particularly in HD, the potential for classifying pre-symptomatic individuals surpasses what can be done with clinical imaging analysis because the subtle abnormalities are not visually detectable at the individual level (Paulsen et al., 2008).

Potential for Binning in More Homogeneous Phenotypes

Unlike diseases with a clear anatomical phenotype, those that embrace heterogeneous anatomical and clinical phenotypes, or subtle abnormalities, or unknown time courses, offer extra challenges for both visual and automated analysis. This is, for instance, the case with AD. To date, there are about 100 models for predicting conversion from mild cognitive impairment to AD, based on imaging. A PubMed search for “Prediction of MCI to AD conversion MRI” reveals 96 publications in the last 10 years; for more recent reviews, please see (Shaffer et al., 2013; Sanchez-Catusas et al., 2017). Either they achieve unsatisfactory accuracy, or high accuracy at the cost of overfitting, or they are late in the disease course. As a result, we can generalize by saying that there is, as yet, no effective prediction useful in clinical scenarios. **Figure 4** may offer some clues about why this happens. Our classification model achieved <70% accuracy. There is substantial overlap between controls and AD in the PLS-DA plot, and there is no predominant weight in the loadings of component 1. Visual inspection of the images reveals that both groups (control and AD) are heterogeneous in terms of atrophy pattern and degree at this age range. This explains why the individual classification, by visual radiological analysis, is also ineffective.

The source of this challenge is two-fold. First, it is possible that anatomy is not encoding enough information to characterize the pathology reliably. Second, because we do not have strongly discriminating factors, both in clinical and imaging information, the stratification of the patient population is incomplete. For example, if AD is actually a syndrome caused by multiple pathologies with multiple anatomical manifestations, AD's common anatomical features cannot be extracted. In this situation, we need to resort to different study designs, using both clinical and imaging features to stratify the population. Models such as AD provide opportunities to investigate the existence of subgroups, with certain anatomical expression, that can behave as specific entities in some clinical domains. For instance, in Ataxia (**Figure 2**) one can see a subtle spread of patients along the component that differentiates the groups (component 1). Hypothetically, this spread may reflect the effect of a correlated feature, such as disease severity. Similarly in AD or other heterogeneous disease models, there may be a non-orthogonal axis that represents an unknown variable. With regression in

this axis, it is possible to detect the subgroups that, for instance, respond differently to therapeutics, or have different prognosis.

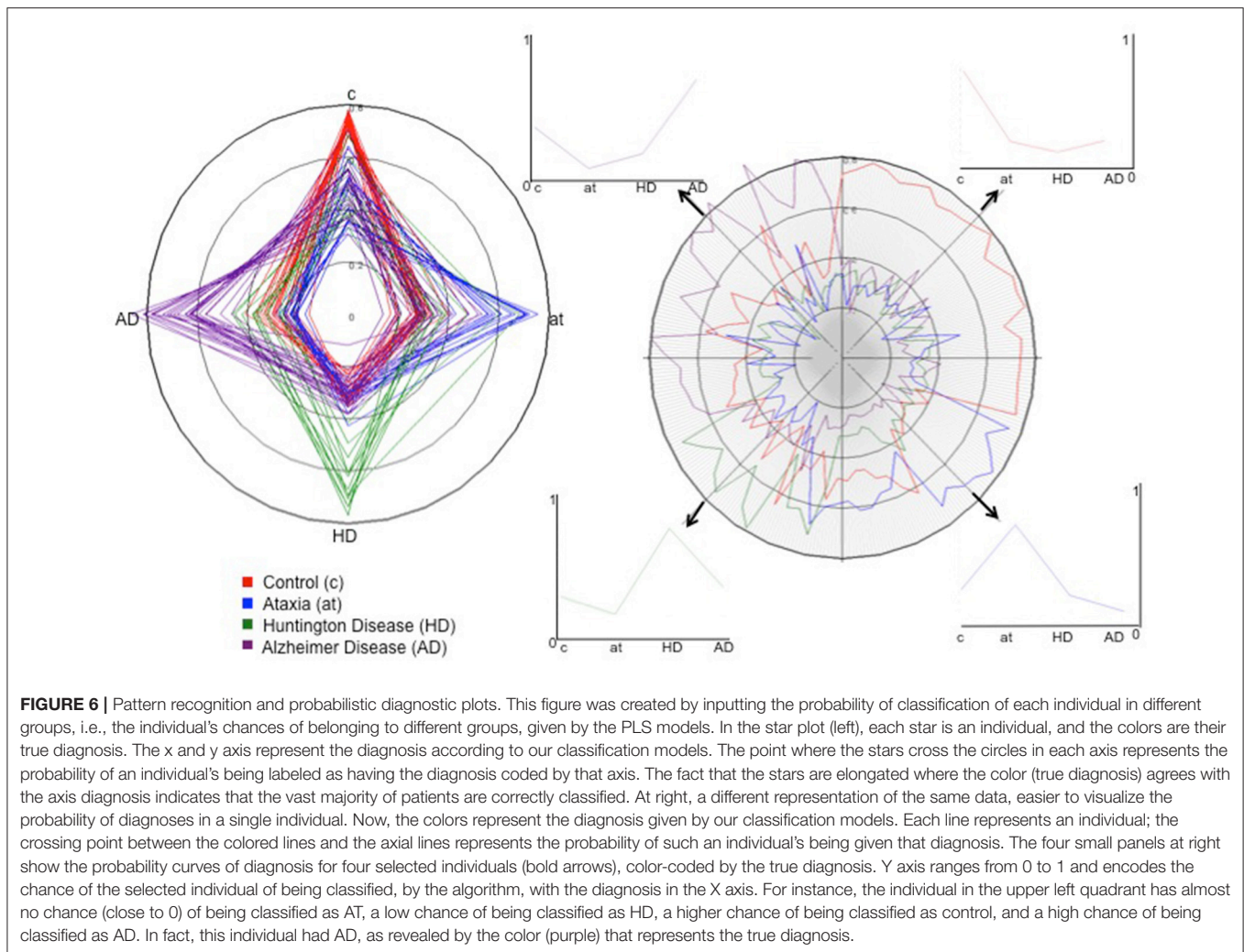
To investigate the potential of the automated structure-based quantification to binning an entity into subgroups of clinical relevance, we used individuals with PPA. PPA, a neurodegenerative clinical syndrome characterized by decline in language ability 2 years before any other cognitive deficit, is an ideal condition to investigate the clustering in sub-phenotypes, since three variants loosely correlated with underlying pathologies and with certain anatomical representation are described (Gorno-Tempini et al., 2011; Rohrer and Rosen, 2013). Although there is still no treatment for PPA, there is hope that certain therapies can be effective for specific variants (Cadorio et al., 2017). Now, suppose that the three variants are yet unknown. An unsupervised PCA plot shows a natural segregation of the data into two or three subgroups (**Figure 5**), but because the variants are hypothetically unknown, one cannot explain the data variance with clinical labels. An unsupervised hierarchical cluster shows the data divided into subgroups that correlate very well with the real variant's diagnosis. The image features selected for classification in these clusters (bottom row, **Figure 5**) agree with those selected for classification according to the real variant's diagnosis (top row, **Figure 5**) and also to those that are described as hallmarks for the variants (Turner et al., 1996; Rohrer et al., 2009; Shim et al., 2012; Zhang et al., 2013; Agosta et al., 2015; Botha et al., 2015; Bisenius et al., 2016), proving the potential of our approach to identify subgroups of clinical relevance.

Deliverables

Subgrouping can be extrapolated to individuals, i.e., the detection of outliers in terms of anatomy may point to individuals who may be unique in additional domains. For instance, in HD (**Figure 3**) anatomical heterogeneity still remains among the genetically homogenized group, as there is at least one individual with visually normal anatomy. It is an open question if this anatomical variability has any predictive value for prognosis, to be answered by quantitative and systematic characterization of this population.

Another potential deliverable is the diagnostic probability map for each individual (**Figure 6**). Given a database large enough to contain various pathologies and the high variability of imaging protocols and age range for controls and patients, it is possible to calculate the probability of differential diagnosis for a new individual, as shown in **Figure 6**. In this example, one can reasonably argue that it is clinically improbable to have HD, Ataxia and AD as differential diagnoses. Again, these diseases were taken as proof of concept, because they all have the same basic anatomic feature (atrophy) and a clear clinical diagnosis used as the gold-standard. The concept of diagnostic probability graphics can be extended to more plausible clinical scenarios.

Finally, the potential for aiding clinical interpretation and education may be a valuable low-hanging fruit. The simple use of z-score maps (**Figure 4**) may confirm or exclude a clinical impression and speed up the radiological reading. Also, having a database big and heterogeneous enough, and coupling image and text information (such as diagnosis, prognosis, response to



treatments, etc.), it is possible to perform a direct image search, producing static reports about similar phenotypes. For example, given a new subject image, it is possible to search in a big dataset for dozens of images with similar features linked to information of clinical relevance.

Limitations

This study is based on a single image variable, the volume. One of the greatest advantages of the structure-based approach is that it allows the combination of many other features, such as T2 contrast, diffusion tensor image indices, functional MRI correlations, metabolite concentration, and others, as we demonstrated in previous studies (Faria et al., 2012). Although there are big challenges in combining features of different domains (e.g., drawbacks on feature concatenation methods, variation among clinical protocols barring the creation of common databases for certain domains, the need for a priori knowledge of noise in order to create models for easy generalization), multi-domain structure-based analysis is a promising strategy for conditions with no single dominant discriminating feature.

An important constraint of the structure-based analysis is that any quantitative characterization and classification model will be limited by the pre-defined space. In other words, if the anatomical pattern does not respect the boundaries of a given parcellation scheme, the abnormality can be overlooked. One strategy to ameliorate this issue is to use different levels of granularity. So one can analyze the data in parcels as big as a hemisphere, or as small as a cranial nerve, which is actually smaller than the gaussian filters traditionally used for voxel-based analysis. However, if the strategy is to replicate the radiological interpretation, then structure-based analysis is intuitively a better solution because visual inspection occurs at the structural level, not at the voxel level.

Perspectives

We explored the performance of structure-based computational analysis in simulated clinical scenarios. The pillars of this approach are automated and accurate quantification, reliability and robustness against artifactual noise, easy interpretation of selected features, and a knowledge repository that is a large database as heterogeneous as possible both in terms of

pathologies and image protocols. The deliverables are diverse, from the image quantification itself, through to the image pattern search, to the diagnostic aid. Although the impact of this method is yet to be tested, it has potential educational value, it may reduce the time for radiological reading, or it may work as second reader in locations where sub-specialized radiologists are not available. In any case, because no such tool can be directly applicable to clinical practice, any positive impact is valuable. In addition, electronic structurized databases and search engines are the basis of high throughput image analysis and may represent the migration of brain MRI to the BigData era, contributing to the emergent field of Precision Medicine.

ETHICS STATEMENT

This study was carried out with data (1) from publically available databases, and (2) from Johns Hopkins Hospital, originally collected for other studies. These studies were approved and carried out in accordance with the recommendations of the local IRB, with written informed consent from all subjects, in accordance with the Declaration of Helsinki.

REFERENCES

- Agosta, F., Ferraro, P. M., Canu, E., Copetti, M., Galantucci, S., Magnani, G., et al. (2015). Differentiation between subtypes of primary progressive aphasia by using cortical thickness and diffusion-tensor MR imaging measures. *Radiology* 276, 219–227. doi: 10.1148/radiol.15141869
- Akgul, C. B., Rubin, D. L., Napel, S., Beaulieu, C. F., Greenspan, H., and Acar, B. (2011). Content-based image retrieval in radiology: current status and future directions. *J. Digit. Imaging* 24, 208–222. doi: 10.1007/s10278-010-9290-9
- Albert, M. S., Dekosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* 7, 270–279. doi: 10.1016/j.jalz.2011.03.008
- Apostolova, L. G., Steiner, C. A., Akopyan, G. G., Dutton, R. A., Hayashi, K. M., Toga, A. W., et al. (2007). Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. *Arch. Neurol.* 64, 1489–1495. doi: 10.1001/archneur.64.10.1489
- Artaechevarria, X., Munoz-Barrutia, A., and Ortiz-De-Solorzano, C. (2009). Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans. Med. Imaging* 28, 1266–1277. doi: 10.1109/TMI.2009.2014372
- Aylward, E. H., Codori, A. M., Rosenblatt, A., Sherr, M., Brandt, J., Stine, O. C., et al. (2000). Rate of caudate atrophy in presymptomatic and symptomatic stages of Huntington's disease. *Mov. Disord.* 15, 552–560. doi: 10.1002/1531-8257(200005)15:3<552::AID-MDS1020>3.0.CO;2-P
- Bisenius, S., Neumann, J., and Schroeter, M. L. (2016). Validating new diagnostic imaging criteria for primary progressive aphasia via anatomical likelihood estimation meta-analyses. *Eur. J. Neurol.* 23, 704–712. doi: 10.1111/ene.12902
- Botha, H., Duffy, J. R., Whitwell, J. L., Strand, E. A., Machulda, M. M., Schwarz, C. G., et al. (2015). Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex* 69, 220–236. doi: 10.1016/j.cortex.2015.05.013
- Butteriss, D., Chinnery, P., and Birchall, D. (2005). Radiological characterization of spinocerebellar ataxia type 6. *Br. J. Radiol.* 78, 694–696. doi: 10.1259/bjr/73834093
- Cadorio, I., Lousada, M., Martins, P., and Figueiredo, D. (2017). Generalization and maintenance of treatment gains in primary progressive aphasia

AUTHOR CONTRIBUTIONS

All the authors are accountable for all aspects of the work and approved the final version. In addition, AF and SM are responsible for the conception and design, interpretation of data, and drafting the work. ZL is responsible for data acquisition and analysis. MIM revised it critically for important intellectual content.

ACKNOWLEDGMENTS

We are grateful to the individuals who participated in this research. We thank Dr. Argye Hillis, Dr. Christopher Ross, and Dr. Sarah Ying, for data sharing. This research was possible due to the following grants, from the National Institutes of Health: P41EB015909 (SM and MIM), R01EB017638 (MIM), and R01NS084957 (SM).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2017.00578/full#supplementary-material>

- (PPA): a systematic review. *Int. J. Lang. Commun. Disord.* 52, 543–560. doi: 10.1111/1460-6984.12310
- Ceritoglu, C., Oishi, K., Li, X., Chou, M. C., Younes, L., Albert, M., et al. (2009). Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging. *Neuroimage* 47, 618–627. doi: 10.1016/j.neuroimage.2009.04.057
- Delmaire, C., Dumas, E. M., Sharman, M. A., van den Bogaard, S. J., Valabregue, R., Jauffret, C., et al. (2013). The structural correlates of functional deficits in early huntington's disease. *Hum. Brain Mapp.* 34, 2141–2153. doi: 10.1002/hbm.22055
- Depeursinge, A., Fischer, B., Muller, H., and Deserno, T. M. (2011). Prototypes for content-based image retrieval in clinical practice. *Open Med. Inform. J.* 5, 58–72. doi: 10.2174/1874431101105010058
- Djamanakova, A., Faria, A. V., Hsu, J., Ceritoglu, C., Oishi, K., Miller, M. I., et al. (2013). Diffeomorphic brain mapping based on T1-weighted images: improvement of registration accuracy by multichannel mapping. *J. Magn. Reson. Imaging* 37, 76–84. doi: 10.1002/jmri.23790
- Djamanakova, A., Tang, X., Li, X., Faria, A. V., Ceritoglu, C., Oishi, K., et al. (2014). Tools for multiple granularity analysis of brain MRI data for individualized image analysis. *Neuroimage* 101, 168–716. doi: 10.1016/j.neuroimage.2014.06.046
- Eichler, L., Bellenberg, B., Hahn, H. K., Koster, O., Schols, L., and Lukas, C. (2011). Quantitative assessment of brain stem and cerebellar atrophy in spinocerebellar ataxia types 3 and 6: impact on clinical status. *AJNR Am. J. Neuroradiol.* 32, 890–897. doi: 10.3174/ajnr.A2387
- Faria, A. V., Joel, S. E., Zhang, Y., Oishi, K., van Zijl, P. C., Miller, M. I., et al. (2012). Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. *Neuroimage* 61, 613–621. doi: 10.1016/j.neuroimage.2012.03.078
- Faria, A. V., Oishi, K., Yoshida, S., Hillis, A., Miller, M. I., and Mori, S. (2015). Content-based image retrieval for brain MRI: an image-searching engine and population-based analysis to utilize past clinical data for future diagnosis. *Neuroimage Clin.* 7, 367–376. doi: 10.1016/j.nicl.2015.01.008
- Faria, A. V., Ratnanather, J. T., Tward, D. J., Lee, D. S., van den Noort, F., Wu, D., et al. (2016). Linking white matter and deep gray matter alterations in premanifest Huntington disease. *Neuroimage Clin.* 11, 450–460. doi: 10.1016/j.nicl.2016.02.014

- Frisoni, G. B., Boccardi, M., Barkhof, F., Blennow, K., Cappa, S., Chiotis, K., et al. (2017). Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. *Lancet Neurol.* 16, 661–676. doi: 10.1016/S1474-4422(17)30159-X
- Georgiou-Karistianis, N., Scallill, R., Tabrizi, S. J., Squitieri, F., and Aylward, E. (2013). Structural MRI in Huntington's disease and recommendations for its potential use in clinical trials. *Neurosci. Biobehav. Rev.* 37, 480–490. doi: 10.1016/j.neubiorev.2013.01.022
- Gorno-Tempini, M. L., Dronkers, N. F., Rankin, K. P., Ogar, J. M., Phengrasamy, L., Rosen, H. J., et al. (2004). Cognition and anatomy in three variants of primary progressive aphasia. *Ann. Neurol.* 55, 335–346. doi: 10.1002/ana.10825
- Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014. doi: 10.1212/WNL.0b013e31821103e6
- Guo, Z., Rudow, G., Pletnikova, O., Codispoti, K. E., Orr, B. A., Crain, B. J., et al. (2012). Striatal neuronal loss correlates with clinical motor impairment in Huntington's disease. *Mov. Disord.* 27, 1379–1386. doi: 10.1002/mds.25159
- Huntington, G. (1872). On Chorea. *Med. Surg. Rep.* 26:2.
- Hwang, K. H., Lee, H., and Choi, D. (2012). Medical image retrieval: past and present. *Healthc. Inform. Res.* 18, 3–9. doi: 10.4258/hir.2012.18.1.3
- Jack, C. R. Jr., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., et al. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* 7, 257–262. doi: 10.1016/j.jalz.2011.03.004
- Jia, H., Yap, P. T., and Shen, D. (2012). Iterative multi-atlas-based multi-image segmentation with tree-based registration. *Neuroimage* 59, 422–430. doi: 10.1016/j.neuroimage.2011.07.036
- Kalpathy-Cramer, J., de Herrera, A. G., Demner-Fushman, D., Antani, S., Bedrick, S., and Muller, H. (2015). Evaluating performance of biomedical image retrieval systems—an overview of the medical image retrieval task at ImageCLEF 2004–2013. *Comput. Med. Imaging Graph.* 39, 55–61. doi: 10.1016/j.compmedimag.2014.03.004
- Klockgether, T., Skalej, M., Wedekind, D., Luft, A. R., Welte, D., Schulz, J. B., et al. (1998). Autosomal dominant cerebellar ataxia type I. MRI-based volumetry of posterior fossa structures and basal ganglia in spinocerebellar ataxia types 1, 2 and 3. *Brain* 121(Pt. 9), 1687–1693. doi: 10.1093/brain/121.9.1687
- Langerak, T. R., van der Heide, U. A., Kotte, A. N., Viergever, M. A., Van Vulpen, M., and Pluim, J. P. (2010). Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Trans. Med. Imaging* 29, 2000–2008. doi: 10.1109/TMI.2010.2057442
- Leyton, C. E., Britton, A. K., Hodges, J. R., Halliday, G. M., and Kril, J. J. (2016). Distinctive pathological mechanisms involved in primary progressive aphasia. *Neurobiol. Aging* 38, 82–92. doi: 10.1016/j.neurobiolaging.2015.10.017
- Liang, Z., He, X., Ceritoglu, C., Tang, X., Li, Y., Kuttan, K. S., et al. (2015). Evaluation of cross-protocol stability of a fully automated brain multi-atlas parcellation tool. *PLoS ONE* 10:e0135533. doi: 10.1371/journal.pone.0135533
- Lotjonen, J. M., Wolz, R., Koikkalainen, J. R., Thurfjell, L., Waldemar, G., Soininen, H., et al. (2010). Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 49, 2352–2365. doi: 10.1016/j.neuroimage.2009.10.026
- Ma, J., Ma, H. T., Li, H., Ye, C., Wu, D., Tang, X., et al. (2015). A fast atlas pre-selection procedure for multi-atlas based brain segmentation. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 3053–3056. doi: 10.1109/EMBC.2015.7319036
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R. Jr., Kawas, C. H., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- Mesulam, M. M. (1982). Slowly progressive aphasia without generalized dementia. *Ann. Neurol.* 11, 592–598. doi: 10.1002/ana.410110607
- Nopoulos, P. C., Aylward, E. H., Ross, C. A., Johnson, H. J., Magnotta, V. A., Juhl, A. R., et al. (2010). Cerebral cortex structure in prodromal Huntington disease. *Neurobiol. Dis.* 40, 544–554. doi: 10.1016/j.nbd.2010.07.014
- Orphanoudakis, S. C., Chronaki, C. E., and Vamvaka, D. (1996). I2Cnet: Content-based similarity search in geographically distributed repositories of medical images. *Comput. Med. Imaging Graph.* 20, 193–207. doi: 10.1016/S0895-6111(96)00013-4
- Paulsen, J. S., Langbehn, D. R., Stout, J. C., Aylward, E., Ross, C. A., Nance, M., et al. (2008). Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *J. Neurol. Neurosurg. Psychiatr.* 79, 874–880. doi: 10.1136/jnnp.2007.128728
- Paulsen, J. S., Long, J. D., Johnson, H. J., Aylward, E. H., Ross, C. A., Williams, J. K., et al. (2014a). Clinical and biomarker changes in premanifest huntington disease show trial feasibility: a decade of the PREDICT-HD Study. *Front. Aging Neurosci.* 6:78. doi: 10.3389/fnagi.2014.00078
- Paulsen, J. S., Long, J. D., Ross, C. A., Harrington, D. L., Erwin, C. J., Williams, J. K., et al. (2014b). Prediction of manifest Huntington's disease with clinical and imaging measures: a prospective observational study. *Lancet Neurol.* 13, 1193–1201. doi: 10.1016/S1474-4422(14)70238-8
- Paulsen, J. S., Nopoulos, P. C., Aylward, E., Ross, C. A., Johnson, H., Magnotta, V. A., et al. (2010). Striatal and white matter predictors of estimated diagnosis for Huntington disease. *Brain Res. Bull.* 82, 201–207. doi: 10.1016/j.brainresbull.2010.04.003
- Penney, J. B. Jr., Vonsattel, J. P., Macdonald, M. E., Gusella, J. F., and Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann. Neurol.* 41, 689–692. doi: 10.1002/ana.410410521
- Pinho, E., Godinho, T., Valente, F., and Costa, C. (2017). A multimodal search engine for medical imaging studies. *J. Digit. Imaging* 30, 39–48. doi: 10.1007/s10278-016-9903-z
- Reetz, K., Costa, A. S., Mirzazade, S., Lehmann, A., Juzek, A., Rakowicz, M., et al. (2013). Genotype-specific patterns of atrophy progression are more sensitive than clinical decline in SCA1, SCA3 and SCA6. *Brain* 136, 905–917. doi: 10.1093/brain/aww369
- Rohrer, J. D., and Rosen, H. J. (2013). Neuroimaging in frontotemporal dementia. *Int. Rev. Psychiatry* 25, 221–229. doi: 10.3109/09540261.2013.778822
- Rohrer, J. D., Warren, J. D., Modat, M., Ridgway, G. R., Douiri, A., Rossor, M. N., et al. (2009). Patterns of cortical thinning in the language variants of frontotemporal lobar degeneration. *Neurology* 72, 1562–1569. doi: 10.1212/WNL.0b013e3181a4124e
- Ross, C. A., Aylward, E. H., Wild, E. J., Langbehn, D. R., Long, J. D., Warner, J. H., et al. (2014). Huntington disease: natural history, biomarkers and prospects for therapeutics. *Nat. Rev. Neurol.* 10, 204–216. doi: 10.1038/nrneurol.2014.24
- Sabuncu, M. R., Yeo, B. T., Van Leemput, K., Fischl, B., and Golland, P. (2010). A generative model for image segmentation based on label fusion. *IEEE Trans. Med. Imaging* 29, 1714–1729. doi: 10.1109/TMI.2010.2050897
- Sanchez-Catusas, C. A., Stormezand, G. N., van Laar, P. J., De Deyn, P. P., Sanchez, M. A., and Dierckx, R. A. (2017). FDG-PET for Prediction of AD Dementia in Mild Cognitive Impairment. A review of the state of the art with particular emphasis on the comparison with other neuroimaging modalities (MRI and Perfusion SPECT). *Curr. Alzheimer. Res.* 14, 127–142. doi: 10.2174/1567205013666160629081956
- Schulz, J. B., Skalej, M., Wedekind, D., Luft, A. R., Abele, M., Voigt, K., et al. (1999). Magnetic resonance imaging-based volumetry differentiates idiopathic Parkinson's syndrome from multiple system atrophy and progressive supranuclear palsy. *Ann. Neurol.* 45, 65–74. doi: 10.1002/1531-8249(199901)45:1<65::AID-ART12>3.0.CO;2-1
- Shaffer, J. L., Petrella, J. R., Sheldon, F. C., Choudhury, K. R., Calhoun, V. D., Coleman, R. E., et al. (2013). Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers. *Radiology* 266, 583–591. doi: 10.1148/radiol.12120010
- Shim, H., Hurley, R. S., Rogalski, E., and Mesulam, M. M. (2012). Anatomic, clinical, and neuropsychological correlates of spelling errors in primary progressive aphasia. *Neuropsychologia* 50, 1929–1935. doi: 10.1016/j.neuropsychologia.2012.04.017
- Spanier, A. B., Cohen, D., and Joskowicz, L. (2017). A new method for the automatic retrieval of medical cases based on the RadLex ontology. *Int. J. Comput. Assist. Radiol. Surg.* 12, 471–484. doi: 10.1007/s11548-016-1496-y
- Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers. Dement.* 7, 280–292. doi: 10.1016/j.jalz.2011.03.003
- Tabrizi, S. J., Langbehn, D. R., Leavitt, B. R., Roos, R. A., Durr, A., Craufurd, D., et al. (2009). Biological and clinical manifestations of Huntington's disease in

- the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol.* 8, 791–801. doi: 10.1016/S1474-4422(09)70170-X
- Tabrizi, S. J., Reilmann, R., Roos, R. A., Durr, A., Leavitt, B., Owen, G., et al. (2012). Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *Lancet Neurol.* 11, 42–53. doi: 10.1016/S1474-4422(11)70263-0
- Tabrizi, S. J., Scahill, R. I., Owen, G., Durr, A., Leavitt, B. R., Roos, R. A., et al. (2013). Predictors of phenotypic progression and disease onset in premanifest and early-stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *Lancet Neurol.* 12, 637–649. doi: 10.1016/S1474-4422(13)70088-7
- Tang, X., Oishi, K., Faria, A. V., Hillis, A. E., Albert, M. S., Mori, S., et al. (2013). Bayesian parameter estimation and segmentation in the multi-atlas random orbit model. *PLoS ONE* 8:e65591. doi: 10.1371/journal.pone.0065591
- Turner, R. S., Kenyon, L. C., Trojanowski, J. Q., Gonatas, N., and Grossman, M. (1996). Clinical, neuroimaging, and pathologic features of progressive nonfluent aphasia. *Ann. Neurol.* 39, 166–173. doi: 10.1002/ana.410390205
- van Rikxoort, E. M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M. A., et al. (2010). Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. *Med. Image Anal.* 14, 39–49. doi: 10.1016/j.media.2009.10.001
- Wang, H., Pouch, A., Takabe, M., Jackson, B., Gorman, J., Gorman, R., et al. (2013). Multi-atlas segmentation with robust label transfer and label fusion. *Inf. Process. Med. Imaging* 23, 548–559. doi: 10.1007/978-3-642-38868-2_46
- Wang, L., Beg, F., Ratnanather, T., Ceritoglu, C., Younes, L., Morris, J. C., et al. (2007). Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the Alzheimer type. *IEEE Trans. Med. Imaging* 26, 462–470. doi: 10.1109/TMI.2006.887380
- Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23, 903–921. doi: 10.1109/TMI.2004.828354
- Wilson, S. M., Demarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Miller, B. L., et al. (2016). Variable disruption of a syntactic processing network in primary progressive aphasia. *Brain* 139:11. doi: 10.1093/brain/aww218
- Wu, D., Ma, T., Ceritoglu, C., Li, Y., Chotiyanonta, J., Hou, Z., et al. (2016). Resource atlases for multi-atlas brain segmentations with multiple ontology levels based on T1-weighted MRI. *Neuroimage* 125, 120–130. doi: 10.1016/j.neuroimage.2015.10.042
- Zhang, X., Mormino, E. C., Sun, N., Sperling, R. A., Sabuncu, M. R., and Yeo, B. T. (2016). Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc. Natl. Acad. Sci. U.S.A.* 113, e6535–e6544. doi: 10.1073/pnas.1611073113
- Zhang, Y., Tartaglia, M. C., Schuff, N., Chiang, G. C., Ching, C., Rosen, H. J., et al. (2013). MRI signatures of brain macrostructural atrophy and microstructural degradation in frontotemporal lobar degeneration subtypes. *J. Alzheimers. Dis.* 33, 431–444. doi: 10.3233/JAD-2012-121156
- Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D. W., Amos, C., et al. (1997). Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat. Genet.* 15, 62–69. doi: 10.1038/ng0197-62

Conflict of Interest Statement: SM and MM own AnatomyWorks. SM is its CEO. This arrangement is being managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Faria, Liang, Miller and Mori. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Hierarchical Bayesian Model for the Identification of PET Markers Associated to the Prediction of Surgical Outcome after Anterior Temporal Lobe Resection

Sharon Chiang^{1,2}, Michele Guindani³, Hsiang J. Yeh⁴, Sandra Dewar⁴, Zulfi Haneef⁵, John M. Stern⁴ and Marina Vannucci^{1*}

¹ Department of Statistics, Rice University, Houston, TX, United States, ² School of Medicine, Baylor College of Medicine, Houston, TX, United States, ³ Department of Statistics, University of California, Irvine, Irvine, CA, United States, ⁴ Department of Neurology, University of California, Los Angeles, Los Angeles, CA, United States, ⁵ Department of Neurology, Baylor College of Medicine, Houston, TX, United States

OPEN ACCESS

Edited by:

P. Thomas Fletcher,
University of Utah, United States

Reviewed by:

Hyunjin Park,
Sungkyunkwan University,
South Korea
Lei Wang,
Northwestern University, United States

*Correspondence:

Marina Vannucci
marina@rice.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 26 April 2017

Accepted: 17 November 2017

Published: 05 December 2017

Citation:

Chiang S, Guindani M, Yeh HJ,
Dewar S, Haneef Z, Stern JM and
Vannucci M (2017) A Hierarchical
Bayesian Model for the Identification
of PET Markers Associated to the
Prediction of Surgical Outcome after
Anterior Temporal Lobe Resection.
Front. Neurosci. 11:669.
doi: 10.3389/fnins.2017.00669

We develop an integrative Bayesian predictive modeling framework that identifies individual pathological brain states based on the selection of fluoro-deoxyglucose positron emission tomography (PET) imaging biomarkers and evaluates the association of those states with a clinical outcome. We consider data from a study on temporal lobe epilepsy (TLE) patients who subsequently underwent anterior temporal lobe resection. Our modeling framework looks at the observed profiles of regional glucose metabolism in PET as the phenotypic manifestation of a latent individual pathologic state, which is assumed to vary across the population. The modeling strategy we adopt allows the identification of patient subgroups characterized by latent pathologies differentially associated to the clinical outcome of interest. It also identifies imaging biomarkers characterizing the pathological states of the subjects. In the data application, we identify a subgroup of TLE patients at high risk for post-surgical seizure recurrence after anterior temporal lobe resection, together with a set of discriminatory brain regions that can be used to distinguish the latent subgroups. We show that the proposed method achieves high cross-validated accuracy in predicting post-surgical seizure recurrence.

Keywords: Bayesian hierarchical model, positron emission tomography (PET), spatially-informed prior, mixture model, variable selection, Pólya-Gamma distribution

1. INTRODUCTION

In the era of precision medicine, in order to deliver targeted therapies for neurological disorders, the development of methods to identify reliable and quantifiable biomarkers that are associated to individual clinical outcomes has become of paramount importance (Insel and Cuthbert, 2015). Temporal lobe epilepsy (TLE) is the most common form of adult epilepsy and the most common epilepsy refractory to anti-epileptic drugs. Surgery provides an effective treatment for many patients, yielding a seven-fold greater probability of seizure freedom 1 year after surgery than patients treated with medications alone (Wiebe et al., 2001). Despite its effectiveness, 30–50% of patients with TLE continue to experience seizures after surgery (Spencer et al., 2005; de Tisi et al., 2011).

As interictal ^{18}F -fluorodeoxyglucose positron emission tomography (FDG-PET) has traditionally been used for seizure focus localization (Wieser, 2004), there is substantial interest in identifying methods that utilize PET for prediction of post-surgical seizure relief (Willmann et al., 2007). Mesial TLE with hippocampal sclerosis is defined by the presence of neuronal cell loss and gliosis in the CA1 region and endfolium of the hippocampus, a particular part of the temporal lobe (Wieser, 2004). Therefore, prediction of post-surgical outcome using FDG-PET has traditionally focused on specific regions selected a priori within the temporal lobe (Dupont et al., 2000; Lin et al., 2007). Such studies have demonstrated predictive value of FDG-PET for identifying mesial TLE. Increasing evidence, however, points at TLE as a network disorder that includes abnormality distributed beyond the temporal lobe, rather than a focal disorder (Bonilha et al., 2005; McDonald et al., 2008; Mueller et al., 2010; Chiang and Haneef, 2014). This suggests that whole-brain statistical approaches may allow for improved identification of quantifiable features from neuroimaging data that can be reliably associated with individual clinical outcomes and improve clinical decision-making.

Traditional predictive modeling approaches for neuroimaging data have included the use of pattern recognition techniques, such as Linear Discriminant Analysis (Haynes and Rees, 2005), Support Vector Machines (Mitchell et al., 2004; LaConte et al., 2005) and Bayesian classifiers (Burge et al., 2009; Arribas et al., 2010). In particular, pattern recognition techniques have been used with varying success to predict post-surgical outcome in TLE, ranging from 50 to 75% accuracy using random forests (Njiwa et al., 2015) to 70% accuracy using elastic net and support vector machines (Munsell et al., 2015). Recently, Bayesian spatial hierarchical models have also been used to improve prediction accuracy from PET data by borrowing strength from spatial correlations between neighboring voxels/regions (Derado et al., 2013). Several approaches for dynamic PET data have also been proposed. O'Sullivan (2006) and Jiang and Ogden (2008), for example, utilize mixture modeling and conditional autoregressive models to incorporate spatial information into PET analysis, while other work has used functional principal components (Jiang et al., 2009) or wavelets (Millet et al., 2000; Alpert et al., 2006) to analyze dynamic PET signal. Although each of these approaches represents an important advance in neuroimaging methods development, these methods do not quantify the relative importance of selected regions, which may impact the effectiveness of related clinical decisions. Recently, Bayesian scalar-on-image regression methods have been proposed that associate a univariate outcome to massive multi-dimensional image predictors, particularly for functional magnetic resonance imaging (fMRI) data (van Gerven et al., 2010; Goldsmith et al., 2014; Li et al., 2015). All the methods above, however, do not consider the heterogeneity of the population of individuals and implicitly assume that, given a set of discriminatory regions, their association to the outcome is the same across the population. In reality, however, the strength of the association can vary across subgroups of subjects.

In this paper, we develop a statistical model to identify whole-brain biomarkers from PET imaging which are associated

to the prediction of post-surgical seizure recurrence following anterior temporal lobe resection. Post-surgical seizure recurrence is thought to be due to incomplete resection of the epileptogenic zone, which is defined as the area of cortex necessary and sufficient for initiating seizures, and whose removal is necessary for seizure abolition (Lüders et al., 1993). While the epileptogenic zone was historically thought to arise from discrete focal sources, more recent evidence suggests that seizure activity arises from the activity of epileptogenic cortical networks that are distributed beyond the temporal lobe (Frasnaszczuk et al., 1994; Frasnaszczuk and Bergey, 1998; Baccalá et al., 2004; Worrell et al., 2004, 2008; Jirsch et al., 2006; Kramer et al., 2008; Chiang et al., 2017a). Patients with different epileptogenic zone configurations are expected to exhibit different likelihoods of post-surgical seizure recurrence. Different epileptogenic zone configurations are also expected to produce different interictal metabolic patterns of FDG uptake, due to the effect of epileptogenic activity on neuronal loss and postictal metabolic depression (Lüders, 2008). The epileptogenic zone, however, cannot be identified pre-operatively, due to the fact that parts of an epileptogenic lesion may not be implicated in the preoperatively recorded seizure, but will continue to generate seizures post-operatively if not resected (Rosenow and Lüders, 2001). In our model formulation, we look at the observed PET brain measurements as the phenotypic manifestation of latent individual pathological states that are assumed to vary across the population. We then factor the joint distribution of the data into the product of two conditionally independent submodels, an *outcome model* that relates the post-surgical outcome to the latent states, and a *measurement model* that relates those latent states to the observed brain measurements. For the latter, we employ mixture models for clustering and variable selection priors that capture spatial correlation among neighboring brain regions. This allows us to cluster subjects into subgroups with different latent pathological states, while simultaneously identifying discriminatory brain regions that characterize the subgroups. A logistic regression model relates the latent states to the binary clinical outcome.

We apply the proposed approach to PET data collected at the University of California, Los Angeles (UCLA) as part of a clinical study on post-surgical outcomes in temporal lobe epilepsy. We also incorporate into the analysis connectivity information from resting-state functional magnetic resonance imaging (fMRI) data, to inform the selection of discriminatory brain regions. Integrative models that take into account neuroscientific information from multi-modal data sources, such as fMRI, electroencephalography (EEG), or diffusion tensor imaging (DTI), are a pressing issue in the field, in particular given the limited number of patient samples collected in many neuroimaging experiments (Bowman et al., 2012; Hinne et al., 2014; Jorge et al., 2014). Bayesian inference provides a powerful way to incorporate multi-modal imaging into computational anatomy by inclusion through network priors. In our case study, we identify a subgroup of patients at high risk for post-surgical seizure recurrence, together with several discriminatory brain regions which can be used in clinical decisions to maximize interventional treatments. Furthermore, we show that the proposed approach achieves high cross-validated accuracy in

predicting post-surgical seizure recurrence. Further assessment of the performance of our method is performed in the Supplementary Material by conducting a comparison study on synthetic data against multi-step approaches and/or approaches that do not condition on latent states.

2. MATERIALS AND METHODS

2.1. Case Study on Temporal Lobe Epilepsy

Positron emission tomography (PET) is a type of *in vivo* nuclear medicine imaging which uses radioactive tracers to quantify tissue function. The subject is injected with a positron-emitting isotope, such as ^{18}F -FDG, and a PET image is reconstructed of the isotope concentration based on the incidence of gamma rays from the positron-electron annihilation. In this work, we analyze data on 19 adult patients with drug resistant MTLE and radiological evidence of unilateral hippocampal sclerosis (MTLE-HS), who underwent pre-operative interictal ^{18}F -FDG PET and anterior temporal lobe resection (ATL) at the UCLA Seizure Disorder Center between 2007 and 2012. Patients were identified from the UCLA video-EEG Epilepsy Monitoring Unit. As the primary outcome of this study was post-operative seizure freedom after epilepsy surgery, a healthy control group was not obtained as anterior temporal lobe resections are not performed in healthy patients without indication for surgery. Diagnostic evaluation included video-EEG monitoring, high resolution MRI, interictal ^{18}F -FDG PET, and neuropsychological testing. PET/CT scans were acquired on a Siemens Biograph scanner as described in Kerr et al. (2013). Patients fasted for at least 6 h before each scan except for water and medications. Patients received 0.14 mCi/kg of ^{18}F -FDG intravenously and rested in a quiet, dimly lit room with their eyes open during the ensuing 40 min uptake period with concomitant EEG monitoring to confirm interictal status. The iterative reconstruction program Ordered Subset Expectation Maximization (OSEM) available through NeuroQ (Syntermed, GA, USA) was used for reconstruction of PET images. Iterative reconstruction was halted after two iterations using eight subsets. CT images were reconstructed using filtered back projection at 3.4 mm axial intervals to match the slice separation of the PET data, and used for attenuation correction. Post-operative seizure freedom was assessed 1 year after surgery and classified as either seizure-free (SF; Engel Class 1) or not seizure-free (NSF; Engel Class 2–4). The binary outcome of complete freedom from disabling seizures (Engel Class 1) is the standard primary outcome of interest evaluated in epilepsy surgery treatment trials (Engel et al., 2012). The use of this primary outcome in epilepsy surgery trials results from the goal of epilepsy surgery, which is complete seizure freedom. In addition, we have available resting state fMRI (rs-fMRI) data collected on a separate set of 32 TLE patients recruited from the UCLA Seizure Disorder Center. Details on fMRI data are described in section 3.1.

2.2. PET Pre-processing

In PET studies, the quantity that is clinically assessed is a scalar rate of regional glucose uptake, based on a method described by Sokoloff et al. (1977). This quantity is then normalized

relative to an internal reference standard, such as the whole-brain or cerebellar activity, and compared to the expected level for a reference normal subject (Silverman et al., 2008). The cerebellum is commonly used as the reference PET region for diseases of interest in which the cerebellum is thought not to be affected, such as diseases involving diffuse forebrain involvement. However, cerebellar atrophy is a very well described phenomenon in epilepsy, and is moreover associated with longer duration of epilepsy as well as younger age of epilepsy onset (Sandok et al., 2000). Given that the cerebellum could be more involved in epilepsy than traditionally thought (Fountas et al., 2010), we chose to normalize by the average whole-brain uptake rather than by the cerebellum. The assessed quantity therefore provides a measure of the level of metabolic activity in each region, relative to that expected in healthy controls. Uptake levels may be quantified on the single-voxel level or based on the mean uptake within fixed regions of interest. However, because single-voxel measurements are adversely affected by noise, the use of regions of interest (ROIs) in FDG-PET has been suggested as a more robust alternative for clinical practice (Wahl et al., 2009), which additionally facilitates standardized comparisons of affected regions across subjects. NeuroQ (Syntermed, GA, USA) is a software approved by the FDA in 2004 for quantitative assessment of brain PET imaging in clinical practice and was used to pre-process PET images. Following transformation into template Montreal Neurological Institute (MNI) space by a method previously described by Tai et al. (1997), images were segmented into 47 predefined regions of interest using a predefined NeuroQ atlas (Silverman and Melega, 2004; Ercoli et al., 2012) which has been previously considered for quantitative assessment of PET data in clinical practice (Smith et al., 2007; McCallum et al., 2010; Torosyan and Silverman, 2012; Kerr et al., 2013; Akdemir et al., 2014). ROI abbreviations are listed in the Supplementary Material. Pre-processing consisted of scalp removal, rigid registration to a reference PET image to correct for head tilt, and reformatting of transaxial slices to fit normal template transaxial slices using 10 iterations. Maximization of the mutual information between the image volumes was used to identify the registration parameter. A mean count was calculated in each ROI, normalized by the whole-brain counts and standardized relative to the mean and standard deviation of each ROI among healthy controls. Greater magnitude of PET image intensities indicate more pathological levels of metabolic activity, with positive values indicating greater levels of hypermetabolism (i.e., greater metabolism than in healthy controls) and negative values indicating greater levels of hypometabolism (i.e., lower metabolism than in healthy controls). Consequently, different patterns of nonzero signal characterize different pathological patterns of metabolic activity. Imaging patterns of hyper- and hypometabolism were of interest in this study rather than the raw PET signal intensities, due to the association of hypermetabolic activity with epileptic activity. Lateralized ROIs were recoded from left and right to ipsilateral or contralateral with respect to the side of subsequent resection. A histogram of the normalized and standardized PET image intensities (Figure not shown) indicated a bell-shaped, unimodal, and fairly symmetrical distribution, with a skewness of -0.39 .

2.3. Statistical Model

Let \mathbf{X}_i denote the $R \times 1$ vector of normalized PET image predictors on R brain regions of interest (ROIs) for subject i and let Y_i denote the corresponding post-surgical outcome, for $i = 1, \dots, n$. We propose to study the association between the PET image predictors and the outcome via a *measurement error* model formulation. As described above, non-zero values of \mathbf{X} indicate the level of PET metabolic activity, with different non-zero intensity patterns indicating different pathological imaging profiles. Accordingly, we assume that the brain's observed profile of metabolic activity is the manifestation of a latent (i.e., unobserved) pathological state. In epilepsy, the latent pathological state represents the configuration of metabolic activity in regions implicated in the underlying epileptogenic zone, which is in turn associated to post-surgical seizure recurrence. Here, we assume a finite number of pathological states due to the expected modular organization of the brain, which is generally decomposed into a finite number of submodules (Meunier et al., 2010). Let η_i denote the latent pathological state of subject i . Then, we propose to factor the joint distribution of $\mathbf{Z}_i = \{Y_i, \mathbf{X}_i\}_{i=1}^n$ into the product of two conditionally independent sub-models: an *outcome model* that relates the clinical outcome to the latent pathological state, and a *measurement model* that relates the latent pathological state to the observed imaging data. Therefore, we consider a non-differential measurement error model, i.e., conditionally upon the latent pathological state η_i , the observed surrogate \mathbf{X}_i contains no additional information on the outcome Y_i (Richardson and Gilks, 1993), $f(Y_i|\eta_i, \mathbf{X}_i) = f(Y_i|\eta_i)$. This model allows us to capture the current understanding in epilepsy that failure of temporal lobe resection results most likely from incomplete resection of the epileptogenic zone (Ryvlin and Kahane, 2005). In other words, if the true epileptogenic zone were known, data contained in the PET image \mathbf{X}_i would not provide any additional information on the probability of post-operative seizure recurrence Y_i . Thus,

$$f(\mathbf{Z}|\boldsymbol{\eta}) = \prod_{i=1}^n f(Y_i|\eta_i)f(\mathbf{X}_i|\eta_i), \quad (1)$$

where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$. We specify the measurement model in Equation (1) as a mixture model with variable selection. Subgroups of patients with different epileptogenic zone configurations may be expected to exhibit different risks of post-surgical seizure recurrence. We therefore specify the outcome model in Equation (1) as a logistic regression model that relates the latent states to the binary clinical outcome. There is extensive literature on the use of measurement error models to model data in which risk factors related to the observed disease or treatment status are unknown, but where surrogate measures, which provide information on the unobserved risk factor, are recorded. A review of measurement error models may be found in Carroll et al. (2006). With respect to existing literature, our model formulation allows us to cluster subjects into subgroups with different latent pathological states, i.e., different epileptogenic zone configurations, while simultaneously identifying discriminatory brain regions. In the selection, we also

capture spatial correlation among neighboring brain regions via a spatial prior, as described in section 2.3.3.

2.3.1. Clustering via Finite Mixture Models

We envision that a subject may be characterized by one of K possible pathological states. Let η_i denote a latent random variable that identifies the state of the i -th subject, $i = 1, \dots, n$. We assume that the latent individual state η_i takes values in $\{1, \dots, K\}$, where one of the states can be assumed as reference. Then, for each subject i we define an allocation vector $\boldsymbol{\rho}_i = (I(\eta_i = 1), \dots, I(\eta_i = K - 1))$, where $I(\eta_i = k)$ indicates that subject i has latent state k , i.e., $I(\eta_i = k) = 1$ if $\eta_i = k$, and 0 otherwise. Then, for the measurement model in Equation (1), we choose a finite mixture model that clusters the n subjects into K possible subgroups as

$$f(\mathbf{X}_i|\eta_i, \boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k f(\mathbf{X}_i|\boldsymbol{\theta}_k),$$

with $\eta_i = k$ if subject i belongs to cluster k and $P[\eta_i = k] = \pi_k$. The η_i 's are assumed to be independent and identically distributed, so that $\boldsymbol{\eta} \sim \text{Multinomial}(1; \pi_1, \dots, \pi_K)$. We assume a Dirichlet prior on the mixture weights, $p(\boldsymbol{\pi}) = \text{Dirichlet}(\alpha_1, \dots, \alpha_K)$. We consider the case where $f(\mathbf{x}_i|\boldsymbol{\theta}_k)$ is Gaussian with parameters $\boldsymbol{\theta}_k = (\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, so that

$$f(\mathbf{X}_i|\boldsymbol{\theta}_k) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (2)$$

with $k = 1, \dots, K$. The component-specific mean $\boldsymbol{\mu}_k$ models the latent state specific random effect and characterizes the mean metabolic profile for subjects with latent state k , whereas $\boldsymbol{\Sigma}_k$ is a variance-covariance matrix that captures general relationships among regions for subjects with latent state k . In summary, the likelihood function for the measurement model is

$$L(\mathbf{X}|\boldsymbol{\eta}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \prod_{k=1}^K (2\pi)^{-n_k R/2} |\boldsymbol{\Sigma}_k|^{-n_k/2} \times \exp \left\{ -\frac{1}{2} \sum_{\{i: \eta_i=k\}} (\mathbf{X}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_k) \right\},$$

with n_k denoting the number of subjects in cluster k . Here we assume diagonal variance-covariance matrices $\boldsymbol{\Sigma}_k = \text{diag}(\sigma_{k,1}, \dots, \sigma_{k,R})$. Even though we make this simplifying assumption at this stage of the hierarchy, our proposed model is still able to capture structural dependencies via the specification of the prior model for the mean components in Equation (4) that we describe in section 2.3.3.

2.3.2. Association with the Treatment Outcome

The outcome model in Equation (1) allows the prediction of the subject-specific outcomes based on the patients' individual latent pathological state η_i . We can relate the latent states with the outcome of interest by employing a generalized linear model. In general, we may have available a vector of baseline covariates \mathbf{U}_i

for subject i . Since the post-surgical outcome is binary, we can then use a logistic regression model

$$p(Y_i = y_i | \eta_i, \boldsymbol{\beta}) = \frac{\exp(\boldsymbol{\xi}_i^T \boldsymbol{\beta})^{y_i}}{1 + \exp(\boldsymbol{\xi}_i^T \boldsymbol{\beta})}, \quad (3)$$

with $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{K-1}, \boldsymbol{\beta}_U)$ and $\boldsymbol{\xi}_i = (1, \boldsymbol{\rho}_i, \mathbf{U}_i)$, where $\boldsymbol{\beta}_U$ is the vector of corresponding regression coefficients for $\mathbf{U} = \{\mathbf{U}_i\}_{i=1}^n$. Here, β_k , $k = 1, \dots, K-1$ captures the “risk” associated to latent state k relative to the baseline latent state. Each β_k can be interpreted as the log-odds of the outcome for subjects in state k relative to subjects in the reference state, and β_0 as an intercept term yielding the log-odds of the outcome for subjects in the reference state.

The analytically intractable form of the likelihood function using a logit link is known to pose challenges for Bayesian inference in logistic regression models. To address this and to improve posterior sampling, we employ the data augmentation approach recently devised by Polson et al. (2013). Let ω be a Pólya-Gamma random variable, $\omega \sim \text{PG}(b, c)$, with parameters $b > 0$ and $c \in \mathbb{R}$,

$$\omega \stackrel{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-1/2)^2 + c^2/4\pi^2},$$

where g_k are independently distributed as $\text{Gamma}(b, 1)$. Augmentation with a Pólya-Gamma random variable allows for the likelihood contribution of the i th observation to be written as

$$\begin{aligned} L_i(\boldsymbol{\beta}) &= \frac{\exp(\boldsymbol{\xi}_i^T \boldsymbol{\beta})^{y_i}}{1 + \exp(\boldsymbol{\xi}_i^T \boldsymbol{\beta})} \\ &= \frac{1}{2} \exp(\kappa_i \boldsymbol{\xi}_i^T \boldsymbol{\beta}) \int_0^\infty \exp\left(-\frac{\omega_i (\boldsymbol{\xi}_i^T \boldsymbol{\beta})^2}{2}\right) p(\omega_i) d\omega_i, \end{aligned}$$

where $\kappa_i = y_i - 1/2$, for $\omega_i \sim \text{PG}(1, 0)$. Combining all n terms then gives the following convenient representation for the conditional likelihood in $\boldsymbol{\beta}$, given $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$:

$$L(\boldsymbol{\beta} | \boldsymbol{\eta}, \boldsymbol{\omega}) \propto \exp\left\{-\frac{1}{2}(\mathbf{z} - \boldsymbol{\Xi}\boldsymbol{\beta})^T \boldsymbol{\Omega}(\mathbf{z} - \boldsymbol{\Xi}\boldsymbol{\beta})\right\},$$

where $\mathbf{z} = (\kappa_1/\omega_1, \dots, \kappa_m/\omega_m)$, $\kappa_i = y_i - 1/2$, $\boldsymbol{\Omega} = \text{diag}(\omega_1, \dots, \omega_n)$, $\boldsymbol{\Xi}$ is the $n \times K$ matrix $\boldsymbol{\Xi} = (\boldsymbol{\xi}_1^T, \dots, \boldsymbol{\xi}_n^T)$, $\boldsymbol{\xi}_i = (1, \rho_{i,1}, \rho_{i,2}, \dots, \rho_{i,K-1})$, and $\rho_{i,k} = I(\eta_i = k) \forall k = 1, \dots, K-1$. See Polson et al. (2013) for details. We complete the model by imposing a conjugate prior on $\boldsymbol{\beta}$, $p(\boldsymbol{\beta}) = \text{N}(\mathbf{m}_\beta, V_\beta)$, where \mathbf{m}_β and V_β denote the prior mean and covariance, respectively.

2.3.3. Spatially-Informed Selection Prior

Not all brain regions are expected to provide information about the subgroup structure of the subjects, in which case the inclusion of non-discriminatory regions in model (Equation 2) may obscure the discovery of true groups. One way to address this issue is through variable selection for clustering. Let $\boldsymbol{\gamma} \in \{0, 1\}^R$ denote a binary vector, where $\gamma_j = 1$ if region j is discriminatory, and $\gamma_j = 0$ otherwise, $\forall j = 1, \dots, R$. We follow Hoff (2006) and

identify discriminatory brain regions by imposing spike-and-slab priors on the random effects $\boldsymbol{\mu}_k = (\mu_{k,1}, \dots, \mu_{k,R})$. Given the spatial contiguity in neuronal glucose consumption, we allow for spatial smoothness among neighboring regions by specifying the slab portion of the prior as an intrinsic conditional autoregressive (ICAR) prior distribution (Banerjee et al., 2014). Our prior on $\mu_{k,j}$ can be written as

$$\begin{aligned} p(\mu_{k,j} | \gamma_j, \boldsymbol{\mu}_{k,\setminus j}) &= \gamma_j \text{N}\left(\frac{\sum_{j'=1}^R S_{jj'} \mu_{k,j'}}{\sum_{j'=1}^R S_{jj'}}, \frac{c_k}{\sum_{j'=1}^R S_{jj'}}\right) \\ &\quad + (1 - \gamma_j) \delta_0(\mu_{k,j}), \end{aligned} \quad (4)$$

where δ_0 denotes a spike at zero, S is an $R \times R$ symmetric neighborhood matrix, with $S_{jj'} = 1$ if regions j and j' are neighbors, and $S_{jj'} = 0$ otherwise, and where $\boldsymbol{\mu}_{k,\setminus j}$ denotes all elements of $\boldsymbol{\mu}_k$ except the j th element. We also impose priors on the diagonal elements of Σ_k in Equation (2) and allow for separate variances for the discriminatory and non-discriminatory regions. In particular, for the parameters corresponding to $\gamma_j = 1$, we have $\sigma_{k,j} = \sigma_k \sim \text{IG}(a_k, b_k)$ for all k , while for $\gamma_j = 0$ we impose $\sigma_{k,j} = \sigma_0 \sim \text{IG}(a_0, b_0)$. Finally, in specifying the prior on the selection indicators, $\boldsymbol{\gamma}$, we allow for external information on the network structure of the brain, for example on connectivity between regions, to be incorporated in the model by imposing an Ising prior of the type

$$p(\boldsymbol{\gamma}) \propto \exp\left\{e \mathbf{1}_R^T \boldsymbol{\gamma} + f \boldsymbol{\gamma}^T S \boldsymbol{\gamma}\right\}, \quad (5)$$

with S denoting the neighborhood matrix. If a connection exists between two regions j and j' , then selection of one region j (i.e., $\gamma_j = 1$) leads to an increased probability that region j' will also be selected (i.e., $\gamma_{j'} = 1$). The hyperparameter $e \in (-\infty, \infty)$ controls the sparsity of the model and represents the prior expected number of discriminatory regions. The hyperparameter $f > 0$ is a smoothing parameter which represents the prior probability of a region being discriminatory given that its neighbors are too. In particular, if a region has no neighbors, then its prior distribution reduces to an independent Bernoulli distribution with probability $\exp(e)/(1 + \exp(e))$, which is a common prior assumed in Bayesian variable selection literature in the case of independent variables.

The prior construction (Equations 4, 5) allows for sparsity while promoting spatial contiguity in the selection. The ICAR prior, in particular, ensures that each cluster's mean metabolic PET profile varies smoothly in space, as each $\mu_{k,j}$ is modeled to vary around the mean of its neighbors, with variance inversely scaled by the number of neighbors. Spatial prior constructions have been used extensively in neuroimaging applications, particularly with fMRI data (Smith and Fahrmeir, 2007; Zhang et al., 2014; Li et al., 2015).

2.3.4. MCMC Algorithm

In order to sample from the joint posterior distribution of all parameters $(\{\sigma_k\}_{k=1}^K, \sigma_0, \boldsymbol{\eta}, \boldsymbol{\pi}, \boldsymbol{\gamma}, \{\boldsymbol{\mu}_k\}_{k=1}^K, \boldsymbol{\beta}, \boldsymbol{\omega})$, we employ Markov Chain Monte Carlo (MCMC) methods that combine

variable selection stochastic search algorithms that use *add-delete-swap* moves (Savitsky et al., 2011) with efficient Pólya-Gamma sampling for logit models (Polson et al., 2013). We provide full details of the implementation in the Supplementary Material.

2.3.5. Prediction

An important characteristic of our model formulation is that it allows for prediction of the outcome status y_f of a future observation \mathbf{x}_f , based on the training data $\{\mathbf{X}, \mathbf{Y}\}$. In the context of pre-surgical evaluation for epilepsy surgery, this allows for probabilistic, patient-specific predictive estimates of the patient's probability of surgery benefit, in order to assist with clinical decision-making. The predictive distribution is given by

$$p(y_f|\mathbf{x}_f, \mathbf{X}, \mathbf{Y}) = \int_{\beta} \sum_{\eta_f \in \{1, \dots, K\}} p(y_f|\eta_f, \beta) p(\beta|\mathbf{X}, \mathbf{Y}) p(\eta_f|\mathbf{x}_f) d\beta, \quad (6)$$

and cannot be computed in closed form. Following standard Bayesian techniques, these steps can be employed to simulate from Equation (6):

1. Sample T values of $\mu_k, \Sigma_k, \pi_k, \beta$ from the joint posterior, using the MCMC algorithm as described in the Supplementary Material.
2. For each posterior draw, $t = 1, \dots, T$:
 - Sample $m \geq 1$ values of $\eta_f \in \{1, \dots, K\}$ from $p(\eta_f|\mathbf{x}_f)$, where $\forall k = 1, \dots, K$

$$p(\eta_f = k|\mathbf{x}_f) \propto p(\mathbf{x}_f|\eta_f = k) p(\eta_f = k) = p(\mathbf{x}_f|\mu_k^{(t)}, \Sigma_k^{(t)}) \pi_k^{(t)}.$$
 - For each sampled value of η_f , sample a value of $y_f \in \{0, 1\}$ from $p(y_f|\eta_f, \beta^{(t)})$.

The posterior predictive probability $p(y_f = 1|\mathbf{x}_f, \mathbf{X}, \mathbf{Y})$ can then be estimated as the proportion of posterior predictive samples for which $y_f = 1$. In the analyses of this paper, given the limited number of samples available, which does not allow a meaningful splitting of the data into training and validation, we implemented cross-validation prediction via the importance-sampling approach, as proposed by Gelfand (1996), and write the cross-validation predictive density for the i th observation as

$$p(Y_i = 1|\mathbf{X}, \mathbf{Y}_{-i}) = \int_{\eta, \beta} p(Y_i = 1|\mathbf{X}, \mathbf{Y}_{-i}, \eta, \beta) p(\eta, \beta|\mathbf{X}, \mathbf{Y}_{-i}) d\beta d\eta$$

where we use $p(\eta, \beta|\mathbf{X}, \mathbf{Y})$ as an importance sampling density for $p(\eta, \beta|\mathbf{X}, \mathbf{Y}_{-i})$, and \mathbf{Y}_{-i} denotes the non-hold out outcomes. Specific details on implementation are provided in the Supplementary Material.

3. RESULTS

We now apply the proposed model to the data we have available from the University of California, Los Angeles Seizure Disorder Center, where we illustrate the utility of our proposed model for predicting a post-surgical outcome among MTLE-HS patients from pre-surgical FDG-PET imaging.

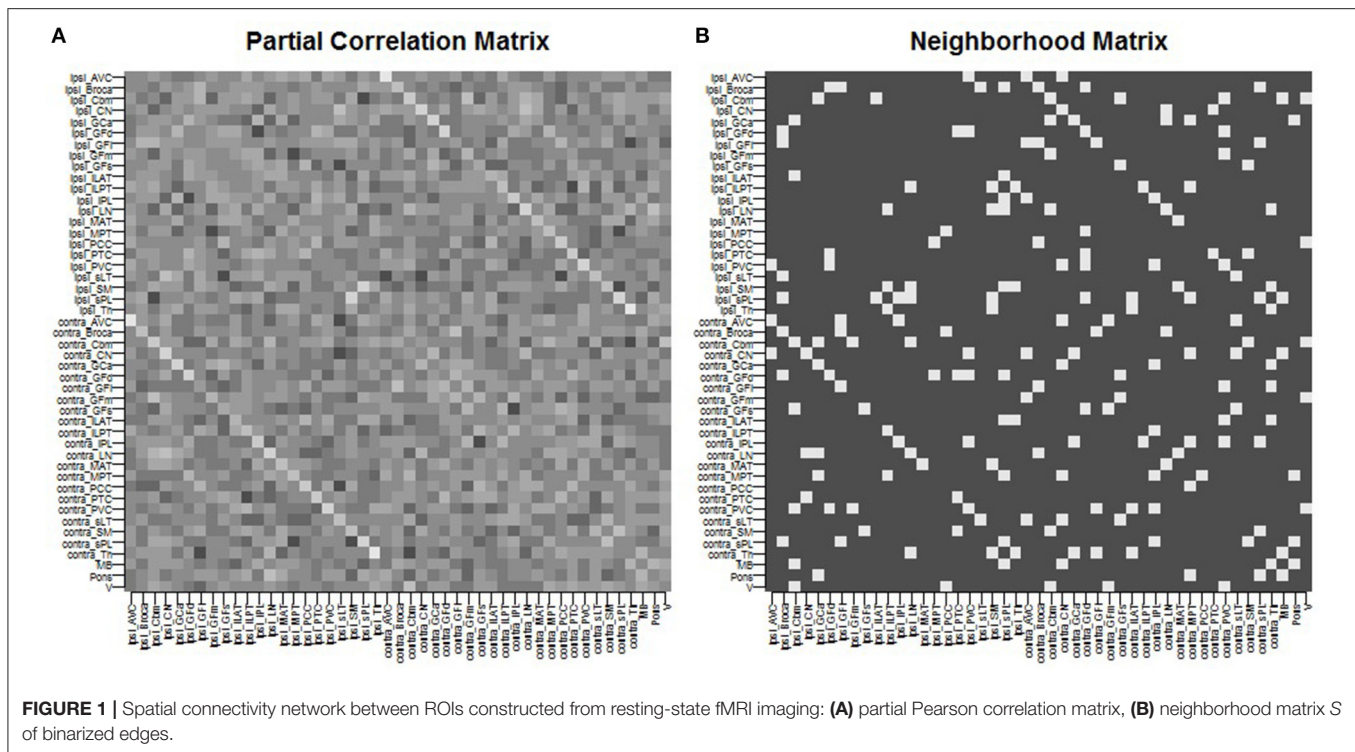
3.1. Prior Connectivity Network

For this analysis, we allowed the spatial network prior Equation (5) to capture information on functional connectivity between the ROIs, which we estimated based on resting-state fMRI data (rs-fMRI), collected on a separate set of 32 unilateral temporal lobe epilepsy patients from the UCLA Seizure Disorder Center. Rs-fMRI was performed on the subjects after a comprehensive epilepsy surgery evaluation and prior to epilepsy surgery. None of the patients had a seizure in the 24 h preceding the imaging or had seizures during the study, as confirmed by the simultaneous EEG obtained during fMRI. There were no post-surgical outcome data available for these patients. External or historical information is often used to formulate priors in Bayesian analysis. There is extensive literature which demonstrates the general replicability of Pearson correlation estimation of functional connectivity from rs-fMRI in temporal lobe epilepsy (Centeno and Carmichael, 2014). Furthermore, despite increasing evidence that functional connectivity is dynamic (Honey et al., 2009; Ma et al., 2014; Chiang et al., 2016), recent research indicates a large proportion of the information present in functional connectivity is contained in static estimates (Chiang et al., 2017b).

We give full details of the rs-fMRI data and the process to estimate a connectivity network in the Supplementary Material. In brief, preprocessing of rs-fMRI imaging was performed using FSL (fMRIB Software Library) version 5.0.7 (Oxford, United Kingdom, www.fmrib.ox.ac.uk/fsl). Functional connectivity between the 47 ROIs was estimated by placing a 6-mm spherical seed in Montreal Neurological Institute (MNI) space at the location of each of the 47 ROIs. Each patient's fMRI BOLD image was registered to the patient's high-resolution structural image using FLIRT (FMRIB's Linear Image Registration Tool) (Jenkinson et al., 2002; Greve and Fischl, 2009), and the high-resolution structural was registered to the standard MNI space using FNIRT (FMRIB's Non-linear Image Registration Tool) (Andersson et al., 2007). Functional connectivity between each pair of nodes was computed as the partial Pearson correlation between the averaged regional time-series. This provided us with a 47×47 correlation matrix. An edge was then considered as included in the connectivity network if the correlation between the regions exceeded a given threshold. The threshold was chosen so that the average number of neighbors for each region was approximately 5, yielding a connectivity structure close to a three-dimensional lattice. The resulting network was used as the neighborhood matrix S in the specification of the MRF prior (Equation 5) on γ and also in the ICAR prior (Equation 4) on the slab portion of the prior on $\mu_{k,j}$. The estimated functional connectivity matrix and resulting neighborhood matrix S are shown in **Figure 1**. We observe several known connectivity relationships, including functional connectivity between regions in the brainstem (midbrain, pons); between the primary and associative visual cortices; between the cerebellar hemispheres and vermis; and between ipsilateral and contralateral ROIs (Quigley et al., 2003).

3.2. Biomarker Selection and Clustering

In our approach to model fitting we consider a grid of values of K to find the number of states K yielding the best model fit



that also provides improved clinical interpretability. For the study of this paper, model fit for each value of K for $K = 2, \dots, 6$ was assessed using the deviance information criterion (DIC) of Spiegelhalter et al. (2002). We found that $K = 2$ clusters allowed for a parsimonious model permitting meaningful clinical characterization of high- and low-risk patients, with minimal to no further improvement in the DIC for larger values of K . This result was confirmed through model comparison using the posterior Bayes factor (Aitkin, 1991), with a posterior Bayes factor greater than 1 from comparisons of the $K = 2$ model to $K = 3, \dots, 6$ models. Results we report here are based on the combined posterior output from two MCMC chains, with each chain initialized with different numbers of discriminatory ROIs and number of subjects in each subgroup. Other initial values were set as $\mu_k^{(0)} = \mathbf{0}$, $\sigma_k^{(0)} = 1 \forall k$, $\sigma_0^{(0)} = 1$, $\beta^{(0)} = \mathbf{0}$. We ran each MCMC chain for 100,000 iterations, with the first 50,000 sweeps discarded as burn-in.

As discussed in section 2.3.3, the hyperparameter e of the MRF prior (Equation 5) regulates the prior sparsity whereas f induces smoothness, with higher values of f yielding a higher prior probability that a region is selected given that its neighbors are selected. The choice of e and f has been discussed by Li and Zhang (2010) and Stingo et al. (2013). It is known that with distributions as in Equation (5) a phase transition boundary problem can be encountered, where the number of selected regions increases sharply for small changes in f (Li and Zhang, 2010). Here we set the sparsity parameter to $e = -4.5$, corresponding to a lower bound on the prior probability of selection of 1%. As for the prior smoothness, f , a plot of the prior over a grid of values $f \in \{0.1, 0.2, 0.3, \dots, 0.9\}$ revealed that the phase transition

starts at a prior smoothness of $f = 0.2$ and becomes severe at around $f = 0.4$. As suggested by Li and Zhang (2010), the prior smoothness parameter f was therefore set to a value far from the phase transition boundary. Here we present results for two values, $f = 0.01$ and $f = 0.1$, representing different levels of small-to-moderate effect of the prior information on connectivity. As for the other hyperparameter settings, we placed a vague prior on the mixing parameters π , that is, $\alpha_k = 1 \forall k$, and fixed the prior shape and scale parameters of the inverse gamma priors on σ_k and σ_0 to be non-informative with $a_k = 2$ and $b_k = 1 \forall k$, and $a_0 = 2$ and $b_0 = 1$. We also set the unscaled variance of the ICAR prior to $c_k = 5$, and the prior mean and covariance of β to $m_\beta = \mathbf{0}$ and $V_\beta = 5\mathbb{I}$, respectively. Age of the patient at surgery, epilepsy duration, and history of generalized tonic clonic seizures were controlled for as baseline covariates in the logistic likelihood.

Convergence of each MCMC chain was assessed using two independent tests: the Raftery-Lewis diagnostic (Raftery and Lewis, 1992) and the Geweke test (Geweke, 1991). In addition, convergence of the multiple chains was assessed using the Gelman-Rubin potential scale reduction factor, based on the implementation in the R package “coda” (Raftery and Lewis, 1992). Convergence diagnostics indicated convergence to the stationary distribution (results reported in the Supplementary Material). Agreement between MCMC chains was assessed through the Pearson correlation between the marginal posterior probabilities of ROI selection and cluster allocation of each pair of chains.

For posterior inference, our primary interest is in the estimation of the discriminatory regions, the latent states, and

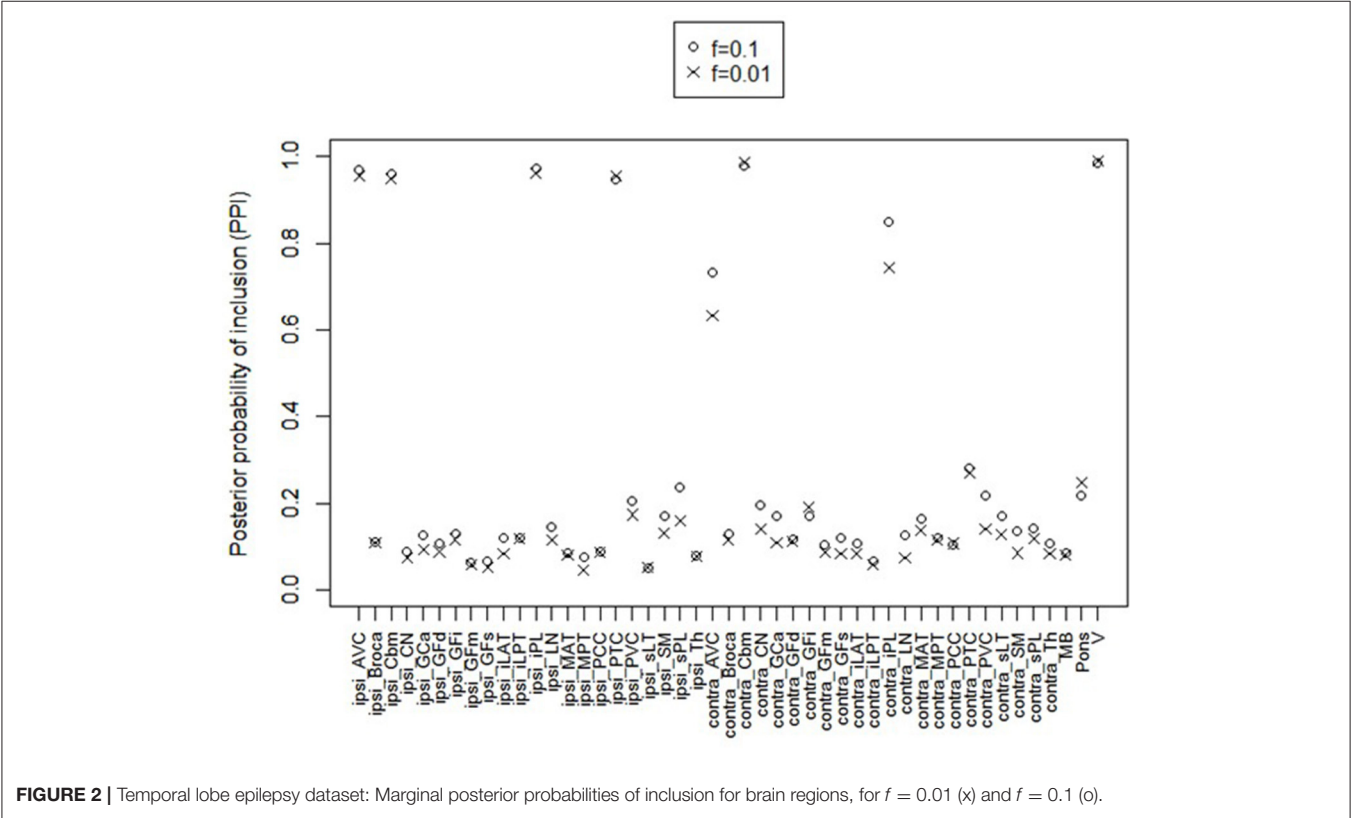
their association with the binary clinical outcome, as captured by the parameters γ , η , and β , respectively. Trace plots for these parameters showed good mixing for all chains (figures not shown). **Figure 2** shows the marginal posterior probabilities of inclusion (PPIs) for each of the 47 brain regions, with different graphical symbols for the settings of $f = 0.01$ (x) and $f = 0.1$ (o). Based on this plot, a selection of the discriminatory regions can be done by thresholding the PPIs. For example, the median model (Barbieri and Berger, 2004) selects the same subset of 8 ROIs under both $f = 0.01$ and $f = 0.1$. The selected brain regions are listed in **Table 1**, and graphically depicted in **Figure 3**. To examine the sensitivity of the selected regions to the formulation of the network prior, we additionally ran the model under a neighborhood matrix S defined by simple Euclidean distance. Selected discriminatory regions were robust to the formulation of the network, with the exception of the contralateral associative visual cortex, which had a marginal PPI of 0.303 ($f = 0.1$) and 0.311 ($f = 0.01$) under a network defined by spatial neighbors. This decrease in posterior probability is an effect of the MRF prior, due to the functional connectivity present between the ipsilateral and contralateral associative visual cortex in **Figure 1B** which is not captured based on spatial distance.

Figure 4 shows the marginal posterior probabilities of sample allocations for each of the 19 MTLE-HS patients. A classification of the subjects into two subgroups can be obtained, for example, by assigning subjects according to the posterior mode of η . For interpretation of the two subgroups, one can examine the PET metabolic activities characterizing the subjects. These are

shown in **Figure 5** for the selected brain regions. Furthermore, posterior inference for the β parameters is summarized in **Table 2**. These results suggest that the two subgroups identify patients at different levels of risk for post-operative seizure recurrence, with one subgroup having a $e^{\beta} = 5.2$ times greater odds of persistent post-operative seizures 1 year after surgery (**Table 2**). This corresponds to a 90% posterior probability of an odds ratio >1 for post-surgical seizure freedom between the two identified subgroups (**Table 2**). **Figure 5** reveals, in particular, that the subgroup with greater odds of post-operative seizure recurrence (Cluster 2) is characterized by lower levels of interictal glucose metabolism in the bilateral associative

TABLE 1 | Temporal lobe epilepsy dataset: Selected brain regions and corresponding marginal posterior probabilities of inclusion (PPI).

ROI	PPI	
	$f = 0.01$	$f = 0.1$
Ipsilateral inferior parietal lobule	0.961	0.973
Ipsilateral parietotemporal cortex	0.955	0.948
Ipsilateral associative visual cortex	0.956	0.969
Contralateral inferior parietal lobule	0.742	0.850
Contralateral associative visual cortex	0.632	0.732
Contralateral cerebellar hemisphere	0.988	0.979
Ipsilateral cerebellar hemisphere	0.950	0.961
Cerebellar vermis	0.989	0.984



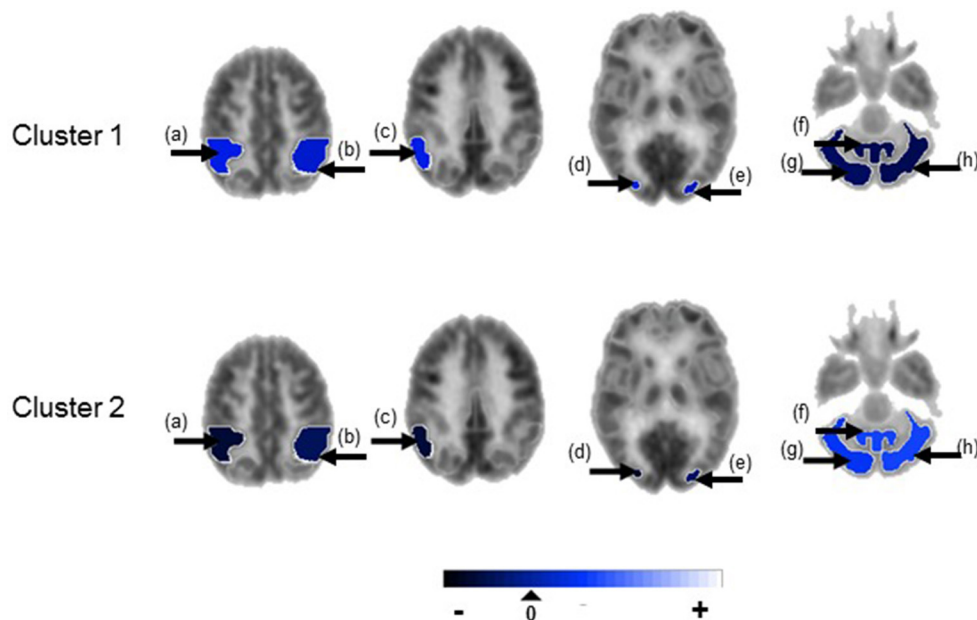


FIGURE 3 | Temporal lobe epilepsy dataset: Mean random effect (μ_k) of PET metabolic activity for ROIs with PPI greater than 0.5, shown on axial sections. (a) Ipsilateral inferior parietal lobule, (b) contralateral inferior parietal lobule, (c) ipsilateral parieto-temporal cortex, (d) ipsilateral associative visual cortex, (e) contralateral associative visual cortex, (f) cerebellar vermis, (g) ipsilateral cerebellar hemisphere, (h) contralateral cerebellar hemisphere. Non-selected ROIs are shown in grayscale.

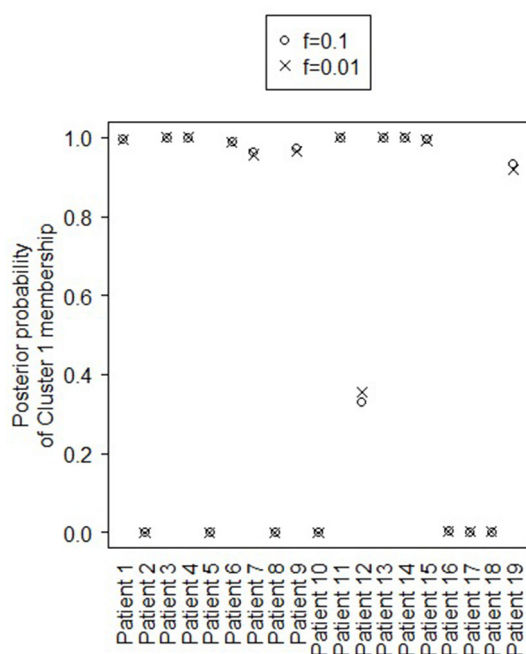


FIGURE 4 | Temporal lobe epilepsy dataset: Marginal posterior probabilities of cluster allocation, for $f = 0.01$ (x) and $f = 0.1$ (o).

visual cortices, ipsilateral parieto-temporal cortex, and bilateral inferior parietal cortices, as well as higher levels of interictal glucose metabolism in the bilateral cerebellar hemispheres

and cerebellar vermis. Our identification of these metabolic patterns may suggest extratemporal gliosis, as well as increased baseline levels of cortical excitability, in patients at higher risk for post-operative seizure recurrence. We provide further comment on the neurological significance of these findings in the Discussion.

3.3. Prediction Results

In addition to the identification of subgroups of subjects, characterized by latent pathologic conditions differentially associated to the outcome of interest, and the selection of imaging biomarkers that characterize the pathologic states of the subjects, our modeling approach allows a probabilistic estimate of an individual patient's risk of post-operative seizure recurrence. Probabilistic assessment of outcome risk may aid pre-surgical decision-making, by facilitating identification of patients with greater probability of seizure recurrence following anterior temporal lobe resection. Such information may potentially be weighed against the known risks of surgery (e.g., infection, bleeding, reactions to general anesthesia) to stratify patients according to predicted outcome. Here, we assessed prediction performance via importance-sampling cross-validation.

Figure 6 shows the receiver operating characteristic (ROC) curve, a plot of the false positive rates vs. the true positive rates, obtained for a grid of threshold values (0:0.05:1) on the estimated posterior predictive probabilities. The area under the curve (AUC) was 0.91. The optimal threshold, selected to maximize the Youden's index (Hiden and Glasziou, 1996), for imbalanced class sizes, resulted in an 84% predictive accuracy,

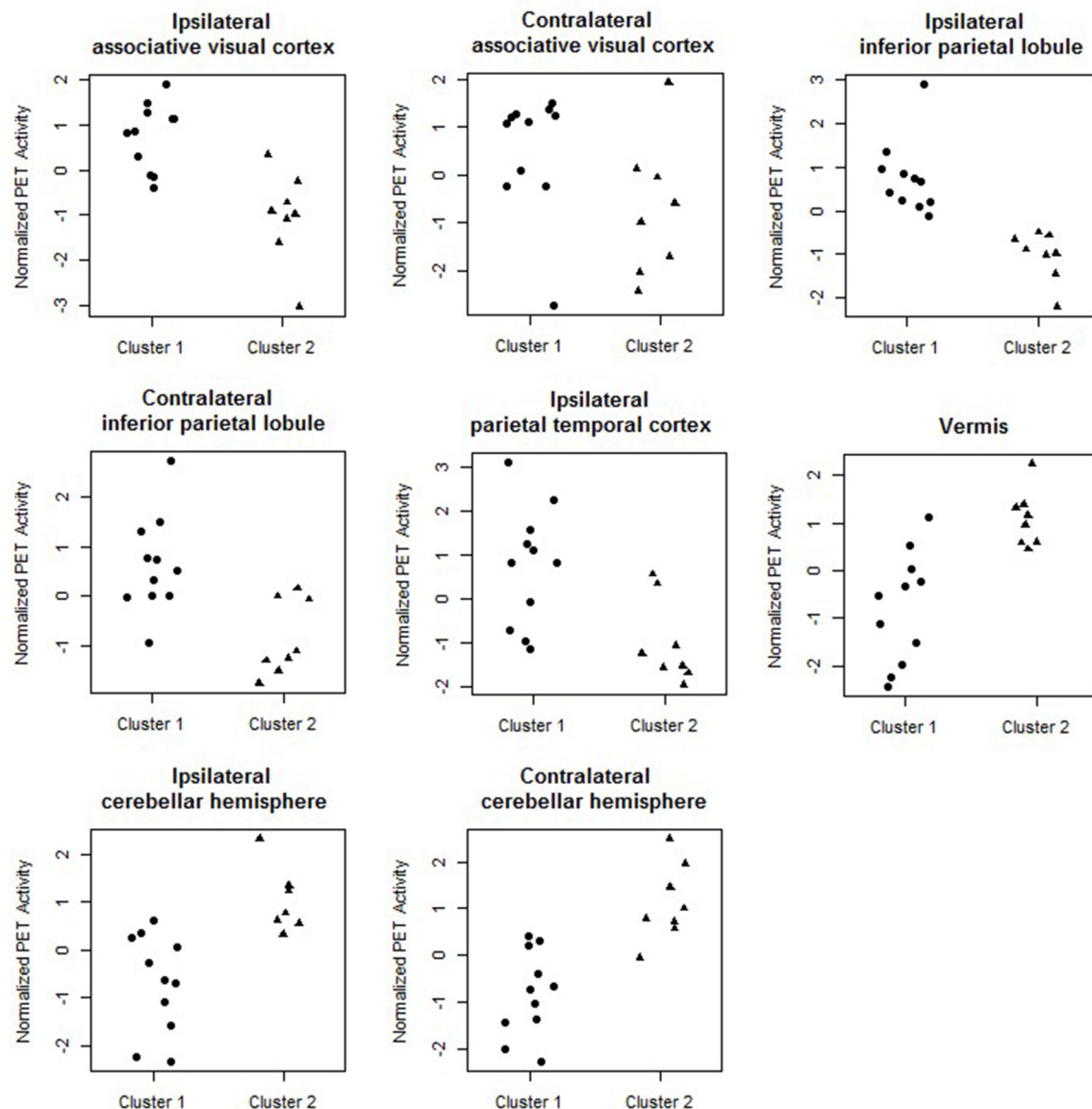


FIGURE 5 | Temporal lobe epilepsy dataset: Distribution of PET metabolic activity in the selected regions for the identified subgroups, for $f = 0.1$.

with correct prediction of post-surgical outcome in 16/19 patients, including 10/12 seizure-free patients and 6/7 non seizure free patients.

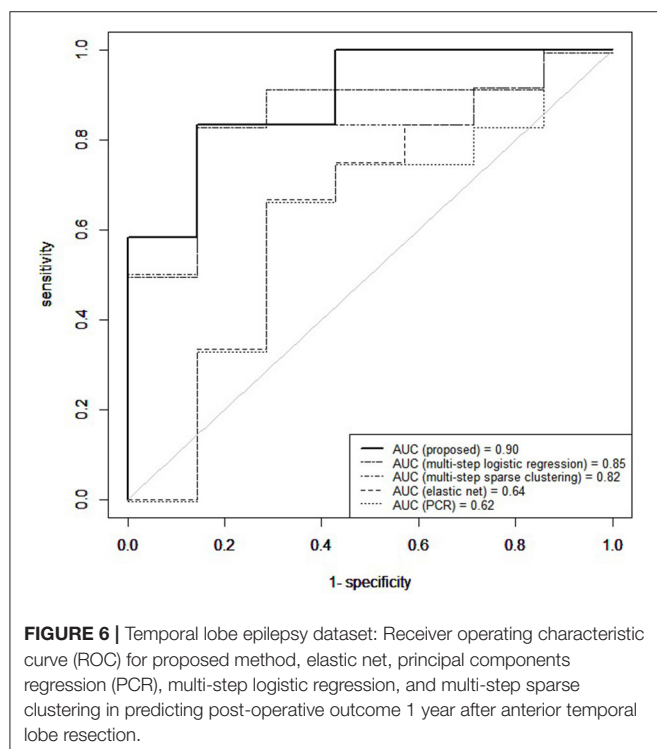
Our prediction results compared favorably to those we obtained on the same data with other analogous methods which predict binary outcomes from an identified underlying latent state. In particular, we compare to three multi-step approaches commonly used in prediction for their simplicity and computational speed. In the first approach, principal components was used to reduce the data to the top eight principal components, collectively explaining 85% of the variance in the data. The reduced principal components of \mathbf{X} were then used as predictors within Bayesian logistic regression.

Predictive accuracy was assessed through the importance-sampling cross-validation prediction approach of Gelfand (1996). In the second approach, a multistep logistic regression approach was used, similarly to what has been done in neuroimaging studies (Versace et al., 2014). In this approach, a filtering approach was performed by calculating permutation p -values for each region and retaining regions with small p -values. Using this reduced subset of regions, patients were clustered using k -means. Bayesian logistic regression was fitted to predict post-surgical outcome from latent class membership, and importance sampling cross-validation used to assess predictive accuracy. In the third comparison, a multi-step version of our approach was used, in which sparse cluster analysis was separated from the

TABLE 2 | Temporal lobe epilepsy dataset: (a) Posterior mean of β ; (b) 95% credible interval (CI) for β ; and (c) posterior probability of odds ratio >1 , e.g., $\mathbb{P}[e^{\beta_j} > 1|X, Y] = \mathbb{P}[\beta_j > 0|X, Y]$, shown for proposed approach ($f = 0.1$), multi-step logistic approach, and multi-step sparse clustering approach.

	Proposed Method			Multistep logistic regression			Multistep sparse clustering		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
β_0	-0.377	(-1.588, 3.587)	0.43	-0.225	(-0.8, 1.665)	0.41	-0.169	(-0.746, 1.718)	0.43
β_1	0.368	(0.275, 0.685)	0.99	0.250	(0.183, 0.48)	0.99	0.262	(0.196, 0.486)	0.99
β_2	-3.726	(-4.639, -0.76)	0.01	-1.25	(-1.742, 0.347)	0.06	-1.247	(-1.737, 0.358)	0.06
β_3	-0.082	(-0.117, 0.032)	0.09	-0.069	(-0.094, 0.011)	0.05	-0.079	(-0.107, 0.014)	0.05
β_4	1.649	(0.832, 4.364)	0.90	0.939	(0.453, 2.552)	0.88	0.41	(-0.113, 2.140)	0.68

Here $\beta_1 \equiv$ Epilepsy duration, $\beta_2 \equiv$ History of GTC, $\beta_3 \equiv$ Age at surgery, $\beta_4 \equiv$ Cluster 1 (v. Cluster 2). Odds are with respect to seizure freedom.



outcome model. In particular, a greedy forward search algorithm was used for simultaneous variable selection and clustering (Raftery and Dean, 2006). Patients were clustered based on the selected variables through a Gaussian mixture model (Fraley et al., 2012) and Bayesian logistic regression then used to predict post-surgical outcome from latent class membership, with predictive accuracy assessed through importance-sampling cross-validation. Prediction results using our unified approach attained superior predictive performance compared to multi-step approaches (Figure 6). Multi-step logistic regression and multi-step sparse clustering approaches attained higher predictive accuracy than PCR. We also compared to methods such as elastic net (Zou and Hastie, 2005), ridge regression (Hoerl and Kennard, 1970), and the Least Absolute Shrinkage and Selection Operator (LASSO) method of Tibshirani (1996) that, in particular, do not condition on latent states, but rather use

the X data as the covariates. Penalized regression approaches that did not condition on a latent state performed poorly in data with underlying latent states (see Supplementary Material). Additionally, in the Supplementary Material, we conduct a full comparison study among competing methods on synthetic data to evaluate results for both prediction and biomarker selection.

4. DISCUSSION

Our results have identified a subgroup of temporal lobe epilepsy patients with 5.8 times greater odds of post-operative seizure recurrence after anterior temporal lobe resection. These patients were characterized by lower levels of interictal metabolism in regions near the ipsilateral parieto-temporal-occipital junction. Lower interictal metabolism in peritemporal regions may suggest structural abnormalities such as gliosis or neuronal loss in these regions, alternatively or in combination with functional abnormality involving a widespread epileptogenic network which extends beyond the temporal lobe. Evidence for such a subgroup has been suggested by previous research, which found limited improvement in seizure outcomes in patients with electrocorticographical (ECoG) evidence of extratemporal involvement of inferior parietal cortex (Aghakhani et al., 2004). The implication of extratemporal brain structures in patients with poorer postsurgical outcomes supports the presence of latent pathologies in patients with epilepsy. Other ECoG studies have also suggested the presence of latent pathology in epilepsy involving spread of the epileptogenic focus and the possible creation of secondary foci (Rougier, 1990; D'Ambrosio et al., 2005). Therefore, lower interictal metabolism in this subset of patients may suggest a subtype of MTLE-HS with parietal involvement, which may lead to post-operative seizure generation if not resected. The involvement of posterior parietal regions in this subset of patients may result from connectivity to other regions clinically involved in MTLE. Structural connectivity exists between the presubiculum and the posterior parietal cortex through the cingulum, for example, and functional connectivity between these regions also exists through the default mode network (Buckner et al., 2008). Pulvinar atrophy has also been found in TLE patients with persistent post-operative seizures (Keller

et al., 2015), so connectivity of posterior parietal regions to the pulvinar nucleus may also play a role in posterior parietal involvement.

Patients at high risk for post-operative seizure recurrence were also characterized by higher levels of interictal glucose metabolism in the cerebellum. The cerebellum's role in inhibiting seizures has been investigated since the early 1940's, following the discovery that cerebellar stimulation may result in seizure modification or even termination (Moruzzi, 1950). Recent technological advances in techniques for cerebellar stimulation have led to renewed interest in the role of cerebellar stimulation in seizure inhibition, with a 41% seizure rate reduction achieved through cerebellar stimulation (Velasco et al., 2005). Direct optogenetic stimulation of the cerebellar Purkinje cells has been found to be sufficient to reduce the duration of seizures in temporal lobe epilepsy (Krook-Magnuson et al., 2014). It is postulated that the mechanism of cerebellar stimulation in seizure inhibition may be through increased inhibitory efferent output from the Purkinje cells to the deep cerebellar nuclei, resulting in increased inhibitory cerebellar output to the thalamocortical projections and thus decreased contralateral cortical excitability (Fountas et al., 2010). Likewise, the cerebral cortex exhibits feedback to the contralateral cerebellar hemispheres through corticopontocerebellar tracts. In our study, we found that the subgroup of MTLE-HS patients at high risk for post-operative seizure recurrence was characterized by higher levels of interictal glucose metabolism in the bilateral cerebellar hemispheres and cerebellar vermis, with slightly larger marginal posterior probability of discriminating high- vs. low-risk patients in the contralateral than the ipsilateral cerebellar hemisphere. Higher interictal glucose metabolism in the cerebellum may be caused by pre-operatively increased baseline levels of cortical excitability in high-risk patients, resulting in increased activity of corticopontocerebellar white matter tracts and increased crossed cerebellar metabolism. The localization of this phenomenon may be similar to that of cerebellar diaschisis, in which supratentorial lesions such as stroke may cause disruption of corticopontocerebellar tracts and therefore contralateral cerebellar hypometabolism. In the case of epilepsy, in which there is over- rather than underactivity of the cortex, overstimulation of the corticopontocerebellar tracts may lead to contralateral cerebellar hypermetabolism. Inhibitory outflow from the Purkinje cells may then result in hypometabolic activity in areas such as the inferior parietal lobule, congruent with the functional abnormality observed in the temporo-parieto-occipital junction as described above. Our observation of bilaterally increased glucose metabolism in the cerebellum suggest bilaterally increased cortical excitability in patients at high risk for post-operative seizure recurrence, with slightly higher cortical excitability ipsilaterally. The greater contralateral cerebellar involvement observed here is also consistent with our observation of ipsilaterally involved temporo-parieto-occipital regions due to crossed cerebello-cortical connections.

In addition to enhancing understanding of the pathophysiology behind post-operative seizure recurrence, our finding that patients at high risk for epilepsy surgery failure are

characterized by lower PET metabolism in peritemporal regions and higher cerebellar metabolism, provides a marker for patients where epilepsy surgery is at high risk for failure. These patients may be better candidates for neuromodulatory treatments for medication-refractory epilepsy, such as direct cortical stimulation, as is being used in responsive neurostimulation (RNS) at regions of seizure onset (Geller et al., 2017). We show that TLE patients at high risk for anterior temporal lobe resection failure have abnormal pre-surgical brain metabolic activity compared to those patients who attain post-surgical seizure freedom, suggesting a difference in the underlying brain networks of the two groups. The approach proposed here provides a method which may potentially allow for pre-surgical differentiation between patients with abnormal underlying brain activity.

In this paper we have developed a general integrative modeling framework to characterize the association between a set of image predictors and an individual clinical outcome that simultaneously (a) identifies subgroups of patients characterized by latent pathologies differentially associated to the outcome of interest, (b) identifies discriminatory brain regions across subjects, and (c) uses prior connectivity information from external data to inform the selection of biomarkers. Our Bayesian measurement error model provides a modeling approach for the prediction of post-surgical treatment response from imaging data which explicitly accounts for the unobserved disease state. As described in section 2.3.5, our model provides an approach in which a new prospective surgery candidate can come in, be scanned with PET imaging, assigned to a latent risk group, and evaluated for their probability of achieving seizure freedom if operated upon. By accounting for heterogeneity in the unobserved state, while allowing for incorporation of external prior information, we have obtained accurate prediction in data where surrogate measures, such as neuroimaging data, are observed. We have shown that our approach achieves superior predictive performance compared to commonly used approaches, such as principal components regression, ROI-based clustering, and ROI-based sparse regression, and additionally leads to accurate inference with respect to identification of latent states and variable selection.

We have used the proposed method to analyze data we have available from the University of California, Los Angeles Seizure Disorder Center, where the interest was in predicting the post-surgical outcome among MTLE-HS patients from pre-operative FDG-PET imaging. In the analysis, we have used resting-state fMRI imaging to inform the prior model. Our analysis has identified several discriminatory ROIs, together with a subgroup of patients at higher risk of post-operative seizures recurrence. Pre-surgical identification of regions pathophysiologically involved in post-operative seizure recurrence may assist in targeting these regions for interruption. Here, patients at higher risk were characterized by lower levels of interictal glucose metabolism in the bilateral associative visual cortices, ipsilateral parietotemporal cortex, and bilateral inferior parietal lobules, and higher levels of interictal glucose metabolism in the bilateral cerebellar hemispheres and cerebellar vermis. Cross-validated prediction of post-operative

seizure freedom has achieved an AUC of 0.91 and 84% predictive accuracy, showing superior predictive performance compared to methods which do not condition on latent states. One caution in interpreting the results of this study is the moderate statistical power due to limited sample size. Future corroboration on larger samples is needed prior to use in clinical practice. Pre-surgical identification of patients at high risk of not benefiting from surgery may improve treatment planning for these patients, including the potential avoidance of surgery risks in cases with low probability of benefit.

In our study, we have utilized standard PET ROIs obtained from quantitative assessment software used in clinical practice, where PET activity in each region of interest is computed by averaging within the ROI. Similar ROI-based approaches are utilized within the standard preprocessing protocol of NeuroQ to aid clinical interpretability, and have demonstrated clinical utility in neurological disorders such as Parkinson's disease (Akdemir et al., 2014), tinnitus (Smith et al., 2007), and epilepsy (Kerr et al., 2013). However, it is important to note that voxel-based data allow for a finer-grained approach to biomarker selection and may be of interest in future applications of our methodology. Use of other well-known atlases to segment PET data, such as the Automated Anatomic Labeling (AAL) atlas, may also be useful for comparing to other studies. Rigid registration and the use of PET-to-PET registration is also susceptible to PET signal variations, with hippocampal atrophy in TLE potentially contributing further to decreased registration accuracy as well as partial voluming effects. Further improvements in predictive accuracy may be seen with alternative pre-processing methods, including registration to high-resolution structural imaging and partial volume correction.

Future applications of our method to pre-operative mapping may wish to investigate finer parcellations of the brain, to better delineate the epileptogenic zone and more directly aid pre-operative mapping. Given the routine use of fMRI and EEG in the management of patients with epilepsy, it might also be possible to extend our general model formulation to the identification of spatial fMRI markers of disease outcome while taking advantage of the temporal resolution of EEG data to construct prior connectivity networks. Finally, even though the motivating example for our proposed model has come from the prediction of post-surgical outcomes in epilepsy surgery, data from other neurological disorders may also be analyzed. In such cases, it may be of interest to extend the treatment outcome to a multinomial likelihood, with larger sample sizes needed if such analysis is desired.

REFERENCES

Aghakhani, Y., Rosati, A., Dubeau, F., Olivier, A., and Andermann, F. (2004). Patients with temporoparietal ictal symptoms and inferomesial EEG do not benefit from anterior temporal resection. *Epilepsia* 45, 230–236. doi: 10.1111/j.0013-9580.2004.43003.x

ETHICS STATEMENT

In this paper we analyze data that were collected as part of a clinical study between 2007 and 2012. No new data were generated for this manuscript. The data were provided to us in computer form and were recorded in such a manner that subjects could not be identified directly or indirectly through identifiers/codes. The protocol that generated the data is now closed to enrollment.

AUTHOR CONTRIBUTIONS

SC, MG, ZH, JS, and MV contributed to the design and analysis of the work; HY, SD, and JS contributed to the data acquisition; SC, MG, and MV wrote the paper; all authors revised the manuscript critically for important intellectual content.

FUNDING

MV and MG are partially supported by NSF SES-1659925 and NSF SES-1659921. SC was supported by the National Library of Medicine Training Fellowship in Biomedical Informatics, Gulf Coast Consortia for Quantitative Biomedical Sciences (Grant #2T15-LM007093-21) and by the National Institute of Health (Grant #5T32-CA096520-07). ZH is partially supported by The Epilepsy Foundation of America (Award ID 244976), the Baylor College of Medicine, Computational and Integrative Biomedical Research Center (CIBR) Seed Grant Awards and the Baylor College of Medicine Junior Faculty Seed Funding Program Grant. JS is partially supported by NIH-NINDS K23 Grant NS044936 and The Leff Family Foundation.

ACKNOWLEDGMENTS

The authors would like to give special thanks to Daniel H. S. Silverman, Stefan T. Nguyen, Navya M. Reddy, and Regina Ahn (University of California, Los Angeles) for provision of the NeuroQ software, pre-processing of PET data, and organizational and software support. The authors also express their grateful appreciation to Wesley Kerr (University of California, Los Angeles) for data management of the PET records and helpful insights.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2017.00669/full#supplementary-material>

Aitkin, M. (1991). Posterior Bayes factors. *J. R. Stat. Soc. Ser. B (Methodol.)*, 53, 111–142.

Akdemir, Ü. Ö., Tokçer, A. B., Karakus, A., and Kapucu, L. Ö. (2014). Brain 18F-FDG PET imaging in the differential diagnosis of Parkinsonism. *Clin. Nucl. Med.* 39, e220–e226. doi: 10.1097/RLU.0000000000000315

- Alpert, N. M., Reilhac, A., Chio, T. C., and Selesnick, I. (2006). Optimization of dynamic measurement of receptor kinetics by wavelet denoising. *Neuroimage* 30, 444–451. doi: 10.1016/j.neuroimage.2005.09.031
- Andersson, J. L. R., Jenkinson, M., and Smith, S. (2007). *Non-linear Registration Aka Spatial Normalisation FMRIB Technical Report TR07JA2*. Technical report, FMRIB Centre, Oxford, UK.
- Arribas, J., Calhoun, V. D., and Adali, T. (2010). Automatic Bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from fMRI data. *IEEE Trans. Biomed. Eng.* 57, 2850–2860. doi: 10.1109/TBME.2010.2080679
- Baccalá, L. A., Alvarenga, M. Y., Sameshima, K., Jorge, C. L., and Castro, L. H. (2004). Graph theoretical characterization and tracking of the effective neural connectivity during episodes of mesial temporal epileptic seizure. *J. Integr. Neurosci.* 3, 379–395. doi: 10.1142/S0219635204000610
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL: Chapman and Hall/CRC.
- Barbieri, M. M., and Berger, J. O. (2004). Optimal predictive model selection. *Ann. Stat.* 32, 870–897. doi: 10.1214/009053604000000238
- Bonilha, L., Rorden, C., Castellano, G., Cendes, F., and Li, L. M. (2005). Voxel-based morphometry of the thalamus in patients with refractory medial temporal lobe epilepsy. *Neuroimage* 25, 1016–1021. doi: 10.1016/j.neuroimage.2004.11.050
- Bowman, F., Zhang, L., Derado, G., and Chen, S. (2012). Determining functional connectivity using fMRI data with diffusion-based anatomical weighting. *NeuroImage* 62, 1769–1779. doi: 10.1016/j.neuroimage.2012.05.032
- Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N.Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011
- Burge, J., Lane, T., Link, H., Qiu, S., and Clark, V. P. (2009). Discrete dynamic Bayesian network analysis of fMRI data. *Hum. Brain Mapp.* 30, 122–137. doi: 10.1002/hbm.20490
- Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Boca Raton, FL: CRC Press.
- Centeno, M., and Carmichael, D. W. (2014). Network connectivity in epilepsy: resting state fMRI and EEG-fMRI contributions. *Front. Neurol.* 5:93. doi: 10.3389/fneur.2014.00093
- Chiang, S., Cassese, A., Guindani, M., Vannucci, M., Yeh, H. J., Haneef, Z., et al. (2016). Time-dependence of graph theory metrics in functional connectivity analysis. *NeuroImage* 125, 601–615. doi: 10.1016/j.neuroimage.2015.10.070
- Chiang, S., and Haneef, Z. (2014). Graph theory findings in the pathophysiology of temporal lobe epilepsy. *Clin. Neurophysiol.* 125, 1295–1305. doi: 10.1016/j.clinph.2014.04.004
- Chiang, S., Haneef, Z., Stern, J. M., and Engel, J. (2017a). Use of resting-state fMRI in planning epilepsy surgery. *Neurol. India* 65, S25–S33. doi: 10.4103/neuroindia.NI_823_16
- Chiang, S., Vankov, E. R., Yeh, H. J., Guindani, M., Vannucci, M., Haneef, Z., et al. (2017b). Temporal and spectral characteristics of dynamic functional connectivity between resting-state networks reveal information beyond static connectivity.
- D'Ambrosio, R., Fender, J. S., Fairbanks, J. P., Simon, E. A., Born, D. E., Doyle, D. L., et al. (2005). Progression from frontal-parietal to mesial-temporal epilepsy after fluid percussion injury in the rat. *Brain* 128, 174–188. doi: 10.1093/brain/awh337
- de Tisi, J., Bell, G. S., Peacock, J. L., McEvoy, A. W., Harkness, W. F., Sander, J. W., et al. (2011). The long-term outcome of adult epilepsy surgery, patterns of seizure remission, and relapse: a cohort study. *Lancet* 378, 1388–1395. doi: 10.1016/S0140-6736(11)60890-8
- Derado, G., Bowman, F. D., Zhang, L., and Alzheimer's Disease Neuroimaging Initiative Investigators (2013). Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* 22, 382–397. doi: 10.1177/0962280212448972
- Dupont, S., Semah, F., Clémenceau, S., Adam, C., Baulac, M., and Samson, Y. (2000). Accurate prediction of postoperative outcome in mesial temporal lobe epilepsy: a study using positron emission tomography with 18-Fluorodeoxyglucose. *Arch. Neurol.* 57, 1331–1336. doi: 10.1001/archneur.57.9.1331
- Engel, J., McDermott, M. P., Wiebe, S., Langfitt, J. T., Stern, J. M., Dewar, S., et al. (2012). Early surgical therapy for drug-resistant temporal lobe epilepsy: a randomized trial. *JAMA* 307, 922–930. doi: 10.1001/jama.2012.220
- Ercoli, L., Small, G., Siddarth, P., Kepe, V., Huang, S.-C., Miller, K., et al. (2012). Assessment of dementia risk in aging adults using both FDG-PET and FDDNP-PET imaging. *Int. J. Geriatr. Psychiatry* 27, 1017–1027. doi: 10.1002/gps.2816
- Fountas, K. N., Kapsalaki, E., and Hadjigeorgiou, G. (2010). Cerebellar stimulation in the management of medically intractable epilepsy: a systematic and critical review. *Neurosurg. Focus* 29:E8. doi: 10.3171/2010.5.FOCUS10111
- Fraley, C., Raftery, A., Murphy, T., and Scrucca, L. (2012). *mclust Version 4 for R: Normal Mixture Modeling for Model-based Clustering, Classification, and Density Estimation*. Technical Report No. 597, Department of Statistics, University of Washington.
- Franaszczuk, P. J., and Bergey, G. K. (1998). Application of the directed transfer function method to mesial and lateral onset temporal lobe seizures. *Brain Topogr.* 11, 13–21. doi: 10.1023/A:1022262318579
- Franaszczuk, P. J., Bergey, G. K., and Kamiński, M. J. (1994). Analysis of mesial temporal seizure onset and propagation using the directed transfer function method. *Electroencephalogr. Clin. Neurophysiol.* 91, 413–427. doi: 10.1016/0013-4694(94)90163-5
- Gelfand, A. E. (1996). "Model determination using sampling-based methods," in *Markov Chain Monte Carlo in Practice*, Chap. 9, eds W. R. Gilks, S. Richardson, and D. Spiegelhalter (London: Chapman and Hall), 145–161.
- Geller, E. B., Skarpaas, T. L., Gross, R. E., Goodman, R. R., Barkley, G. L., Bazil, C. W., et al. (2017). Brain-responsive neurostimulation in patients with medically intractable mesial temporal lobe epilepsy. *Epilepsia* 58, 994–1004. doi: 10.1111/epi.13740
- Geweke, J. (1991). *Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments*, vol. 196. Minneapolis, MN: Federal Reserve Bank of Minneapolis, Research Department Minneapolis.
- Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comput. Graph. Stat.* 23, 46–64. doi: 10.1080/10618600.2012.743437
- Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060
- Haynes, J.-D., and Rees, G. (2005). Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* 15, 1301–1307. doi: 10.1016/j.cub.2005.06.026
- Hidden, J., and Glasziou, P. (1996). Regret graphs, diagnostic uncertainty and Youden's Index. *Stat. Med.* 15, 969–986.
- Hinne, M., Ambrogioni, L., Janssen, R. J., Heskes, T., and van Gerven, M. A. (2014). Structurally-informed bayesian functional connectivity analysis. *Neuroimage* 86, 294–305. doi: 10.1016/j.neuroimage.2013.09.075
- Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12, 55–67. doi: 10.1080/00401706.1970.10488634
- Hoff, P. D. (2006). Model-based subspace clustering. *Bayesian Anal.* 1, 321–344. doi: 10.1214/06-BA111
- Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106
- Insel, T., and Cuthbert, B. (2015). Brain disorders? Precisely. *Science* 348, 499–500. doi: 10.1126/science.aab2358
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17, 825–841. doi: 10.1006/nimg.2002.1132
- Jiang, C.-R., Aston, J. A., and Wang, J.-L. (2009). Smoothing dynamic positron emission tomography time courses using functional principal components. *NeuroImage* 47, 184–193. doi: 10.1016/j.neuroimage.2009.03.051
- Jiang, H., and Ogden, R. T. (2008). Mixture modeling for dynamic PET data. *Stat. Sin.* 18, 1341–1356.
- Jirsch, J., Urrestarazu, E., LeVan, P., Olivier, A., Dubeau, F., and Gotman, J. (2006). High-frequency oscillations during human focal seizures. *Brain* 129, 1593–1608. doi: 10.1093/brain/awl085

- Jorge, J., van der Zwaag, W., and Figueiredo, P. (2014). EEG-fMRI integration for the study of human brain function. *Neuroimage* 102, 24–34. doi: 10.1016/j.neuroimage.2013.05.114
- Keller, S. S., Richardson, M. P., Schoene-Bake, J.-C., O'Muirheartaigh, J., Elkomos, S., Kreilkamp, B., et al. (2015). Thalamotemporal alteration and postoperative seizures in temporal lobe epilepsy. *Ann. Neurol.* 77, 760–774. doi: 10.1002/ana.24376
- Kerr, W. T., Nguyen, S. T., Cho, A. Y., Lau, E. P., Silverman, D. H., Douglas, P. K., et al. (2013). Computer-aided diagnosis and localization of lateralized temporal lobe epilepsy using interictal FDG-PET. *Front. Neurol.* 4:31. doi: 10.3389/fneur.2013.00031
- Kramer, M. A., Kolaczky, E. D., and Kirsch, H. E. (2008). Emergent network topology at seizure onset in humans. *Epilepsy Res.* 79, 173–186. doi: 10.1016/j.eplepsyres.2008.02.002
- Krook-Magnuson, E., Szabo, G. G., Armstrong, C., Oijala, M., and Soltesz, I. (2014). Cerebellar directed optogenetic intervention inhibits spontaneous hippocampal seizures in a mouse model of temporal lobe epilepsy. *eNeuro* 1, 1–15. doi: 10.1523/ENEURO.0005-14.2014
- LaConte, S., Strother, S., Cherkassky, V., Anderson, J., and Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *NeuroImage* 26, 317–329. doi: 10.1016/j.neuroimage.2005.01.048
- Li, F., and Zhang, N. R. (2010). Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J. Am. Stat. Assoc.* 105, 1202–1214. doi: 10.1198/jasa.2010.tm08177
- Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L., Coan, J. A., et al. (2015). Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *Ann. Appl. Stat.* 9, 687–713. doi: 10.1214/15-AOAS818
- Lin, T. W., de Aburto, M. A. K., Dahlbom, M., Huang, L. L., Marvi, M. M., Tang, M., et al. (2007). Predicting seizure-free status for temporal lobe epilepsy patients undergoing surgery: prognostic value of quantifying maximal metabolic asymmetry extending over a specified proportion of the temporal lobe. *J. Nucl. Med.* 48, 776–782. doi: 10.2967/jnumed.106.034249
- Lüders, H., Engel, J., and Munari, C. (1993). “General principles,” in *Surgical Treatment of the Epilepsies, 2nd Edn.* eds J. Engel Jr. (New York, NY: Raven Press), 609–622.
- Lüders, H. O. (2008). *Textbook of Epilepsy Surgery*. Boca Raton, FL: CRC Press.
- Ma, S., Calhoun, Vince D., and Phlypo, R., and Adali, T. (2014). Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis. *Neuroimage* 90, 196–206. doi: 10.1016/j.neuroimage.2013.12.063
- McCallum, R., Dusing, R., Sarosiek, I., Cocjin, J., Forster, J., and Lin, Z. (2010). Mechanisms of symptomatic improvement after gastric electrical stimulation in gastroparetic patients. *Neurogastroenterol. Motil.* 22:161. doi: 10.1111/j.1365-2982.2009.01389.x
- McDonald, C. R., Hagler, D. J., Ahmadi, M. E., Tecoma, E., Iragui, V., Gharapetian, L., et al. (2008). Regional neocortical thinning in mesial temporal lobe epilepsy. *Epilepsia* 49, 794–803. doi: 10.1111/j.1528-1167.2008.01539.x
- Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* 4:200. doi: 10.3389/fnins.2010.00200
- Millet, P., Ibáñez, V., Delforge, J., Pappata, S., and Guimón, J. (2000). Wavelet analysis of dynamic PET data: application to the parametric imaging of benzodiazepine receptor concentration. *Neuroimage* 11, 458–472. doi: 10.1006/nimg.2000.0563
- Mitchell, T., Hutchinson, R., Niculescu, R., Pereira, F., Wang, X., Just, M., et al. (2004). Learning to decode cognitive states from brain images. *Mach. Learn.* 57, 145–175. doi: 10.1023/B:MACH.0000035475.85309.1b
- Moruzzi, G. (1950). Effects at different frequencies of cerebellar stimulation upon postural tonus and myotatic reflexes. *Electroencephalogr. Clin. Neurophysiol.* 2, 463–469. doi: 10.1016/0013-4694(50)90083-6
- Mueller, S. G., Laxer, K. D., Barakos, J., Cheong, I., Finlay, D., Garcia, P., et al. (2010). Involvement of the thalamocortical network in the with and without mesiotemporal sclerosis. *Epilepsia* 51, 1436–1445. doi: 10.1111/j.1528-1167.2009.02413.x
- Munsell, B. C., Wee, C.-Y., Keller, S. S., Weber, B., Elger, C., da Silva, L. A. T., et al. (2015). Evaluation of machine learning algorithms for treatment outcome prediction in patients with epilepsy based on structural connectome data. *NeuroImage* 118, 219–230. doi: 10.1016/j.neuroimage.2015.06.008
- Njiwa, J. Y., Gray, K., Costes, N., Mauguire, F., Ryvlin, P., and Hammers, A. (2015). Advanced [18 F] FDG and [11 C] flumazenil PET analysis for individual outcome prediction after temporal lobe epilepsy surgery for hippocampal sclerosis. *NeuroImage* 7, 122–131. doi: 10.1016/j.nicl.2014.11.013
- O'Sullivan, F. (2006). Locally constrained mixture representation of dynamic imaging data from PET and MR studies. *Biostatistics* 7, 318–338. doi: 10.1093/biostatistics/kxj010
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *J. Am. Stat. Assoc.* 108, 1339–1349. doi: 10.1080/01621459.2013.829001
- Quigley, M., Cordes, D., Turski, P., Moritz, C., Haughton, V., Seth, R., et al. (2003). Role of the corpus callosum in functional connectivity. *Am. J. Neuroradiol.* 24, 208–212.
- Raftery, A. E., and Dean, N. (2006). Variable selection for model-based clustering. *J. Am. Stat. Assoc.* 101, 168–178. doi: 10.1198/016214506000000113
- Raftery, A. E., and Lewis, S. (1992). How many iterations in the Gibbs sampler. *Bayesian Stat.* 4, 763–773.
- Raftery, A. E., and Lewis, S. M. (1992). One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. *Stat. Sci.* 7, 493–497. doi: 10.1214/ss/1177011143
- Richardson, S., and Gilks, W. (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Stat. Med.* 12, 1703–1722. doi: 10.1002/sim.4780121806
- Rosenow, F., and Lüders, H. (2001). Presurgical evaluation of epilepsy. *Brain* 124, 1683–1700. doi: 10.1093/brain/124.9.1683
- Rougier, A. (1990). The epileptic focus versus the pathological focus. *Acta Neurochir. Suppl. (Wien)* 50, 1–5.
- Ryvlin, P., and Kahane, P. (2005). The hidden causes of surgery-resistant temporal lobe epilepsy: extratemporal or temporal plus? *Curr. Opin. Neurol.* 18, 125–127. doi: 10.1097/01.wco.0000162852.22026.6f
- Sandok, E. K., O'Brien, T. J., Jack, C. R., and So, E. L. (2000). Significance of cerebellar atrophy in intractable temporal lobe epilepsy: a quantitative MRI study. *Epilepsia* 41, 1315–1320. doi: 10.1111/j.1528-1157.2000.tb04611.x
- Savitsky, T., Vannucci, M., and Sha, N. (2011). Variable selection for nonparametric Gaussian process priors: models and computational strategies. *Stat. Sci.* 26, 130–149. doi: 10.1214/11-STS354
- Silverman, D. H., and Melega, W. P. (2004). “Molecular imaging of biological processes with PET: evaluating biologic bases of cerebral function,” in *PET molecular imaging and its biological applications*, ed M. E. Phelps (New York, NY: Springer-Verlag), 509–583.
- Silverman, D. H., Mosconi, L., Ercoli, L., Chen, W., and Small, G. W. (2008). Positron emission tomography scans obtained for the evaluation of cognitive dysfunction. *Semin. Nucl. Med.* 38, 251–261. doi: 10.1053/j.semnuclmed.2008.02.006
- Smith, J. A., Mennemeier, M., Bartel, T., Chelette, K. C., Kimbrell, T., Triggs, W., et al. (2007). Repetitive transcranial magnetic stimulation for tinnitus: a pilot study. *Laryngoscope* 117, 529–534. doi: 10.1097/MLG.0b013e31802f4154
- Smith, M., and Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *J. Am. Stat. Assoc.* 102, 417–431. doi: 10.1198/016214506000001031
- Sokoloff, L., Reivich, M., Kennedy, C., Des Rosiers, M., Patlak, C. S., Pettigrew, K., et al. (1977). The [14c] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J. Neurochem.* 28, 897–916. doi: 10.1111/j.1471-4159.1977.tb10649.x
- Spencer, S., Berg, A., Vickrey, B., Sperling, M., Bazil, C., Shinnar, S., et al. (2005). Predicting long-term seizure outcome after resective epilepsy surgery: the multicenter study. *Neurology* 65, 912–918. doi: 10.1212/01.wnl.0000176055.45774.71
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B (Stat. Methodol.)* 64, 583–639. doi: 10.1111/1467-9868.00353
- Stingo, F. C., Guindani, M., Vannucci, M., and Calhoun, V. D. (2013). An integrative Bayesian modeling approach to imaging genetics. *J. Am. Stat. Assoc.* 108, 876–891. doi: 10.1080/01621459.2013.804409

- Tai, Y.-C., Lin, K. P., Hoh, C. K., Huang, S. H., and Hoffman, E. J. (1997). Utilization of 3-D elastic transformation in the registration of chest X-ray CT and whole body PET. *IEEE Trans. Nucl. Sci.* 44, 1606–1612. doi: 10.1109/23.632740
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B (Methodol.)* 58, 267–288.
- Torosyan, N., and Silverman, D. H. (2012). Neuronuclear imaging in the evaluation of dementia and mild decline in cognition. *Semin. Nucl. Med.* 42, 415–422.
- van Gerven, M. A. J., Cseke, B., de Lange, F., and Heskes, T. (2010). Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior. *NeuroImage* 50, 150–161. doi: 10.1016/j.neuroimage.2009.11.064
- Velasco, F., Carrillo-Ruiz, J. D., Brito, F., Velasco, M., Velasco, A. L., Marquez, I., et al. (2005). Double-blind, randomized controlled pilot study of bilateral cerebellar stimulation for treatment of intractable motor seizures. *Epilepsia* 46, 1071–1081. doi: 10.1111/j.1528-1167.2005.70504.x
- Versace, F., Engelmann, J. M., Robinson, J. D., Jackson, E. F., Green, C. E., Lam, C. Y., et al. (2014). Prequit fMRI responses to pleasant cues and cigarette-related cues predict smoking cessation outcome. *Nicotine Tobacco Res.* 16, 697–708. doi: 10.1093/ntr/ntt214
- Wahl, R. L., Jacene, H., Kasamon, Y., and Lodge, M. A. (2009). From recist to perclist: evolving considerations for pet response criteria in solid tumors. *J. Nucl. Med.* 50(Suppl. 1), 122S–150S. doi: 10.2967/jnumed.108.057307
- Wiebe, S., Blume, W. T., Girvin, J. P., and Eliasziw, M. (2001). A randomized, controlled trial of surgery for temporal lobe epilepsy. *New Eng. J. Med.* 345, 311–318. doi: 10.1056/NEJM200108023450501
- Wieser, H.-G. (2004). Ilae commission report: mesial temporal lobe epilepsy with hippocampal sclerosis. *Epilepsia* 45, 695–714. doi: 10.1111/j.0013-9580.2004.09004.x
- Willmann, O., Wennberg, R., May, T., Woermann, F., and Pohlmann-Eden, B. (2007). The contribution of 18F-FDG PET in preoperative epilepsy surgery evaluation for patients with temporal lobe epilepsy: a meta-analysis. *Seizure* 16, 509–520. doi: 10.1016/j.seizure.2007.04.001
- Worrell, G. A., Gardner, A. B., Stead, S. M., Hu, S., Goerss, S., Cascino, G. J., et al. (2008). High-frequency oscillations in human temporal lobe: simultaneous microwire and clinical macroelectrode recordings. *Brain* 131, 928–937. doi: 10.1093/brain/awn006
- Worrell, G. A., Parish, L., Cranstoun, S. D., Jonas, R., Baltuch, G., and Litt, B. (2004). High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain* 127, 1496–1506. doi: 10.1093/brain/awh149
- Zhang, L., Guindani, M., Versace, F., and Vannucci, M. (2014). A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* 95, 162–175. doi: 10.1016/j.neuroimage.2014.03.024
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B (Stat. Methodol.)* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chiang, Guindani, Yeh, Dewar, Haneef, Stern and Vannucci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multivariate Heteroscedasticity Models for Functional Brain Connectivity

Christof Seiler* and Susan Holmes

Department of Statistics, Stanford University, Stanford, CA, United States

OPEN ACCESS

Edited by:

Xiaoying Tang,
Center for Imaging Science,
United States

Reviewed by:

Thomas Vincent,
Concordia University, PERFORM
Center, Canada
Xiaoyun Liang,
Florey Institute of Neuroscience and
Mental Health, Australia

*Correspondence:

Christof Seiler
christof.seiler@stanford.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 27 December 2016

Accepted: 27 November 2017

Published: 12 December 2017

Citation:

Seiler C and Holmes S (2017)
Multivariate Heteroscedasticity
Models for Functional Brain
Connectivity. *Front. Neurosci.* 11:696.
doi: 10.3389/fnins.2017.00696

Functional brain connectivity is the co-occurrence of brain activity in different areas during resting and while doing tasks. The data of interest are multivariate timeseries measured simultaneously across brain parcels using resting-state fMRI (rfMRI). We analyze functional connectivity using two heteroscedasticity models. Our first model is low-dimensional and scales linearly in the number of brain parcels. Our second model scales quadratically. We apply both models to data from the Human Connectome Project (HCP) comparing connectivity between short and conventional sleepers. We find stronger functional connectivity in short than conventional sleepers in brain areas consistent with previous findings. This might be due to subjects falling asleep in the scanner. Consequently, we recommend the inclusion of average sleep duration as a covariate to remove unwanted variation in rfMRI studies. A power analysis using the HCP data shows that a sample size of 40 detects 50% of the connectivity at a false discovery rate of 20%. We provide implementations using R and the probabilistic programming language Stan.

Keywords: Bayesian analysis, functional connectivity, heteroscedasticity, covariance regression, sleep duration

1. INTRODUCTION

Functional connectivity focuses on the exploration of neurophysiological measures of brain activity between brain regions (Friston, 2011; Smith, 2012; Varoquaux and Craddock, 2013). Functional connectivity studies have increased our understanding of the basic structure of the brain (Eguíluz et al., 2005; Sporns et al., 2004; Bassett and Bullmore, 2006; Fox and Raichle, 2007; Bullmore and Sporns, 2009; Van Den Heuvel and Pol, 2010) and provided insights into pathologies (Greicius et al., 2003; Greicius, 2008; Biswal et al., 2010; Fox and Greicius, 2010).

From a statistical viewpoint, functional connectivity is the problem of estimating covariance matrices, precision matrices, or correlation matrices from timeseries data. These matrices encode the level of connectivity between any two brain regions. The timeseries are derived from resting-state fMRI (rfMRI) by averaging individual voxels over parcels in the gray matter. We define parcels manually or with data-driven brain parcellation algorithms. The final goal can be an exploratory or a differential analysis comparing connectivity across regions between experimental conditions and time (Preti et al., 2016). Many statistical methods are available to estimate covariance matrices, precision matrices, or correlation matrices from multivariate data. The sample covariance and its inverse, or the sample correlation matrix are usually poor estimators because of the high-dimensionality of the data (large number of parcels p and small number of subjects). The number of parameters grows quadratically in the number of regions with $p(p - 1)/2$ possible pairwise connections between parcels. Therefore more elaborate estimators need to be employed, such as

the Graphical Lasso (Friedman et al., 2008) estimator for inverse-covariance matrices or the Ledoit-Wolf shrinkage estimator (Ledoit and Wolf, 2004) for correlation matrices. Application of these methods to rfMRI are available (Varoquaux et al., 2010a,b; Smith et al., 2011; Ryali et al., 2012; Varoquaux et al., 2012; Liang et al., 2016).

The estimation of connectivity is usually only the first step and leads to downstream differential analyses comparing connectivity between experimental conditions or between subgroups. For instance, we will compare the connectivity of short sleepers with conventional sleepers available as preprocessed timeseries from the Human Connectome Project (Van Essen et al., 2013). One approach is massive univariate testing of each of the $p(p - 1)/2$ connections by linear modeling. Such an approach allows us to test different contrasts and include batch factors or random effect terms (Lewis et al., 2009; Grillon et al., 2013). It lacks statistical power because it ignores possible dependencies between elements in the connectivity matrix. An alternative is to assess selected functionals or summary statistics rather than individual elements in the connectivity matrix (Stam, 2004; Salvador et al., 2005; Achard et al., 2006; Marrelec et al., 2008; Bullmore and Sporns, 2009; Ginestet et al., to appear). Another approach is to flip response variable and explanatory variable and predict experimental condition (or subgroup) from connectivity matrices (or functionals of matrices) through machine learning (Dosenbach et al., 2010; Craddock et al., 2012). These approaches lack interpretability in terms of brain function.

The problem boils down to modeling heteroscedasticity. Heteroscedasticity is said to occur when the variance of the unobservable error, conditional on explanatory variables, is not constant. For example, consider the regression problem predicting expenditure on meals from income. People with higher income will have greater variability in their choices of food consumption. A poorer person will have less choice, and be constrained to inexpensive foods. In functional connectivity, heteroscedasticity is multivariate and variances become covariance matrices. In other words, heteroscedasticity co-occurs among brain parcels and can be explained as a function of explanatory variables.

In this article, we propose a low-dimensional multivariate heteroscedasticity model for functional connectivity. Our model is of intermediary complexity, between modeling all $p(p - 1)/2$ connections and only using global summary statistics. Our model builds on the covariance regression model introduced by Hoff and Niu (2012). It includes a random effects term that describes heteroscedasticity in the multivariate response variable. We adapt it for functional connectivity and implement it using the statistical programming language Stan. Additionally, we perform preliminary thinning of the observed multivariate timeseries from N to the effective sample size n . Using n reduces false positives and speeds up computations by a factor of N/n . To find the appropriate n , we compute the autocorrelation as it is common in the Markov chain Monte Carlo literature. We compare our low-dimensional model to a full covariance model contained in the class of linear covariance models introduced by Anderson (1973). Both models are used to analyze real data from

HCP comparing connectivity between short and conventional sleepers.

From a neuroscience viewpoint, our low-dimensional model is applicable if we believe that multiple brain parcels work together to accomplish cognitive tasks. Even if this assumption is not entirely true, our low-dimensional model can serve a way to simplify functional connectivity analyses and improve interpretability. One can think of a low-dimensional model as a way to reduce the dimensions of the original data to an interpretable number of variables.

2. MATERIALS AND METHODS

2.1. Data

We analyzed data from the WU-Minn HCP 1200 Subjects Data Release. We focus on the functional-resting fMRI (rfMRI) data of 820 subjects. The images were acquired in four runs of approximately 15 min each. Acquisition ranged over 13 periods (Q01, Q02, ..., Q13). We separated the subjects into two groups: short sleepers (≤ 6 h) or conventional sleepers (7–9 h) as defined by the National Sleep Foundation (Hirshkowitz et al., 2015). This results in 489 conventional and 241 short sleepers. The HCP 1200 data repository contains images processed at different levels: spatially registered images, functional timeseries, and connectivity matrices. We work with the preprocessed timeseries data. In particular, the rfMRI preprocessing pipeline includes both spatial (Glasser et al., 2013) and temporal preprocessing (Smith et al., 2013). The spatial preprocessing uses tools from FSL (Jenkinson et al., 2012) and FreeSurfer (Fischl et al., 1999) to minimize distortions and align subject-specific brain anatomy to reference atlases using volume-based and surface-based registration methods. After spatial preprocessing, artifacts are removed from each subject individually (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014), then the data are temporally demeaned and variance stabilized (Beckmann and Smith, 2004), and further denoised using a group-PCA (Smith et al., 2014). Components of a spatial group-ICA (Hyvärinen, 1999; Beckmann and Smith, 2004) are mapped to each subject defining parcels (Glasser et al., 2013). The ICA-weighted voxelwise rfMRI signal are averaged over each component. Each weighted average represents one row in the multivariate timeseries. Note that parcels obtained in this way are not necessarily spatially contiguous, in particular, they can overlap and include multiple spatially separated regions. HCP provides a range of ICA components 15, 25, 50, 100, 200, and 300. We choose 15 (Figure 1) for our analysis to allow for comparison with prior sleep related findings on a partially overlapping dataset (Curtis et al., 2016).

2.2. Low-Dimensional Covariance Regression

In this section, we introduce a low-dimensional linear model to compare connectivity between experimental conditions or subgroups.

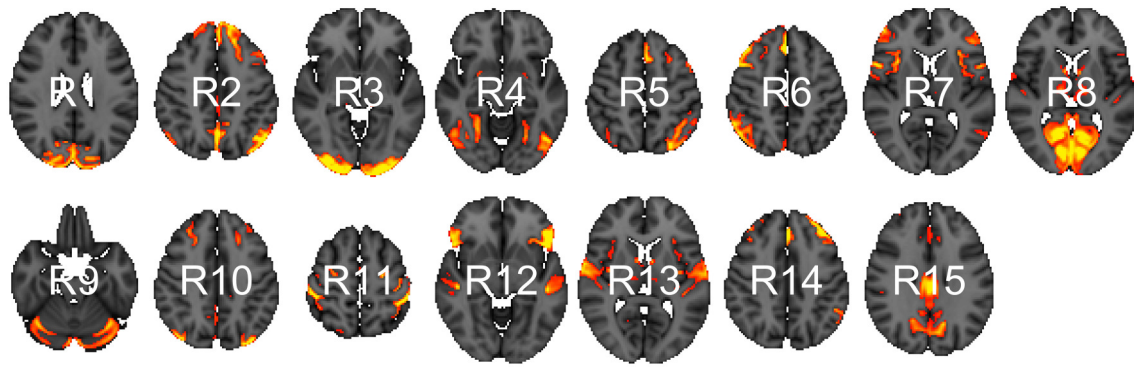


FIGURE 1 | Parcels derived from spatial group-ICA. Created at the most relevant axial slices in MNI152 space. According to Smith et al. (2009), these parcels map to visual areas (R1, R3, R4, and R8), sensorimotor (R7 and R11), cognition-language (R2, R5, R10, and R14), perception-somesthesis-pain (R2, R6, R10, and R14), cerebellum (R9), executive control (R12), auditory (R12 and R13), and default network (R15).

2.2.1. Model

The data we observe are p -dimensional multivariate vectors y_1, \dots, y_N . We assume that the observations are mean-centered so that $\frac{1}{N} \sum_{i=1}^N y_i = 0$. After centering, we subsample each timeseries at $n < N$ time points to remove temporal dependencies between observations (section 2.2.2). We are given a set of explanatory variables x_i that encode experimental conditions or subgroups, e.g., element one is the intercept 1 and element two is 0 for conventional and 1 for short sleepers. We bind the x_i 's row-wise into the usual design matrix X . Our model:

$$y_i = \gamma_i \times Bx_i + \epsilon_i \quad \text{for } i = 1, \dots, n$$

has a random effects term $\gamma_i \times Bx_i$ and an independent and identically distributed error term ϵ_i . We suppose the two random variables to have:

$$\begin{aligned} E(\epsilon_i) &= 0, & \text{Cov}(\epsilon_i) &= \sigma^2 I_p \\ E(\gamma_i) &= 0, & \text{Var}(\gamma_i) &= 1, & E(\gamma_i \times \epsilon_i) &= 0. \end{aligned}$$

Then, the expected covariance is of the form:

$$E(y_i y_i^T) = Bx_i x_i^T B^T + \sigma^2 I_p = \Sigma_{x_i}.$$

resulting from the inclusion of the random variable γ_i . The covariance matrix Σ is indexed by x_i to indicate that it changes as a function of the explanatory variables. As with usual univariate linear modeling, we can interpret the coefficients B as explaining differences between experimental conditions. The matrix B is $p \times J$ dimensional, where J is the number of columns in the $n \times J$ dimensional design matrix X . Here $J = 2$ and the second column encodes the contrast between short sleepers and conventional sleepers. The interpretation of B is that small values indicate little heteroscedasticity, identical signs indicates positive correlation, and opposite signs indicate negative correlation. For instance, assume that the second column of B is $b_2 = (-1, 3, 0, 2)^T$. The interpretation

for these four regions is as follows: region one and two are negatively correlated, so are region one and four, region two and four are positively correlated, and region three is uncorrelated.

The general form of this model was introduced by Hoff and Niu (2012) with the idea of decomposing covariance matrices into covariates explained and unexplained terms. In this original form the unexplained part is parametrized as a full covariance matrix scaling quadratically in the number of regions, i.e., $p(p-1)/p$ parameters. Instead, we parametrize it as a diagonal matrix with independent variance terms for each region. This simplified model scales linearly in the number of regions p and can therefore be applied to large brain parcellations.

We use flat priors on both parameters σ and B . The elements of the B matrix have a uniform prior on $(-\infty, \infty)$, and the elements of σ vector have a uniform prior on $(0, \infty)$. These priors are improper and do not integrate to one over their support. In case of prior knowledge, it is preferable to use more informative priors. For large p , we can add an additional hierarchical level to adjusting for multiple testing by including a common inclusion probability per column in B (Scott and Berger, 2006, 2010).

As is common in univariate linear modeling, it is possible to encode additional explanatory variables such as subject ID and possible batch factors. It would also be possible to extend the model to include temporal dependencies in the form of spline coefficients. We have not done so here because we wanted to focus explicitly on functional connectivity between regions.

2.2.2. Effective Sample Size

We subsample n time points to obtain the Effective Sample Size (ESS). This n is smaller than the total number N of time points because it accounts for temporal dependency. We propose a procedure to automatically choose n using an autocorrelation estimate of the timeseries. This is current practice in the field of Markov chain Monte Carlo and implemented in R package coda (Plummer et al., 2006). The ESS describes how much

a dependent sample is worth with respect to an independent sample of the same size. Kass et al. (1998) define ESS via the lag t autocorrelation $\text{Corr}(y_1^{(j)}, y_{1+t}^{(j)})$ as:

$$n = \min_{j=1, \dots, p} \left(\frac{N}{1 + 2 \sum_{t=1}^{\infty} \text{Corr}(y_1^{(j)}, y_{1+t}^{(j)})} \right).$$

This is a component-wise definition and we follow a conservative approach by taking the minimum over all p components as the overall estimator. Intuitively, the larger the autocorrelation the lower is our ESS because we can predict future from current time points. A convenient side-product of subsampling is reduced computational costs.

2.2.3. Inference

We implement our model in the probabilistic programming language Stan (Carpenter et al., 2017) using R. Stan uses Hamiltonian Monte Carlo to sample efficiently from posterior distributions using automatic differentiation. It removes the need for manually deriving gradients of the posterior distributions, thus making it easy to extend models. Our Stan code is available in our new R package `CovRegFC` from our GitHub repository. Alternatively, using conjugate priors it is possible to derive a Gibbs sampler to sample from the posterior distribution of a related model as in Hoff and Niu (2012). However, this makes it harder to extend the model.

Due to the non-identifiability of matrix \mathbf{B} up to random sign changes, \mathbf{B} and $-\mathbf{B}$ corresponding to the same covariance function, we need to align the posterior samples coming from multiple chains. A general option is to use Procrustes alignment. Procrustes alignment (Korth and Tucker, 1976) is a method for landmark registration (Kendall, 1984; Bookstein, 1986) in the shape statistics literature and an implementation is available in the R package `shape` (Dryden and Mardia, 1998).

2.3. Full Covariance Regression

In this section, we introduce a full covariance linear model.

2.3.1. Model

Here we do not subsample and deal with temporal dependencies in a different way. In this model, the number of observations are the number of subjects $k = 1, \dots, K$. After column-wise centering of each $N \times p$ (recall that N is the total number of time points) timeseries $\mathbf{Y}_1, \dots, \mathbf{Y}_K$, we compute sample covariance matrices for each subject $\mathbf{S}_1 = \mathbf{Y}_1^T \mathbf{Y}_1, \dots, \mathbf{S}_K = \mathbf{Y}_K^T \mathbf{Y}_K$. We take this as our “observed” response. Additionally, we have one explanatory vector $\mathbf{x}_1, \dots, \mathbf{x}_n$ for each response covariance matrix. In our HCP data subset, we have 730 subjects, so $K = 730$ and we have K data point pairs $(\mathbf{S}_1, \mathbf{x}_1), \dots, (\mathbf{S}_K, \mathbf{x}_K)$. We assume that the explanatory vector has two elements: the first element $x_k^{(1)}$ representing the intercept and is equal to one, and the second element $x_k^{(2)}$ is one for short and zero for conventional sleepers. Our regression model:

$$\mathbf{S}_k \sim \text{Wishart}(\mathbf{x}_k^{(1)} \boldsymbol{\Sigma}^{(1)} + \mathbf{x}_k^{(2)} \boldsymbol{\Sigma}^{(2)}, \nu)$$

decomposes the “observed” covariance matrix into an intercept term and a term encoding the functional connectivity between sleepers. The second parameter in the Wishart distribution describes the degrees of freedom and has support $(p - 1, \infty)$.

We will now describe how to draw samples from the Wishart distribution, this will give us a better intuition for the proposed model. Matrices following a Wishart distribution can be generated by drawing vectors $\mathbf{y}_1, \dots, \mathbf{y}_N$ independently from a $\text{Normal}(0, \boldsymbol{\Sigma})$, storing vectors in a $N \times p$ matrices \mathbf{Y}_i , and computing the sample covariance matrix $\mathbf{S}_i = \mathbf{Y}_i^T \mathbf{Y}_i$. Then, the constructed \mathbf{S}_i 's are distributed according to a Wishart distribution with parameters $\boldsymbol{\Sigma}$ and degrees of freedom N . If the ESS is smaller than N it will be reflected in the degrees of freedom parameter ν . In our model, we will estimate ν from the data. In this way, we account for the temporal dependencies in the timeseries. The marginal posterior distribution of ν will be highly concentrated around a small degree of freedom (close to p) for strongly dependent samples and concentrated around a large degree of freedom (close to N) for weakly dependent samples.

To complete our model description, we need to put priors on covariance matrices and the degrees of freedom. We decompose the covariance prior into a standard deviation $\boldsymbol{\sigma}$ vector and a correlation matrix $\boldsymbol{\Omega}$ for each term:

$$\boldsymbol{\Sigma}^{(1)} = \boldsymbol{\sigma}^{(1)} \mathbf{I}_p \boldsymbol{\Omega}^{(1)} \boldsymbol{\sigma}^{(1)} \mathbf{I}_p \quad \text{and} \quad \boldsymbol{\Sigma}^{(2)} = \boldsymbol{\sigma}^{(2)} \mathbf{I}_p \boldsymbol{\Omega}^{(2)} \boldsymbol{\sigma}^{(2)} \mathbf{I}_p$$

and put a Lewandowski, Kurowicka, and Joe (LKJ) prior on the correlation matrix (Lewandowski et al., 2009) independently for each term:

$$\boldsymbol{\Omega}^{(1)} \sim \text{LKJcorr}(\eta) \quad \text{and} \quad \boldsymbol{\Omega}^{(2)} \sim \text{LKJcorr}(\eta).$$

This correlation matrix prior has one parameter η that defines the amount of expected correlations. To gain intuition about η , we draw samples from the prior for a range of dimensions and parameter settings (Figure 2). The behavior in two dimension is similar to a beta distribution putting mass on either the boundary of the support of the prior or in the center. As we move toward higher dimensions, we can see that the distribution is less sensitive to the parameter η . For our model, we set $\eta = 1$ to enforce a flat prior. We complete our prior description by putting independent flat priors on both the vector of standard deviations $\boldsymbol{\sigma}$ and the degrees of freedom ν , i.e., uniform prior on $(0, \infty)$ and uniform prior on $(p - 1, N - 1)$, respectively.

2.3.2. Inference

The number of parameters in the model scales quadratically in the number of regions making this model applicable in the classical statistical setting where we have larger sample sizes than number of predictors. In section 3.1, we will show an application to the HCP data with $K = 730$ subjects and $p = 15$ regions. Note, Hoff (2009) devised a Gibbs sampler for a similar model using an eigenmodel for the subject-level covariance matrices.

2.3.3. Posterior Analysis and Multiplicity Control

After drawing samples from the posterior, we can evaluate the marginal posterior distributions of standard deviations $\boldsymbol{\sigma}$, correlations $\boldsymbol{\Omega}$, and degrees of freedom ν . As mentioned, we

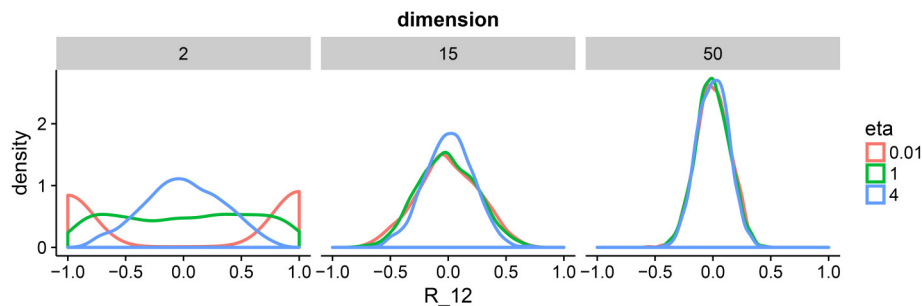


FIGURE 2 | Density plots of 1,000 draws from the LKJ prior in 2, 15, and 50 dimensions. We plot one off-diagonal element in the correlation matrix for each matrix R drawn from the LKJ prior. We pick element (1, 2) denoted by R_{12} . This is representative of the other elements in the matrix because the LKJ prior is symmetric. We evaluate different shape parameters η . In the low-dimensional setting, a small η assigns larger probability mass to correlations close to -1 and 1 , whereas a large η assigns larger mass around 0 . In the high-dimensional setting, probability mass is concentrated around 0 irrespectively of η . The special case of $\eta = 1$ represents the uniform distribution on correlation matrices.

assume that the second element in the explanatory vector encodes whether a subject is a short or a conventional sleepers. In this setting, $\Omega^{(2)}$ represents the difference in correlation between short and conventional sleepers. As we have the marginal posterior distribution for every $\Omega_{ij}^{(2)}$, we can evaluate the probability:

$$P_{ij} = \left| 2 \text{Prob} \left(\Omega_{ij}^{(2)} > 0 \right) - 1 \right|.$$

Our interpretation in terms of connectivity is as follows: If P_{ij} is zero then the correlation is equally probable to be negative or positive. In this case, we are unable to clearly classify the sign of the correlation difference as negative or positive. If P_{ij} is close to one then the correlation is more probable to be either negative or positive. In this case, we can say that parcel i can be seen to be differentially connected to parcel j .

There are $p(p-1)/2$ pairwise correlations and we wish to find correlations that are different between the two groups. If the probability P_{ij} is large, we will report the connection as significantly different. To control for multiple testing, we declare correlations only as significant if they pass a threshold λ . We choose λ to control the posterior expected FDR (Mittra et al., 2016):

$$\text{FDR}_\lambda = \frac{\sum_{ij} (1 - P_{ij}) I(P_{ij} > \lambda)}{\sum_{ij} I(P_{ij} > \lambda)}.$$

We find λ through grid search for a fixed FDR. This allow us to report only correlations that survive the threshold at a given FDR.

3. RESULTS

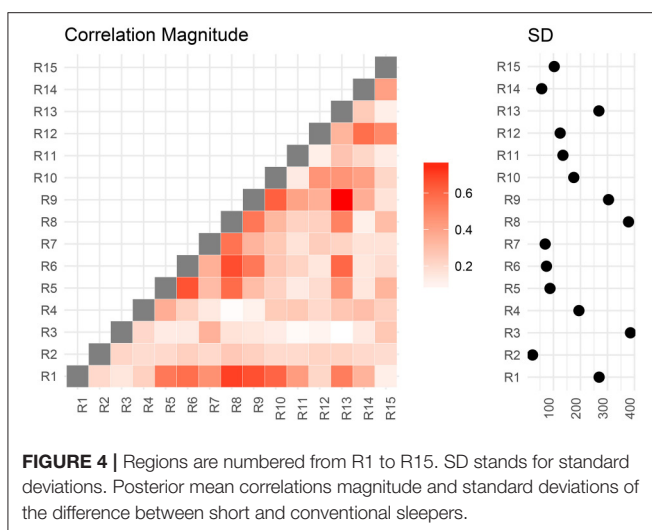
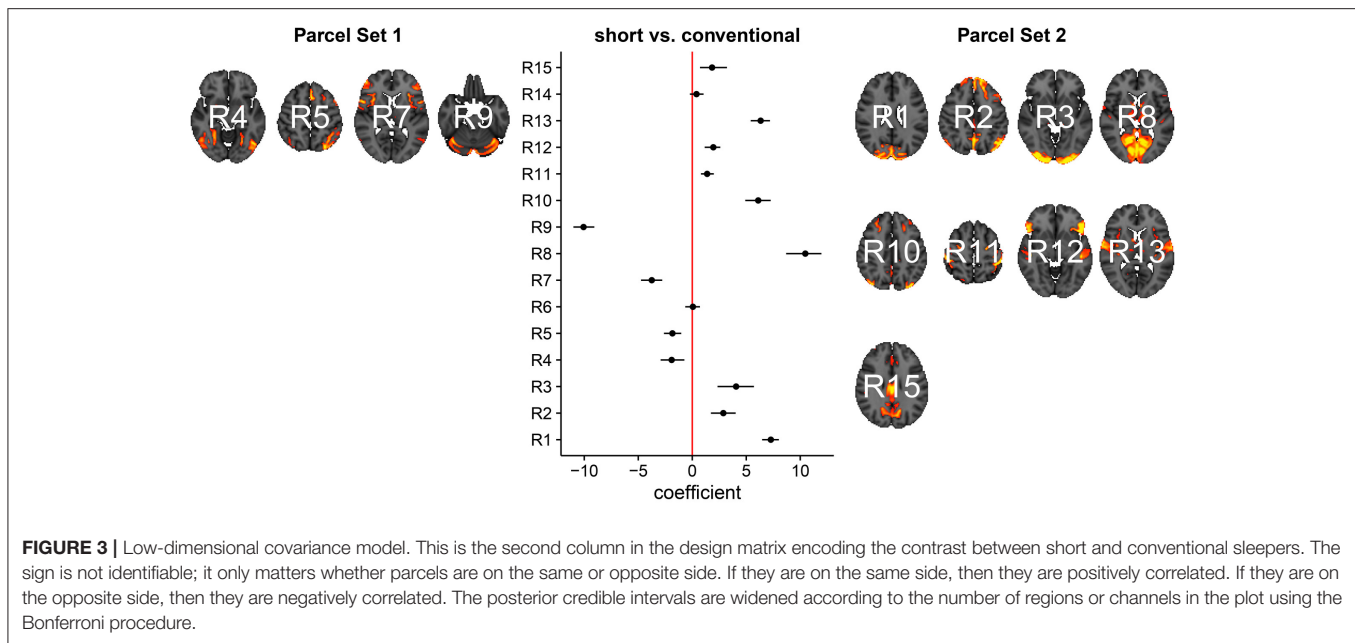
The HCP released a dataset with 820 timeseries of normal healthy subjects measured during resting-state fMRI (rfMRI). The imaging data is accompanied by demographic and behavioral data including a sleep questionnaire. Approximately 30% Americans are reported short sleepers with 4–6 h of sleep per night. The National Sleep Foundation recommends that adults sleep between 7 and 9 h. We use both models to

analyze the HCP data on 730 participants (after subsetting to short and conventional sleepers) to elucidate difference in functional connectivity between short and conventional sleepers. As mentioned before, the design matrix X has an intercept 1 and a column encoding short sleepers 1 and conventional sleepers 0, i.e., conventional sleepers are the reference condition. We use a burn-in of 500 steps during which Stan optimizes tuning parameters for the HMC sampler, e.g., the mass matrix and the integration step length. After burn-in, we run HMC for additional 500 steps. To check convergence, we assess traceplots of random parameter subsets. We obtain an effective sample size of 167 for the 15 regions ICA-based parcellation. We now analyze the marginal posterior distribution of each of the parameters.

3.1. Differential Analysis

3.1.1. Fifteen Parcels

In **Figure 3**, we summarize and visualize the marginal posterior distribution of the second column in B . In the center part of the plot, we show the posterior distribution as posterior medians (dot) and credible intervals containing 95% of the posterior density (segments). The credible intervals are Bonferroni corrected by fixing the segment endpoints at the 0.05/15 and $(1 - 0.05/15)$ quantiles. Care has to be taken when interpreting the location of segments with respect to the zero coefficient line (red line). Due to the sign non-identifiability of B , we have to ignore on which side the segments are located. Recall that regions on the same side are positively correlated, regions on opposite sides are negatively correlation, and regions overlapping the red line are undecided. To relate the region name back to the anatomy, we plotted the most relevant axial slice in the MNI152 space on the left and the right of the coefficient plot, depending on their sign, respectively. We can make the following observations: Parcels in set 1 (R4, R5, R7, and R9) are positively correlated. Keep in mind that the sign of the coefficient carries no information about the sign of the correlation. So, even though the coefficients are negative the correlations are positive, because they are on the same side of the red line. Parcels in set 2 (R1–R3, R8, R10–R13, and R15) are also positively correlated, for the same



reason as before. In contrast, the two parcel sets are negatively correlated, because they are on opposite sides. The connectivity of R6 and R14 are not different from conventional sleepers because their credible intervals overlap the red line. According to the meta analysis in Smith et al. (2009), parcel set 1 is associated with visual, cognition-language, sensorimotor areas, and the cerebellum; and parcel set 2 with visual, cognition-language, auditory areas, and the default network.

We now compare the result from the low-dimensional model with results from the full model. First, we compute the posterior marginal mean of the standard deviations vector $\sigma^{(2)}$ and the correlation matrix magnitude $|\Omega^{(2)}|$ encoding the difference between short and conventional sleepers (Figure 4). The standard deviation plot on the right shows that parcel

R3 varies the most, and that region R2 varies the least. The magnitude correlation plot on the left shows that parcel pair R9 and R13 exhibit the strongest correlation. This is consistent with the low-dimensional model results, where R9 and R13 are in opposite parcel sets. Similarly, parcels R1 and R8 have a strong correlation magnitude in the full model and large effect sizes in the low-dimensional results.

In Figure 5, we assess the significance of differential correlations. The color code indicates different FDR levels. Overall strong differences in the correlation structure are visible with a large portion of connections at an FDR of 0.001. In contrast to the low-dimensional model, these are differences in correlations and not whether they are more positively or more negatively correlated.

3.1.2. Fifty Parcels

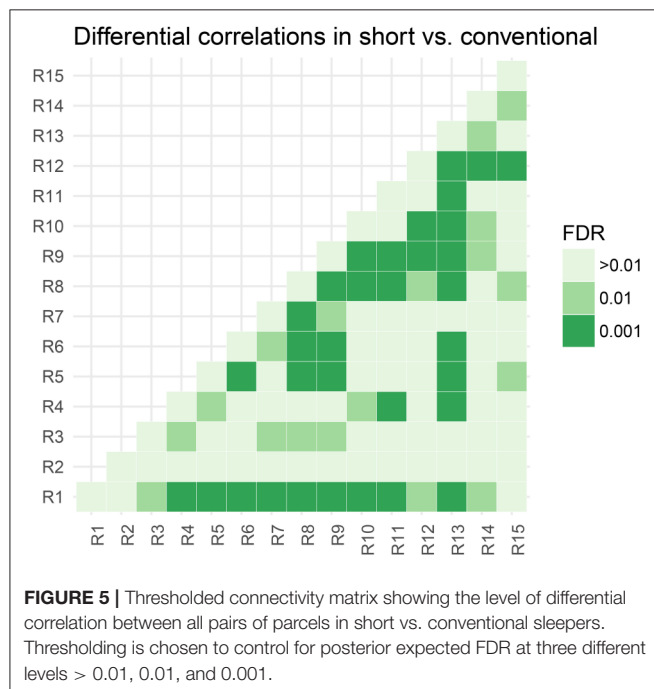
Modeling the data in a more compact representation makes it easier for us to interpret the results and easier to estimate parameters. For instance, consider analyzing $p = 50$ parcels of 160 randomly sampled subjects from the HCP (Figure 6). All the information fits on one plot similar as in the $p = 15$ parcel case. For $p = 50$ it starts to get harder to interpret the full model because we have now $50(50 - 1)/2 = 1225$ possible pairwise correlations. It will be hard to interpret a plot of the full correlation matrix. One way to make sense of it is to cluster rows and columns of the correlation matrix. Even though such post-processing approaches are useful, it is unclear how to propagate uncertainty from the correlation estimation to the clustering step. A low-dimensional model is therefore our preferred analysis approach.

3.1.3. Note on Computation Time

For the low-dimensional model and the available 730 subjects, the computation time for the HMC sampler is around 20 h on

a single core on a modern CPU. For a subsample of 40 subjects, the computation time is around 20–25 min, and for 80 subjects around 50–55 min. It is possible to run more chains in parallel to increase the sample size. To combine each run, we need to align the posterior samples using Procrustes alignment as indicated in the section 2.

The full model takes about 1 h on a single core, and we run four chains in parallel to increase sample size.



3.2. Power Analysis

We design a power analysis (Figure 7) for low-dimensional covariance regression with 15 parcels. As the population we take the available 730 subjects in the HCP data repository that are either short or conventional sleepers and have preprocessed timeseries. We sample 100 times from this population keeping the same ratio between the number of observations for each group, i.e., two thirds conventional and one third short sleepers. We report the average True Positive Rate (TPR) and the False Discovery Rate (FDR) over the 100 samples. We assign a parcel to a parcel set if its credible interval is located on one side of the zero red line and does not overlap the line. The credible intervals contain $100 \times (1 - \alpha)\%$ of the marginal posterior distribution with end points evaluated using quantiles. We need to take into consideration that parcel sets are non-identifiable. We denote the i th predicted parcel set as Z_i^{pred} and the true parcel set as Z_i^{true} . The index i can be either 1 or 2. Parcel sets are subsets of $\{R1, R2, \dots, R15\}$.

With these definitions, we are now ready to calculate TPR and FDR. The TPR measures the power of our procedure to detect true parcels. We define true positives as:

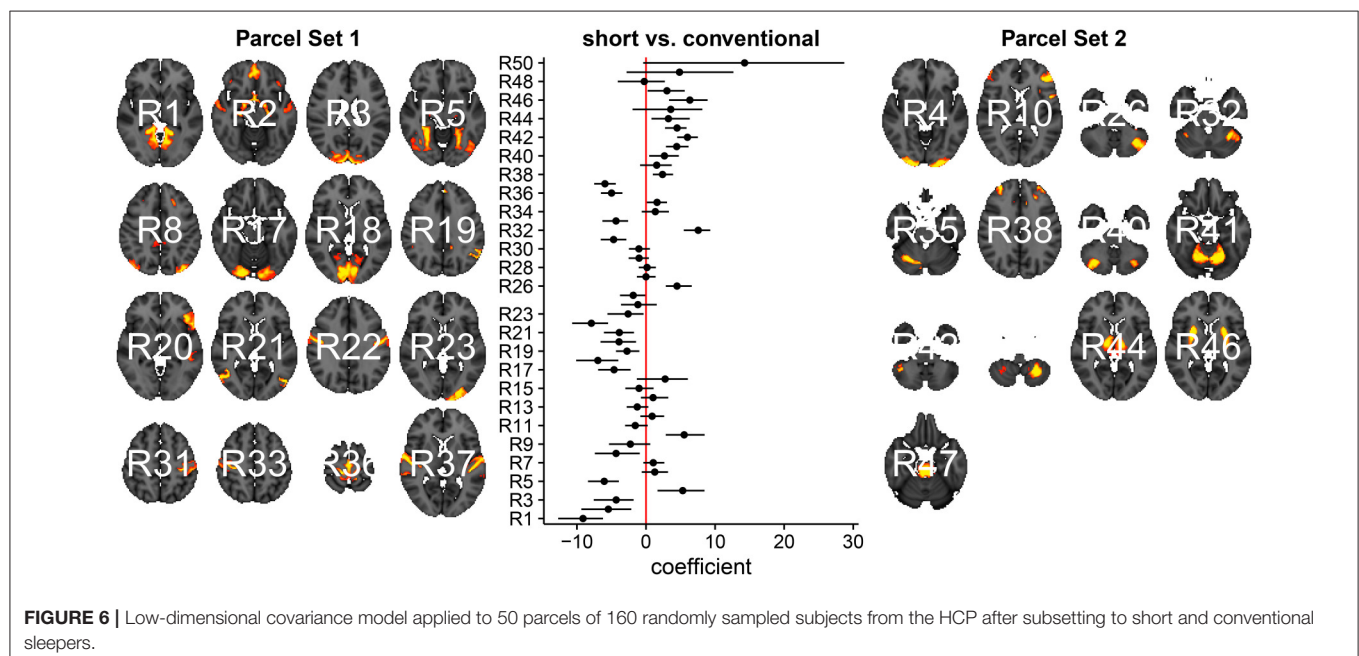
$$TP_{ijkl} = \#(Z_i^{\text{pred}} \cap Z_j^{\text{true}}) + \#(Z_k^{\text{pred}} \cap Z_l^{\text{true}}).$$

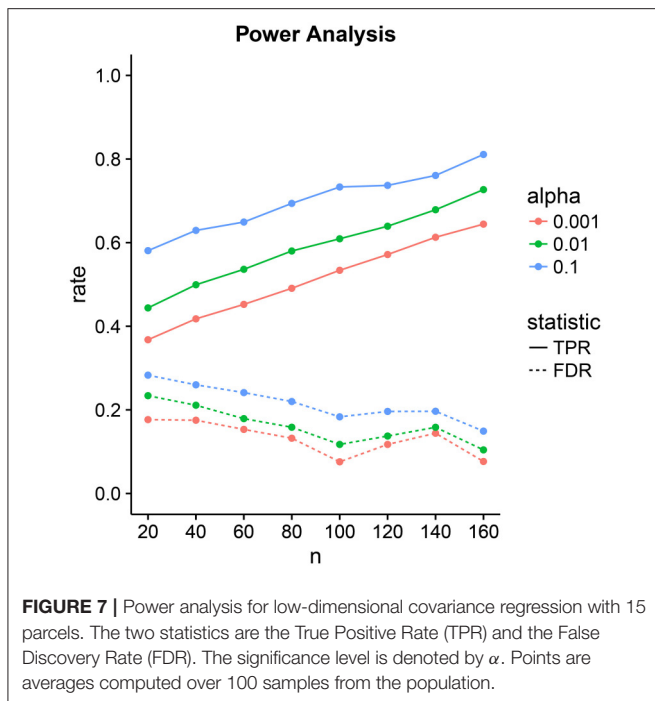
To obtain the rate, we take the maximum of both possible comparisons:

$$\text{Correctly Predicted Parcels} = \max(TP_{1122}, TP_{1221})$$

and divide by the total number of true parcels:

$$TPR = \frac{\text{Correctly Predicted Parcels}}{\text{Total True Parcels}}.$$





The FDR measures the amount of falsely predicted parcels as:

$$FP_{ijkl} = \#(Z_i^{\text{pred}} \setminus Z_j^{\text{true}}) + \#(Z_k^{\text{pred}} \setminus Z_l^{\text{true}})$$

by taking the minimum

$$\text{Falsely Predicted Parcels} = \min(FP_{1122}, FP_{1221})$$

divided by the total number of positives

$$\text{FDR} = \frac{\text{Falsely Predicted Parcels}}{\text{Correctly Predicted Parcels} + \text{Falsely Predicted Parcels}}.$$

The tradeoff between the two can be controlled through the significance level α . Power increases linearly with sample size. FDR decrease linearly but at a lower rate with sample size. At samples size 40, we have a power of 50% with an FDR of 20%. This improves to a power of 80% with an FDR of 10% at sample size 160.

4. DISCUSSION

We introduced two new models for functional connectivity. In particular, the low-dimensional covariance model is able to discover 50% of the correlation differences at a FDR of 20% in a sample size as little as 40. Our Stan implementations make it easy for others to extend our models. We applied both models to the HCP data subset to compare functional connectivity between short and conventional sleepers. Our findings are consistent with Curtis et al. (2016) and Killgore et al. (2012) reporting increases in functional connectivity in short sleepers for primary auditory, primary motor, primary somatosensory, and primary visual

cortices. A similar neural signature was observed in experiments examining the transition from resting wakefulness to sleep onset using EEG and rfMRI (Larson-Prior et al., 2009; Tagliazucchi and Laufs, 2014; Davis et al., 2016). Therefore, we recommend the inclusion of the average sleep duration of a subject as a “batch” covariate in the experimental design of rfMRI studies.

In addition to group comparisons encoded as a design matrix with two columns, it is possible to extend our low-dimensional model to more complicated experimental designs by appending more columns to the design matrix. We can encode batch factors and subject-specific variability by binding one column per factor level. Besides categorical variables, we can model continuous variables such as head-motion measurement made using an accelerometer. Adding covariates to explain unwanted variation in the data can move some of the preprocessing steps to the functional connectivity analysis step. Such joint modeling can enable the propagation of uncertainty to the downstream analyses. Additional columns in the design matrix are called blocking factors and can improve the statistical power. Without modeling the blocking factor, the variability in the data is absorbed by the noise term. The higher level of noise leads to higher uncertainty in our parameter estimates. In contrast, a model with additional blocking factors has more parameters that need to be estimated. As in most practical problems, the right modeling choice depends on the data.

A main challenge in covariance regression is the positive definiteness constraint. A solution is to transform the covariance estimation problem into an unconstrained problem thus making estimation and inference easier (Pourahmadi, 2011). One possible transformation starts with a spectral decomposition where the covariance matrix is decomposed into a diagonal matrix of eigenvalues and an orthogonal matrix with normalized eigenvectors as columns. The procedure continues with a global log-transformation to the covariance matrix, which results in a log-transformation of individual eigenvalues and removes the constraint. However, mathematically and computationally tempting this approach seems, it remains hard to interpret the log-transformations statistically (Brown et al., 1994; Liechty et al., 2004). An alternative transformation uses a Cholesky decomposition of the covariance matrix. For the Cholesky decomposition, we need a natural ordering of the variables not known a priori for functional connectivity data—a natural ordering could be given if the chronology is known.

Modeling of covariance matrices builds on important geometrical concepts and the medical image analysis community has made significant progress in terms of mathematical descriptions and practical applications motivated by data in diffusion tensor imaging (Pennec, 1999, 2006; Moakher, 2005; Arsigny et al., 2006/2007; Lenglet et al., 2006; Fletcher and Joshi, 2007; Fillard et al., 2007; Schwartzman et al., 2008; Dryden et al., 2009). The underlying geometry is called Lie group theory and it appears when we consider the covariance matrices as elements in a non-linear space. The matrix log-transformation from the previous paragraph maps covariance matrices to the tangent space where unconstrained operations can be performed; for instance we create a mean by simple elementwise averaging. After computing the mean in tangent space, this mean is mapped

back to the constrained space of covariance matrices. Despite the mathematical beauty and algorithmic convenience, statistical interpretations are still unwieldy. However, this does provide a fundamental geometric formulation and enables the use of handy geometrical tools (Absil et al., 2008 for a book-length treatment).

We approach the problem from a statistical viewpoint and frame functional connectivity in terms of modeling heteroscedasticity. This allows us to take advantage of the rich history in statistics and led us to the covariance regression model introduced by Hoff and Niu (2012). We simplify the model to meet the large p requirement in neuroscience. The running time for 500 posterior samples on 80 subjects is less than an hour on a single core. This makes our approach applicable to many neuroimaging studies. For larger studies, such as the HCP with 730 subjects, further speed improvements using GPU's are desirable to reduce computation time.

One possible future application is functional Near-Infrared Spectroscopy (fNIRS), which has gained in popularity due its portability and high temporal resolution. A common approach is to set up a linear model between brain responses at channels locations (Huppert et al., 2009; Ye et al., 2009; Tak and Ye, 2014) and experimental conditions. Thus, our models apply to fNIRS experiments. An additional challenge in fNIRS experiments is channel registration across multiple participants (Liu et al., 2016). Connectivity differences could be due artifacts created by channel misalignments not biology. In the absence of structural MRI, we could add an additional hierarchical level in our low-dimensional model to handle measurement errors accounting for possible misalignments between channels.

We use a conservative component-wise estimate of the ESS. Less conservative multivariate estimators (Vats et al., 2015) might be able to increase statistical power at the cost of an increase in the false discovery rate.

REPRODUCIBILITY AND SUPPLEMENTARY MATERIAL

The entire data analysis workflow is available on our GitHub repository:

- https://github.com/ChristofSeiler/CovRegFC_HCP

We also provide a new R package `CovRegFC` with Stan code:

- <https://github.com/ChristofSeiler/CovRegFC>

Data preparation and statistical analyses are contained in Rmd files:

- `Low_Dimensional.Rmd`
- `Full.Rmd`
- `Power.Rmd`

By running these files all results and plots can be completely reproduced as html files:

- `Low_Dimensional.html`
- `Full.html`
- `Power.html`

The HCP data is available here:

- <https://www.humanconnectome.org/data/>

AUTHOR CONTRIBUTIONS

CS wrote an initial draft, performed and implemented the statistical analysis. SH wrote the final manuscript and provided statistical tools.

ACKNOWLEDGMENTS

CS was partially funded by two Swiss NSF postdoctoral fellowships 146281 and 158500. SH was partially supported by NSF DMS grant 1501767. We thank the NIRS lab at the Center for Interdisciplinary Brain Sciences in the Stanford School of Medicine for introducing us to functional neuroimaging data in the context of fNIRS experiments.

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

REFERENCES

- Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ: Princeton University Press.
- Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006
- Anderson, T. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Stat.* 1, 135–141. doi: 10.1214/aos/1193342389
- Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006/2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 29, 328–347. doi: 10.1137/050637996
- Bassett, D. S., and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* 12, 512–523. doi: 10.1177/1073858406293182
- Beckmann, C. F., and Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23, 137–152. doi: 10.1109/TMI.2003.822821
- Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., et al. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739. doi: 10.1073/pnas.0911855107
- Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions. *Stat. Sci.* 1, 181–222. doi: 10.1214/ss/1177013696
- Brown, P. J., Le, N. D., and Zidek, J. V. (1994). “Inference for a covariance matrix,” in *Aspects of Uncertainty: A Tribute to D.V. Lindley*, eds P. Freeman and A. Smith (Chichester: Wiley), 77–92.
- Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Stat. Soft.* 76, 1–32. doi: 10.18637/jss.v076.i01

- Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P., and Mayberg, H. S. (2012). A whole brain fmri atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33, 1914–1928. doi: 10.1002/hbm.21333
- Curtis, B. J., Williams, P. G., Jones, C. R., and Anderson, J. S. (2016). Sleep duration and resting fmri functional connectivity: examination of short sleepers with and without perceived daytime dysfunction. *Brain Behav.* 6:e00576. doi: 10.1002/brb3.576
- Davis, B., Tagliazucchi, E., Jovicich, J., Laufs, H., and Hasson, U. (2016). Progression to deep sleep is characterized by changes to bold dynamics in sensory cortices. *Neuroimage* 130, 293–305. doi: 10.1016/j.neuroimage.2015.12.034
- Dosenbach, N. U., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361. doi: 10.1126/science.1194144
- Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Ann. Appl. Stat.* 3, 1102–1123. doi: 10.1214/09-AOAS249
- Dryden, I. L., and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley Series in Probability and Statistics: Probability and Statistics. Chichester: John Wiley & Sons, Ltd.
- Eguíluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. (2005). Scale-free brain functional networks. *Phys. Rev. Lett.* 94:018102. doi: 10.1103/PhysRevLett.94.018102
- Fillard, P., Pennec, X., Arsigny, V., and Ayache, N. (2007). Clinical dt-mri estimation, smoothing, and fiber tracking with log-Euclidean metrics. *IEEE Trans. Med. Imaging* 26, 1472–1482. doi: 10.1109/TMI.2007.899173
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396
- Fletcher, P. T., and Joshi, S. (2007). Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.* 87, 250–262. doi: 10.1016/j.sigpro.2005.12.018
- Fox, M. D., and Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* 4:19. doi: 10.3389/fnsys.2010.00019
- Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711. doi: 10.1038/nrn2201
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441. doi: 10.1093/biostatistics/kxm045
- Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., and Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *Ann. Appl. Stat.* 11, 725–750. doi: 10.1214/16-AOAS1015
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127
- Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr. opin. Neurol.* 21, 424–430. doi: 10.1097/WCO.0b013e328306f2c5
- Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100
- Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, et al. (2014). ICA-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *Neuroimage* 95, 232–247. doi: 10.1016/j.neuroimage.2014.03.034
- Grillon, M.-L., Oppenheim, C., Varoquaux, G., Charbonneau, F., Devauchelle, A.-D., Krebs, M.-O., et al. (2013). Hyperfrontality and hypoconnectivity during refreshing in schizophrenia. *Psychiatry Res.* 211, 226–233. doi: 10.1016/j.psychres.2012.09.001
- Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., et al. (2015). National sleep foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health* 1, 40–43. doi: 10.1016/j.sleh.2014.12.010
- Hoff, P. D. (2009). A hierarchical eigenmodel for pooled covariance estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71, 971–992. doi: 10.1111/j.1467-9868.2009.00716.x
- Hoff, P. D., and Niu, X. (2012). A covariance regression model. *Statist. Sinica* 22, 729–753. doi: 10.5705/ss.2010.051
- Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48, D280–D298. doi: 10.1364/AO.48.00D280
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10, 626–634. doi: 10.1109/72.761722
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov chain monte carlo in practice: a roundtable discussion. *Am. Stat.* 52, 93–100. doi: 10.2307/2685466
- Kendall, D. G. (1984). Shape manifolds, Procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* 16:81–121. doi: 10.1112/blms/16.2.81
- Killgore, W. D., Schwab, Z. J., and Weiner, M. R. (2012). Self-reported nocturnal sleep duration is associated with next-day resting state functional connectivity. *Neuroreport* 23, 741–745. doi: 10.1097/WNR.0b013e3283565056
- Korth, B., and Tucker, L. R. (1976). Procrustes matching by congruence coefficients. *Psychometrika* 41, 531–535. doi: 10.1007/BF02296973
- Larson-Prior, L. J., Zempel, J. M., Nolan, T. S., Prior, F. W., Snyder, A. Z., and Raichle, M. E. (2009). Cortical network functional connectivity in the descent to sleep. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4489–4494. doi: 10.1073/pnas.0900924106
- Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multiv. Anal.* 88, 365–411. doi: 10.1016/S0047-259X(03)00096-4
- Lenglet, C., Rousson, M., and Deriche, R. (2006). DTI segmentation by statistical surface evolution. *IEEE Trans. Med. Imaging* 25, 685–700. doi: 10.1109/TMI.2006.873299
- Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multiv. Anal.* 100, 1989–2001. doi: 10.1016/j.jmva.2009.04.008
- Lewis, C. M., Baldassarre, A., Comitteri, G., Romani, G. L., and Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17558–17563. doi: 10.1073/pnas.0902455106
- Liang, X., Connelly, A., and Calamante, F. (2016). A novel joint sparse partial correlation method for estimating group functional networks. *Hum. Brain Mapp.* 37, 1162–1177. doi: 10.1002/hbm.23092
- Liechty, J. C., Liechty, M. W., and Müller, P. (2004). Bayesian correlation estimation. *Biometrika* 91, 1–14. doi: 10.1093/biomet/91.1.1
- Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., and Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Front. Hum. Neurosci.* 10:82. doi: 10.3389/fnhum.2016.00082
- Marrelec, G., Bellec, P., Krainik, A., Duffau, H., Péligrini-Issac, M., Lehericy, S., et al. (2008). Regions, systems, and the brain: hierarchical measures of functional integration in fmri. *Med. Image Anal.* 12, 484–496. doi: 10.1016/j.media.2008.02.002
- Mitra, R., Müller, P., and Ji, Y. (2016). Bayesian graphical models for differential pathways. *Bayesian Anal.* 11, 99–124. doi: 10.1214/14-BA931
- Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* 26, 735–747. doi: 10.1137/S0895479803436937
- Pennec, X. (1999). “Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements,” in *NSIP*, eds A. Enis Çetin, L. Akarun, A. Ertüzün, M. N. Gurcan, and Y. Yardimci (Bogaziçi University Printhouse), 194–198.
- Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vis.* 25, 127–154. doi: 10.1007/s10851-006-6228-4
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R. News* 6, 7–11.
- Pourahmadi, M. (2011). Covariance estimation: the GLM and regularization perspectives. *Stat. Sci.* 26, 369–387. doi: 10.1214/11-STS358

- Preti, M. G., Bolton, T. A., and Ville, D. V. D. (2016). The dynamic functional connectome: state-of-the-art and perspectives. *Neuroimage* 160, 41–54. doi: 10.1016/j.neuroimage.2016.12.061
- Ryali, S., Chen, T., Supekar, K., and Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* 59, 3852–3861. doi: 10.1016/j.neuroimage.2011.11.054
- Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046
- Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., and Bullmore, E. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332–1342. doi: 10.1093/cercor/bhi016
- Schwartzman, A., Dougherty, R. F., and Taylor, J. E. (2008). False discovery rate analysis of brain diffusion direction maps. *Ann. Appl. Stat.* 2, 153–175. doi: 10.1214/07-AOAS133
- Scott, J. G., and Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing. *J. Stat. Plann. Infer.* 136, 2144–2162. doi: 10.1016/j.jspi.2005.08.031
- Scott, J. G., and Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.* 38, 2587–2619. doi: 10.1214/10-AOS792
- Smith, S. M. (2012). The future of fMRI connectivity. *Neuroimage* 62, 1257–1266. doi: 10.1016/j.neuroimage.2012.01.022
- Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168. doi: 10.1016/j.neuroimage.2013.05.039
- Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045. doi: 10.1073/pnas.0905267106
- Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., and Beckmann, C. F. (2014). Group-PCA for very large fMRI datasets. *Neuroimage* 101, 738–749. doi: 10.1016/j.neuroimage.2014.07.051
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063
- Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci.* 8, 418–425. doi: 10.1016/j.tics.2004.07.008
- Stam, C. J. (2004). Functional connectivity patterns of human magnetoencephalographic recordings: a small-worldnetwork? *Neurosci. Lett.* 355, 25–28. doi: 10.1016/j.neulet.2003.10.063
- Tagliazucchi, E., and Laufs, H. (2014). Decoding wakefulness levels from typical fmri resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* 82, 695–708. doi: 10.1016/j.neuron.2014.03.020
- Tak, S., and Ye, J. C. (2014). Statistical analysis of fNIRS data: a comprehensive review. *Neuroimage* 85, 72–91. doi: 10.1016/j.neuroimage.2013.06.016
- Van Den Heuvel, M. P., and Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010.03.008
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041
- Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., and Thirion, B. (2010a). “Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Beijing: Springer), 200–208.
- Varoquaux, G., and Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *Neuroimage* 80, 405–415. doi: 10.1016/j.neuroimage.2013.04.007
- Varoquaux, G., Gramfort, A., Poline, J.-B., and Thirion, B. (2010b). “Brain covariance selection: better individual functional connectivity models using population prior,” in *Advances in Neural Information Processing Systems* (Vancouver), 2334–2342.
- Varoquaux, G., Gramfort, A., Poline, J. B., and Thirion, B. (2012). Markov models for fMRI correlation structure: is brain functional connectivity small world, or decomposable into networks? *J. Physiol.* 106, 212–221. doi: 10.1016/j.jphysparis.2012.01.001
- Vats, D., Flegal, J. M., and Jones, G. L. (2015). Multivariate output analysis for Markov chain Monte Carlo. *arXiv preprint arXiv:1512.07713*.
- Ye, J. C., Tak, S., Jang, K. E., Jung, J., and Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage* 44, 428–447. doi: 10.1016/j.neuroimage.2008.08.036

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Seiler and Holmes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Bayesian Spatial Model to Predict Disease Status Using Imaging Data From Various Modalities

Wenqiong Xue^{1*}, F. DuBois Bowman² and Jian Kang³

¹ Boehringer Ingelheim Pharmaceuticals Inc., Ridgefield, CT, United States, ² Department of Biostatistics, The Mailman School of Public Health, Columbia University, New York, NY, United States, ³ Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:

Xiaoying Tang,
Center for Imaging Science,
United States

Reviewed by:

Russell Shinohara,
University of Pennsylvania,
United States
Qingpeng Zhang,
City University of Hong Kong,
Hong Kong

*Correspondence:

Wenqiong Xue
nora.xue@gmail.com

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 19 January 2017

Accepted: 06 March 2018

Published: 26 March 2018

Citation:

Xue W, Bowman FD and Kang J
(2018) A Bayesian Spatial Model to
Predict Disease Status Using Imaging
Data From Various Modalities.
Front. Neurosci. 12:184.
doi: 10.3389/fnins.2018.00184

Relating disease status to imaging data stands to increase the clinical significance of neuroimaging studies. Many neurological and psychiatric disorders involve complex, systems-level alterations that manifest in functional and structural properties of the brain and possibly other clinical and biologic measures. We propose a Bayesian hierarchical model to predict disease status, which is able to incorporate information from both functional and structural brain imaging scans. We consider a two-stage whole brain parcellation, partitioning the brain into 282 subregions, and our model accounts for correlations between voxels from different brain regions defined by the parcellations. Our approach models the imaging data and uses posterior predictive probabilities to perform prediction. The estimates of our model parameters are based on samples drawn from the joint posterior distribution using Markov Chain Monte Carlo (MCMC) methods. We evaluate our method by examining the prediction accuracy rates based on leave-one-out cross validation, and we employ an importance sampling strategy to reduce the computation time. We conduct both whole-brain and voxel-level prediction and identify the brain regions that are highly associated with the disease based on the voxel-level prediction results. We apply our model to multimodal brain imaging data from a study of Parkinson's disease. We achieve extremely high accuracy, in general, and our model identifies key regions contributing to accurate prediction including caudate, putamen, and fusiform gyrus as well as several sensory system regions.

Keywords: Bayesian spatial model, prediction, MCMC, posterior predictive probability, importance sampling, Parkinson's disease

1. INTRODUCTION

Functional and structural neuroimaging play important roles in understanding the neurological basis for major psychiatric and neurological disorders such as Parkinson's disease (PD), schizophrenia, depression, and Alzheimer's diseases. There is emerging interest in using imaging and other clinical data to forecast or blindly classify subjects into subgroups, for example, defined by disease status or more refined diagnostic categories. Classification or prediction of disease status based on imaging data remains an active area of research and holds promise for making a significant clinical impact. Prediction models may have a range of applications and be beneficial for clinical diagnosis, determining antecedents to a standard diagnosis, forecasting prognosis, and revealing the underlying neural basis of disease, thus informing the development of future treatments.

We use data from a study of PD as a motivating example for our proposed methods (see section 2). Neuroimaging has revealed various functional and structural alterations associated with PD. There have been reports of cortical thinning in PD patients determined from T1-MRI scans (Lee et al., 2013; Zarei et al., 2013; Zhang et al., 2015), decreased fractional anisotropy in the substantia nigra revealed by diffusion tensor imaging (DTI) (Vaillancourt et al., 2009), and functional connectivity, structural connectivity, and volumetric PD-related changes revealed by a multimodal imaging analysis (Bowman et al., 2016). These and other related studies suggest the utility of imaging data in revealing neuropathophysiology related the loss of dopamine producing neurons in PD and prompt the need for new methods to accommodate high-dimensional multimodal data.

Regularization and variable selection methods such as the least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) and elastic-net (Zou and Hastie, 2005) as well as support vector classifiers are commonly used to predict a single scalar-valued outcome from high-dimensional data. Support vector classifiers, which arise from support vector machines (SVM), classify the data by constructing an optimal separating hyperplane in a high dimensional space to which the data are mapped (Cortes and Vapnik, 1995). Gaussian process (GP) models provide an alternative approach, which finds the posterior function that is closest to the training data based on Bayesian theory (Marquand et al., 2010). Ham and Kwak (2012) propose a boosted-principal component analysis (PCA) algorithm for binary classification problems that combines feature selection and classification.

Several methods have been proposed to predict follow-up imaging scans from baseline scans (Guo et al., 2008; Derado et al., 2013). Guo et al. (2008) propose a Bayesian hierarchical model for functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) data; Derado et al. (2013) extends the model by introducing both spatial correlations between voxels and temporal correlations between baseline and follow-up functional imaging scans. For structural data, Stonnington et al. (2010) propose a relevance vector regression (RVR) model to predict the clinical scores using MRI T1 weighted scans.

Predicting disease status utilizes a potentially massive number of independent variables that exhibit unknown patterns of correlation. The prediction and classification models described above do not estimate the spatial correlations in imaging scans or capture the associations between different imaging modalities. We build on ideas of spatial modeling for correlated imaging data for our prediction framework. Specifically, we propose a novel Bayesian hierarchical model to predict disease status using imaging scans of different modalities in both gray and white matter to reflect the functional as well as the structural properties of the brain. We consider a two-level brain parcellation, dividing the brain into defined regions as well as subregions within regions, and assume different spatial correlation structures between voxels within a subregion, within a region, and in different regions. We perform Markov Chain Monte Carlo (MCMC) estimations via Gibbs sampling. The predictions for disease status are conducted based on the predictive posterior

probabilities. Both whole-brain and voxel-level predictions are performed using leave-one-out cross validation (LOOCV). Also, we conduct feature selection to identify the regions that are associated with the disease based on the voxel-level prediction results. We apply our approach to a PD study and conduct simulation studies to evaluate its performance.

2. PARKINSON'S DISEASE DATA

This research qualifies as Research of Existing Data, Records, Specimens [Basic Exempt Criteria 45 CFR 46.101(b)(4)], and has been deemed Not Human Subjects Research (HS Code 10 in IPMAC II as reference in the manual chapter 7410) by NIH and Columbia University Medical Center Institutional Review Board (Protocol: IRB-AAA00062).

The data were originally collected at Emory University (P50-NS071669) and were supplied to Columbia with all subjects' records de-identified. Written and informed consent was obtained from all research participants at the time of data collection.

A total of 20 subjects, 11 of which are diagnosed as early to moderate PD patients and the rest are healthy controls, are included in the study. The average age is 66 (*s.d.* = 11) years, and 12 of the subjects are males. The mean duration of disease was 8.4 years (*s.d.* = 3.3). The average height is 175 cm and the average weight is 79 kg. Resting-state fMRI scans, and T1-weighted MRI scans, and diffusion tensor imaging (DTI) scans are obtained.

A Siemens Trio Tim 3T MRI scanner was used to capture all the imaging scans. MPRAGE was used to acquire the structural T1 scans (*TR* = 2,600 ms, *TE* = 3 ms, 192 sagittal slices at 1 mm; 256×232 1 mm isotropic pixels). The resting-state fMRI scans were acquired using echo planar imaging (EPI) (*TR* = 3,000 ms, *TE* = 30 ms, 48 axial slices at 3 mm, 128×128 2 mm isotropic pixels) for each subject. DTI data were captured using a biphasic approach with consecutive left-to-right and right-to-left phase scans. The subjects followed a DTI protocol (*TR* = 8,700 ms, *TE* = 94 ms, 64 axial slices at 2 mm, 128×128 2 mm isotropic pixels) comprised of 64 directions (*B* = 1,000 s/mm²), with three leading and three trailing B0 scans.

We extract voxel-level information from these three imaging modalities, including fractional amplitude of low-frequency fluctuation (fALFF) from resting-state fMRI scans, voxel based morphometry (VBM) from T1-weighted MRI scans, and fractional anisotropy (FA) from DTI scans. fALFF reflects the amplitude of spontaneous blood-oxygen-level-dependent (BOLD) signal fluctuations of each voxel. VBM measures the localized gray matter volume changes in each voxel after spatially normalizing all the images to a standard space, and extracting gray matter from the normalized images (Ashburner and Friston, 2000). FA has a single value for each voxel, measuring the difference in directions along different axes of the random motion of water molecules in the brain, which reflects the physical orientation of white-matter fibers at that location. In summary, fALFF provides functional information, while FA and VBM describe structural properties of the brain.

The image preprocessing was performed using statistical parametric mapping (SPM) (Wellcome Department of Cognitive

Neurology, <http://www.fil.ion.ucl.ac.uk/spm>) and FMRIB (Functional Magnetic Resonance Imaging of the Brain) Software Library (FSL) (Smith et al., 2004). Resting state preprocessing consisted of a despiking stage, slice time correction, motion correction, spatial normalization to MNI and smoothing by 6 mm FWHM. The time courses were filtered to the band 0.01–0.1 Hz.

3. METHODS

We propose a novel Bayesian hierarchical model to predict disease status using imaging data from different modalities, including fALFF, VBM, and FA. For resting-state fMRI scans and DTI scans, the functional and structural information lies in gray matter and white matter, respectively. Most VBM analyses focus on gray matter, which will be the focus of our upcoming data example; however, applications of VBM in white matter has also been found to be associated with psychiatric diseases such as Alzheimer's diseases and schizophrenia (Di et al., 2009; Li et al., 2011). Potentially, our prediction model involves the voxels in gray and/or white matter for different imaging modalities.

3.1. Model and Estimation

We consider a two-level brain parcellation, initially consisting of $G = 90$ brain regions defined by the automated anatomic labeling (AAL) system (Tzourio-Mazoyer et al., 2002). In each region g , we define L_g subregions, ranging from 1 to 9, for $g = 1, \dots, G$. The subregions are built based on the brain parcellation algorithms described in Appendix 1 (Supplementary Material). Each subregion l is composed of V_l voxels. Let $X_{ilg}(v)$, $Y_{ilg}(v)$ and $Z_{ilg}(v)$ respectively denote the observed fALFF, FA and VBM measures for subject i at voxel v in subregion l and region g , for $i = 1, \dots, n$, $v = 1, \dots, V_l$, $l = 1, \dots, L_g$. Let $N_g(l) \subseteq \{1, \dots, L_g\}$ denote the neighboring subregions of subregion l , constrained to fall within region g , and n_{lg} is the number of members in $N_g(l)$. In our model, all the subregions in region g are considered as neighbors of subregion l ; therefore, we have $N_g(l) = \{1, \dots, L_g\}$, and $n_{lg} = L_g$. Let $D_i \in \{0, 1\}$ denote the disease status (here, PD), where 1 indicates PD; and $\mathbf{W}_i = (W_{i1}, \dots, W_{iQ})$ denotes the vector of Q covariates. Let \mathcal{B} , \mathcal{W} and \mathcal{G} respectively represent the whole brain region, the white matter region and the gray matter region.

We propose a model that accounts for the spatial correlations between voxels within the same subregion, between subregions within the same region, and between regions. Building spatial correlations into our model captures associations between different brain regions and generally improves the precision of estimates by borrowing strength from other (sub)regions. First, our model assumes consistent correlations between voxels in a same subregion. Then the spatial correlations between subregions within the same AAL region are described by a conditional autoregressive (CAR) model, which allows the estimates at subregion levels to borrow strength from their neighbors within the same AAL region. In addition, we introduce unstructured spatial correlations between AAL regions.

Our model reflects the assumption that for each voxel v in the gray matter, the fALFF value $X_{ilg}(v)$ follows a normal distribution

conditioning on the VBM value $Z_{ilg}(v)$; for each voxel v in the white matter, the FA value $Y_{ilg}(v)$ follows a normal distribution conditioning on the VBM value $Z_{ilg}(v)$; and for each voxel v included in the analysis, the VBM value $Z_{ilg}(v)$ follows a normal distribution. The proposed model has the following hierarchical structure:

$$\begin{aligned} \text{For any } v \in \mathcal{G}, [X_{ilg}(v) | Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^{xz}(v), \delta_{lg}^{xz}\right\}, \\ \text{for any } v \in \mathcal{W}, [Y_{ilg}(v) | Z_{ilg}(v), D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^{yz}(v), \delta_{lg}^{yz}\right\}, \\ \text{for any } v \in \mathcal{B}, [Z_{ilg}(v) | D_i, \mathbf{W}_i, \bullet] &\sim N\left\{\mu_{lg}^z(v), \delta_{lg}^z\right\}, \end{aligned}$$

where

$$\begin{aligned} \mu_{lg}^{xz}(v) &= \sum_{k=0,1} [c_{klg}^{xz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i \boldsymbol{\gamma}_{klg}^x(v) + \beta_{klg}^x(v) \\ &\quad + \alpha_{ilg}^x + \eta_{kg}^x] I(D_i = k), \\ \mu_{lg}^{yz}(v) &= \sum_{k=0,1} [c_{klg}^{yz}(v)(Z_{ilg}(v) - \bar{Z}_{lg}(v)) + \mathbf{W}_i \boldsymbol{\gamma}_{klg}^y(v) + \beta_{klg}^y(v) \\ &\quad + \alpha_{ilg}^y + \eta_{kg}^y] I(D_i = k), \\ \mu_{lg}^z(v) &= \sum_{k=0,1} (\mathbf{W}_i \boldsymbol{\gamma}_{klg}^z(v) + \beta_{klg}^z(v) + \alpha_{ilg}^z + \eta_{kg}^z) I(D_i = k). \end{aligned}$$

We assume that the probability of disease status $P(D_i = k_i)$ is a constant, and independent of all the parameters. Also, we assume conditional independence among voxel measures of the same modality within the same subregion. The mean structure of the model is composed of several parameters, conditional on disease status. $c_{klg}(v)$ is the slope term for centered VBM values; $\boldsymbol{\gamma}_{klg}(v) = (\gamma_{klg1}(v), \dots, \gamma_{klgQ}(v))'$ is the parameter vector for covariates; $\beta_{klg}(v)$, α_{ilg} , and η_{kg} are the voxel-level intercept term, subregion level random effect, and region level intercept term, respectively. Each imaging modality is assumed to have the same subregion-level variance δ_{lg} for both subject groups.

The prior beliefs about the parameters included in the likelihood function are expressed in the second or lower levels of the model.

We also assume that

$$\begin{aligned} c_{klg}^{xz}(v) &\sim N(\zeta_{klg}^{xz}, \omega_{klg}^{xz}), & \zeta_{klg}^{xz} &\sim N(a_\zeta, b_\zeta), & \omega_{klg}^{xz} &\sim \text{InvG}(a_\omega, b_\omega), \\ c_{klg}^{yz}(v) &\sim N(\zeta_{klg}^{yz}, \omega_{klg}^{yz}), & \zeta_{klg}^{yz} &\sim N(a_\zeta, b_\zeta), & \omega_{klg}^{yz} &\sim \text{InvG}(a_\omega, b_\omega), \\ \gamma_{klgq}^m(v) &\sim N(0, s_{klg}^m), & s_{klg}^m &\sim \text{InvG}(a_s, b_s), \\ \beta_{klg}^m(v) &\sim N(\{\beta_{klg}^m, \lambda_{klg}^m\}), & \lambda_{klg}^m &\sim \text{InvG}(a_\lambda, b_\lambda), \end{aligned}$$

The slope $c_{klg}(v)$ follows a normal distribution, whose mean and variance are drawn from noninformative hyperpriors. Each covariate parameter $\gamma_{klgq}(v)$ is assumed to arise from a normal mean-zero distribution with variance s_{klg} , which has a noninformative hyperprior distribution. Parameters $\beta_{klg}(v)$ that fall within the same subregion are assumed to follow normal distributions with common mean β_{klg} , and variance λ_{klg} . We assume a noninformative distribution for λ_{klg} , and as described

in detail below, we use a spatial prior for β_{klg} to incorporate spatial correlations between subregions. η_k follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ_k whose off-diagonal elements capture spatial dependence between AAL regions. Spatial associations between voxels within each subregion are introduced by the individualized random effect term α_{ilg} , which follows a mean-zero normal distribution with variance τ_{lg} , thus assuming the same spatial correlations between voxels in the same subregion.

We assume a CAR model for β_{klg}^m as follows:

$$[\beta_{klg}^m | \{\beta_{kl'g}^m\}_{l' \neq l, \bullet}] \sim N \left\{ \frac{\rho_g^m}{n_{lg}} \sum_{l' \in N_g(l)} \beta_{kl'g}^m, \frac{\phi_g^m}{n_{lg}} \right\},$$

$$\rho_g^m \sim U(\{0, 0.05, 0.1, \dots, 0.8, 0.81, \dots, 0.9, 0.91, \dots, 0.99\}),$$

$$\phi_g^m \sim \text{InvG}(a_\phi, b_\phi),$$

$$\alpha_{ilg}^m \sim N(0, \tau_{lg}^m),$$

$$\tau_{lg}^m \sim \text{InvG}(a_\tau, b_\tau),$$

$$\eta_k^m = (\eta_{k1}^m, \dots, \eta_{kG}^m)' \sim N(0, \Sigma_k^m), \quad \Sigma_k^m \sim \text{InvW}(\Lambda, \nu),$$

$$\delta_{lg}^{xz} \sim \text{InvG}(a_\delta, b_\delta), \quad \delta_{lg}^{yz} \sim \text{InvG}(a_\delta, b_\delta),$$

$$\delta_{lg}^z \sim \text{InvG}(a_\delta, b_\delta), \quad \text{where } m \in \{x, y, z\}.$$

By assuming a subregion level CAR model, we capture the spatial dependence between subregions within each AAL region. In the model, ρ_g represents the overall degree of spatial dependence in region g and $\frac{\phi_g}{L_g}$ is the conditional variance of β_{klg} . The neighborhood of subregion $l \in g$, is defined as all the other subregions in AAL region g . The spatial neighborhood effect ρ_g is assumed to follow a discrete uniform distribution (Gelfand and Vounatsou, 2003). As we would like to identify the similarity of the neighboring subregions, we impose $0 \leq \rho_g < 1$. Specifically, equal mass is put on the following 36 values: 0, 0.05, 0.1, ..., 0.8, 0.81, 0.82, ..., 0.90, 0.91, 0.92, ..., 0.99, which includes a more refined set of values in the upper range of ρ_g since estimation of ρ_g for imaging data often tends toward large values.

For any disease status k , the covariance between the voxels within a same subregion l in region g is contributed by the variance from three components: β_{klg} , α_{ilg} , and η_{kg} ; the covariance between the voxels in two subregions l and l' , but the same AAL region g , comes from the covariance between β_{klg} and $\beta_{kl'g}$, and the variance of η_{kg} ; and the covariance between the voxels in two AAL regions g and g' is determined by the covariance of η_{kg} and $\eta_{kg'}$.

We perform estimation using Markov chain Monte Carlo (MCMC) implemented via Gibbs sampling. The full conditional posterior distributions are shown in Appendix 2 (Supplementary Material).

3.2. Prediction

3.2.1. Whole Brain Prediction

The objective of our model is to predict PD status, given imaging data and other covariates. To achieve this goal, we use

the posterior samples drawn from estimation to calculate the posterior predictive probability of disease status.

Let θ denote the parameter space, $\mathbf{B}_i = (\mathbf{X}_{ilg}, \mathbf{Y}_{ilg}, \mathbf{Z}_{ilg})$ denote the observed imaging data for subject i , and $\mathbf{A}_i = (\mathbf{B}_i, D_i)$ denote the combination of the imaging data and disease status. Suppose we have n training subjects, and we want to predict the disease status D_{n+1} for a new subject indexed by $n + 1$. The posterior predictive distribution for D_{n+1} is given by

$$\begin{aligned} P(D_{n+1} = k | \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) \\ &= \frac{P(D_{n+1} = k, \mathbf{B}_{n+1} | \{\mathbf{A}_i\}_{i=1}^n)}{\sum_{k'=0,1} P(D_{n+1} = k', \mathbf{B}_{n+1} | \{\mathbf{A}_i\}_{i=1}^n)} \\ &= \frac{P(D_{n+1} = k) \int_{\theta} P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta) P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta}{\sum_{k'=0,1} P(D_{n+1} = k') \int_{\theta} P(\mathbf{B}_{n+1} | D_{n+1} = k', \theta) P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta}, \end{aligned} \quad (1)$$

where

$$\begin{aligned} P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta) &= \prod_{v \in G} P(X_{n+1}(v) | Z_{n+1}(v), D_{n+1} = k, \theta) \\ &\quad P(Z_{n+1}(v) | D_{n+1} = k, \theta) \\ &\quad \prod_{v \in W} P(Y_{n+1}(v) | Z_{n+1}(v), D_{n+1} = k, \theta) \\ &\quad P(Z_{n+1}(v) | D_{n+1} = k, \theta), \end{aligned} \quad (2)$$

Suppose we draw T posterior samples, denoted $\theta^{(t)}$, from $P(\theta | \{\mathbf{A}_i\}_{i=1}^n)$, for $t = 1, \dots, T$. Letting $\pi_k^{(t)} = P(\mathbf{B}_{n+1} | D_{n+1} = k, \theta^{(t)})$, the posterior predictive probability can be expressed by

$$\hat{P}(D_{n+1} = k | \mathbf{B}_{n+1}, \{\mathbf{A}_i\}_{i=1}^n) = \frac{P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)}}{\sum_{k'=0,1} P(D_{n+1} = k') \sum_{t=1}^T \pi_k^{(t)}}. \quad (3)$$

Then ultimately the prediction of D_{n+1} is given by

$$\hat{D}_{n+1} = \arg \max_k \left(P(D_{n+1} = k) \sum_{t=1}^T \pi_k^{(t)} \right). \quad (4)$$

To evaluate the performance of our method, we calculate the prediction accuracy using LOOCV.

Applied directly, LOOCV is very computational expensive because it involves multiple posterior simulations with tens of thousands voxels included in the analysis. Therefore, we employ an importance sampling approach to reduce the computation for LOOCV of our model (Gelfand et al., 1992; Gelfand, 1996; Alqallaf and Gustafson, 2001; Vehtari and Lampinen, 2002). Specifically, the LOOCV predictive probabilities can be expressed by

$$P(D_i = k | \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = k) Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k') Q_{k'd_i}}, \quad (5)$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i | D_i = k, \theta)}{P(\mathbf{B}_i | D_i = d_i, \theta)} P(\theta | \{\mathbf{A}_i\}_{i=1}^n) d\theta, \quad (6)$$

and d_i is the observed disease status for subject i . Next, we provide the details of how Q_{kd_i} is derived. The posterior predictive probability can be written as follows:

$$\begin{aligned} P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) \\ = \int P(D_i = k \mid \mathbf{B}_i, \boldsymbol{\theta}) \frac{P(\boldsymbol{\theta} \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i})} \\ P(\boldsymbol{\theta} \mid \mathbf{B}_i, D_i = d_i, \mathbf{A}_{-i}) d\boldsymbol{\theta}. \end{aligned} \quad (7)$$

Therefore,

$$\frac{P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i})}{P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i})} = \frac{P(D_i = k)}{P(D_i = d_i)} Q_{kd_i}. \quad (8)$$

By using the fact that $\sum_{k=0,1} P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}) = 1$, we have

$$P(D_i = d_i \mid \mathbf{B}_i, \mathbf{A}_{-i}) = \frac{P(D_i = d_i)}{\sum_{k=0,1} P(D_i = k) Q_{kd_i}}, \quad (9)$$

thus leading to the above LOOCV predictive probability Equation (5). For $i = 1, \dots, n$ and $k = 0, 1$, compute

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^T \frac{P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})}. \quad (10)$$

The estimate of D_i is

$$\hat{D}_i = \arg \max_k (P(D_i = k) Q_{kd_i}). \quad (11)$$

Since there are only two possible values for D_i , we only need to calculate $P(\mathbf{B}_i \mid D_i = k, \boldsymbol{\theta}^{(t)})$ and $P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)})$, where $k \neq d_i$, for each subject i .

3.2.2. Voxel-Level Prediction

We also consider the use of imaging data $\mathbf{B}_i(v) = (X_{ilg}(v), Y_{ilg}(v), Z_{ilg}(v))$ for subject i at voxel v to predict the disease status D_i . Similar to Equation (5), the voxel-level LOOCV predictive probabilities can be expressed by

$$P(D_i = k \mid \mathbf{B}_i(v), \mathbf{A}_{-i}) = \frac{P(D_i = k) Q_{kd_i}}{\sum_{k'=0,1} P(D_i = k') Q_{k'd_i}}, \quad (12)$$

where

$$Q_{kd_i} = \int \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta}) P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta}) P(D_i = k')} P(\boldsymbol{\theta} \mid \{\mathbf{A}_i\}_{i=1}^n) d\boldsymbol{\theta}, \quad (13)$$

which is estimated by

$$\hat{Q}_{kd_i} = \frac{1}{T} \sum_{t=1}^T \frac{P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i(v) \mid D_i = k', \boldsymbol{\theta}^{(t)}) P(D_i = k')}{P(\mathbf{B}_i \mid D_i = d_i, \boldsymbol{\theta}^{(t)}) / \sum_{k'=0,1} P(\mathbf{B}_i \mid D_i = k', \boldsymbol{\theta}^{(t)}) P(D_i = k')}. \quad (14)$$

Then the estimate of D_i is

$$\hat{D}_i = \arg \max_k (P(D_i = k) Q_{kd_i}), \quad (15)$$

which is equivalent to

$$\hat{D}_i = \arg \max_k \left(P(D_i = k) \frac{1}{T} \sum_{t=1}^T P(\mathbf{B}_i(v) \mid D_i = k, \boldsymbol{\theta}^{(t)}) \right). \quad (16)$$

Q_{kd_i} is derived in the similar way as in the whole brain analysis. The derivation of Equation (14) is described in Appendix 3 (Supplementary Material).

The voxel-level prediction result can be used as a way to select the regions that are highly associated with PD if the prediction accuracy is high in these regions. An alternative approach to perform feature selection using our model is discussed in section 5.

4. RESULTS

4.1. Parkinson's Disease Data

We applied our proposed Bayesian spatial model to PD data, which has T1 and resting-state fMRI images available; therefore, our model reduces to one which includes two imaging modalities, VBM and fALFF, and only considers data in the gray matter. We generate predictions of PD based on multimodal imaging data aggregated across the whole brain, and we provide voxel-level predictions as well. By evaluating the prediction accuracy at each voxel, we are able to identify brain regions that are highly associated with Parkinson's disease as an alternative to performing feature selection.

In the estimation procedure, the hyperparameters for the prior distribution are set to provide vague information to ensure that the results are dominated by the information in the data. Specifically, all the hyperparameters in the inverse-gamma distribution are set to 10^{-3} (Spiegelhalter et al., 1994/2003), the normal prior for ζ_{klg} is assumed to have mean $a_\zeta = 0$ and variance $b_\zeta = 10^5$. In the inverse-Wishart distribution, the degrees of freedom ν should be greater than $G - 1$ to build a proper distribution, so we set $\nu = G$, which provides the least information based on our data. The scale matrix Λ is set as $10^{-3} \times \mathbf{I}_G$, where \mathbf{I}_G is a $G \times G$ identity matrix.

We perform a total of 10,000 MCMC iterations including 5,000 burn-in iterations, and store the results thinning by 10. Due

to the massive number of parameters in our model, we randomly check trace plots for parameters at the voxel-level, subregion-level, and region-level, respectively. We provide some examples in Appendix 4 (Supplementary Material).

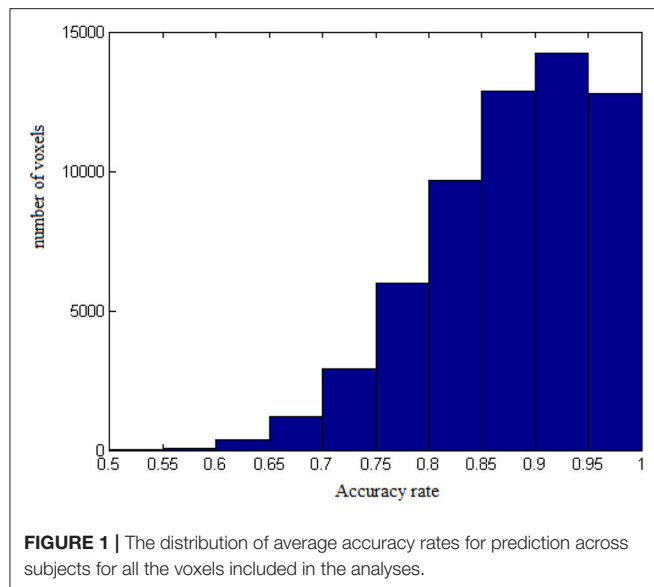


TABLE 1 | Summary of average accuracy rates for prediction across subjects.

Accuracy rate	[80%, 85%)	[85%, 90%)	[90%, 95%)	[95%, 100%)	100%
Number of voxels	5,993	9,663	12,878	14,236	12,764
(Percentage)	(9.97)	(16.07)	(21.42)	(23.68)	(21.23)

After estimating the model parameters, we perform a whole-brain and voxel-level prediction using posterior samples based on procedures described in section 3.2. Here, we have a total of 500 posterior samples after thinning. By assuming an equal probability for classification as a PD patient and a control subject, our model achieves 100% accuracy from the whole-brain prediction based on LOOCV.

The results from voxel-level prediction provide interesting information as well. The highest voxel-level accuracy rate is 100%, and the lowest is near 50%. **Figure 1** shows the distribution of the average accuracy rate across subjects for all the voxels included in the analysis. **Table 1** gives the number of voxels (percentage) achieving accuracy rates higher than 80%. Also, an average whole-brain prediction map based on the results from voxel-level prediction across subjects are presented in **Figure 2**.

To identify the regions which are predictive for disease status, we compute the average accuracy rates across voxels within a region, and **Table 2** shows the regions that have accuracy rate above 95%. **Table 2** also shows the percentage of voxels exceeding 90% accuracy rates for those regions. The right rectus gyrus, which is associated with cognitive impairment in PD patients, and is shown to have different gray matter density between PD and controls (Nagano-Saito et al., 2005), is identified in our analysis. The precentral gyrus, which is part of the primary motor cortex, is identified among the most accurate brain regions, and its performance is consistent with the involvement of this region in planning and initiating motor movements, which is critically impaired in patients with PD. We also find the

TABLE 2 | List of regions with above 95% average accuracy rate across voxels.

Region	Accuracy rate	Percentage of voxels with accuracy rate > 90%
Left postcentral	99.9%	42.6%
Right rectus	99.3%	52.2%
Left inferior parietal	99.3%	90.2%
Right superior medial frontal	99.0%	61.1%

bilateral caudate and the left putamen as regions with accurate predictions. The caudate and putamen, two regions comprising the dorsal striatum, exhibit marked pathologic changes from PD, linked to the loss of dopaminergic neurons in the substantia nigra which projects to striatal neurons in the caudate nucleus and putamen (Spencer et al., 1992). The right fusiform gyrus, which is believed to be related to impaired ability to correctly identify negative facial expressions (Geday et al., 2006), and the left inferior parietal lobule which is involved in the perception of emotions in facial stimuli, may play a role of differentiating healthy controls and PD patients as well. Other regions which are involved in face perception such as the right mid-temporal pole are also identified. The left postcentral gyrus, the left superior parietal lobule, and the right superior medial frontal gyrus also stand out since all of them are part of the sensory system. A region-level prediction map based on the average accuracy rates across voxels within a region is shown in **Figure 3**.

4.2. Simulation Studies

We conduct a simulation study to evaluate the performance of our proposed model. The purpose of this simulation study is to show that the MCMC generated samples from our model accurately target the true values and that the whole-brain prediction is accurate. In addition, we demonstrate that our model can distinguish regions that are predictive of disease status.

We assume that the imaging data are generated from the likelihood function of our model. We simulate data for 25 subjects from three AAL regions, the number of subregions within an AAL region has a mean and variance of 3, and the number of voxels within a subregion has a mean and variance of 50. We specify the true values for the parameters in the likelihood function, i.e., $c_{klg}(v)$, $\gamma_{klg}(v)$, $\beta_{klg}(v)$, α_{ilg} , η_{kg} , and δ_{lg} , which are the most relevant parameters for voxel-level inference and future prediction. In this way, we can compare our posterior estimations with specified true values. All the other parameters are updated from the posterior distributions. And the hyperparameters are set to be the same as in data from PD study. We select some subregions to be the ones that are associated with PD, and a region is classified into this category if it contains those selected subregions. We set different true values of parameters for disease and non-disease group if they are within the pre-specified regions and otherwise assume that the true values are the same for two groups. A total of 100 data sets are drawn in the simulation study. The programming is implemented in Matlab, and the computation is performed on a Linux cluster with 16 GB of RAM. Execution time is approximately 3–4 h for one data set.

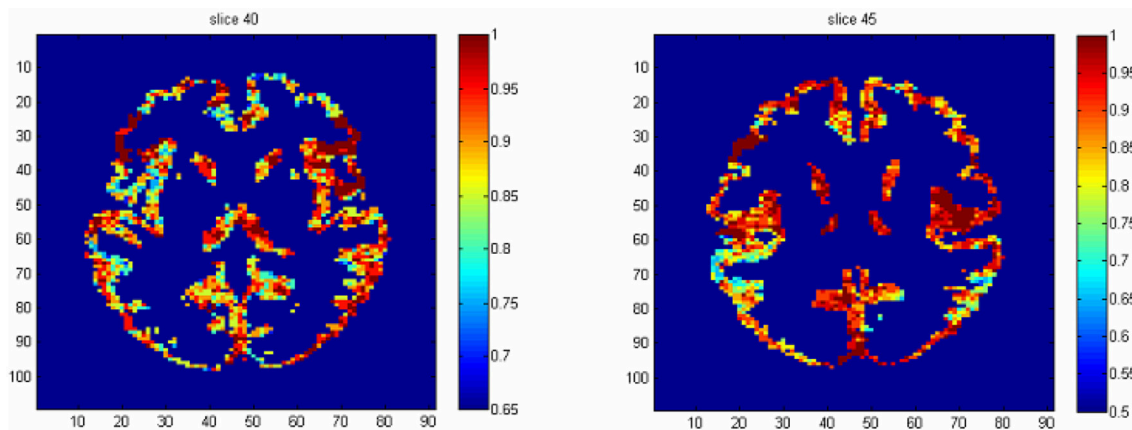


FIGURE 2 | The average prediction map based on the voxel-level prediction results across subjects.

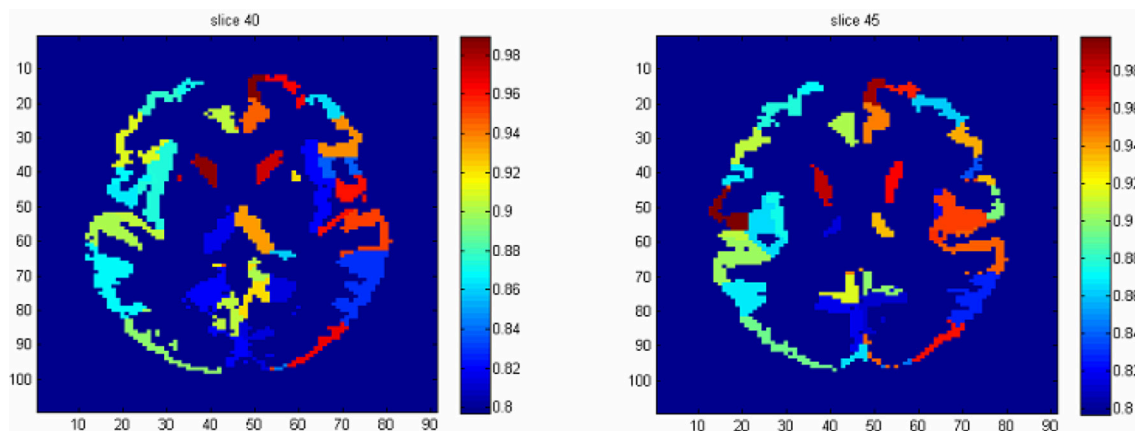


FIGURE 3 | The region-level prediction map based on the average accuracy rates across voxels within a region.

First, we evaluate the posterior estimates by comparing the posterior means to the true values. Instead of examining a total of five thousand parameters which have known true values separately, we calculate the mean structure and variance of the likelihood function from posterior samples and compare them to the truth since they are the most essential for inferences and predictions. The average bias (percentage change) in mean structure is 3.52×10^{-2} (0.54%), and in variance is 1.04×10^{-5} (1.04%). Secondly, we calculate the accuracy of whole-brain prediction. The LOOCV achieves 100% for the whole-brain prediction for all 100 simulated data sets. Thirdly, we identify the regions that are highly associated with disease status by evaluating the voxel-level accuracy rates for prediction. We compare the average accuracy for voxel-level prediction between the pre-specified regions and the others. Within the pre-specified regions, the average accuracy rate is 99.8%; for voxels which are in the other regions, the average accuracy rate is 71.7%. Here, we can see an improvement in prediction when voxels are from the pre-specified regions.

In comparison, we apply the elastic-net model to the simulated data as described above, and the LOOCV achieves an average of 86% accuracy rate for the whole-brain prediction.

In summary, our model accurately performs posterior estimation with small bias, provides accurate prediction of PD status using whole-brain imaging data, and correctly identifies the regions that are highly associated with disease.

5. DISCUSSION

We propose a Bayesian spatial model to predict PD using different modalities of imaging data, including fALFF, VBM, and FA in gray and white matter. Our framework performs voxel-level estimation for imaging data and conducts whole-brain and voxel-level prediction of disease status based on posterior predictive probabilities. Our model estimates both the mean and covariance structures of imaging data, predicting disease status using whole-brain imaging data, and identifying the regions which are highly associated with the disease based on voxel-level prediction results.

In our framework, we consider spatial correlations at voxel level, subregion level, and region level, and specify different correlation structures such as exchangeable, CAR, and unstructured correlation matrices for them. The rich hierarchical spatial correlation structures captured by our model extends previous spatial modeling frameworks by Bowman et al. (2008) and Derado et al. (2013). The intra-subregion correlation in our model is described by a single value within each subregion; the inter-subregion correlation is modeled by a CAR model which borrows information from the subregions within the same parent AAL region; the inter-region correlation is assumed to have a unstructured correlation matrix.

We derive the posterior predictive probability using whole brain data and data from a single voxel. Due to the complexity of computation, we adopt an importance sampling strategy to conduct LOOCV. The importance sampling techniques estimate the LOOCV error rate based only on one-model fitting using all samples and produces very accurate estimate on the LOOCV error rate. We evaluate the accuracy rate of the whole-brain prediction and identify the regions that are predictive for disease based on the results from voxel-level accuracy rates. Our model accounts for spatial correlations embedded in the data; however, additional multiple testing strategies could be explored to account for potential dependence inherent in the data. Our model increases localization compared to some approaches by offering voxel-level predictions. While we incorporate information from multiple modalities, we are unable to dissociate the relative predictive accuracy generated by each modality.

One weakness of our method is computational time since we use a joint model that performs estimation at the voxel-level. However, by applying the importance sampling strategy, we only need to perform the posterior estimation once, and then the posterior predictive probabilities can be computed fairly efficiently.

Compared to the existing feature selection methods, e.g., LASSO or elastic-net, our model uses a different modeling strategy and different criteria for selection. LASSO and elastic-net model the probability of PD, while our method starts from modeling the imaging data. This distinction leads to an important advantage that we are able to estimate and borrow strength from the spatial correlations in the data, whereas highly correlated predictors often lead to poor performance of the LASSO and related methods. Also, we use posterior predictive probability as the criteria to select the features, which is the exact target of prediction problems; on the other hand, LASSO and elastic-net, from a Bayesian perspective, use posterior modes to perform feature selection. Our model also has interpretive advantages over SVM and GP models by identifying particular voxels, subregions, or regions that contribute significantly to accurate prediction. Compared to the methodology of scalar-on-image regression

(Goldsmith et al., 2014; Reiss et al., 2015; Kang et al., 2016; Wang et al., 2017), our method models the images as the response, which is a natural generative process, and then we predict the disease distribution given the imaging scans.

In summary, the advantages of our proposed Bayesian method are three fold. First, it is more straightforward to incorporate prior knowledge regarding brain function and structure, which is extremely useful to improve the prediction accuracy and to provide a better understanding of the etiology. Second, it yields estimates and inference from the full posterior distribution, e.g., rather than point estimates. In particular, it can provide measures of uncertainty of the predictions based on the posterior predictive distribution. In addition, the posterior computation based on the MCMC algorithm is more robust to complex imaging data, while the optimization algorithms for other frequentist prediction methods are more likely to be trapped at the local modes, which may reduce the prediction accuracy.

In our method, we select features based on the posterior predictive probability of each voxel; ideally, we would like to identify the voxels $v \in \mathcal{V}$ s.t.

$$P(D_i = k \mid \{\mathbf{B}_i(v)\}_{v \in \mathcal{V}}, \mathbf{A}_{-i}) = P(D_i = k \mid \mathbf{B}_i, \mathbf{A}_{-i}), \quad (17)$$

which could be a possible extension of our proposed approach.

For Parkinson's disease, our model may not immediately supplant current clinical standards to diagnose patients at or near the manifestation of motor symptoms. However, our model stands to provide insights into the useful information for the diagnosis of PD, underlying neurophysiological basis of the disease, potentially early pre-motor alterations, and effective strategies to design studies examining potential neuroprotective treatments with consideration of the cost and complexity as well as extensive validation and comparison to current standards.

AUTHOR CONTRIBUTIONS

WX: methodology, simulation, and writing. FB and JK: supervising and editing the manuscript.

FUNDING

This research was funded by a grant from the NINDS (U18 NS082143) at NIH as part of the Parkinson's Disease Biomarker Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00184/full#supplementary-material>

REFERENCES

- Alqallaf, F., and Gustafson, P. (2001). On cross-validation of Bayesian models. *Can. J. Stat.* 29, 333–340. doi: 10.2307/3316081
- Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry—the methods. *NeuroImage* 11, 805–821. doi: 10.1006/nimg.2000.0582

- Bowman, F. D., Caffo, B. S., Spear Bassett, S., and Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* 39, 146–156. doi: 10.1016/j.neuroimage.2007.08.012
- Bowman, F. D., Drake, D. F., and Huddleston, D. (2016). Multimodal imaging signatures of Parkinson's disease. *Front. Neurosci.* 10:131. doi: 10.3389/fnins.2016.00131

- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Derado, G., Bowman, F. D., and Zhang, L. (2013). Predicting brain activity using a Bayesian spatial model. *Stat. Methods Med. Res.* 22, 382–397. doi: 10.1177/0962280212448972
- Di, X., Chan, R. C. K., and Gong, Q. (2009). White matter reduction in patients with schizophrenia as revealed by voxel-based morphometry: an activation likelihood estimation meta-analysis. *Prog. Neuro-psychopharmacol. Biol. Psychiatry* 33, 1390–1394. doi: 10.1016/j.pnpbp.2009.08.020
- Geday, J., Ostergaard, K., and Gjedde, A. (2006). Stimulation of subthalamic nucleus inhibits emotional activation of fusiform gyrus. *NeuroImage* 33, 706–714. doi: 10.1016/j.neuroimage.2006.06.056
- Gelfand, A. E. (1996). “Inference and monitoring convergence,” in *Markov Chain Monte Carlo in Practice*, eds W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman & Hall), 131–144.
- Gelfand, A. E., Dey, D. K., and Chang, H. (1992). “Model determination using predictive distributions with implementation via sampling-based methods (with discussion),” in *Bayesian Statistics*, Vol. 4, eds J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford University Press), 147–167.
- Gelfand, A. E., and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4, 11–15. doi: 10.1093/biostatistics/4.1.11
- Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *J. Comput. Graph. Stat.* 23, 46–64. doi: 10.1080/10618600.2012.743437
- Guo, Y., Bowman, F. D., and Kilts, C. (2008). Predicting the brain response to treatment using a Bayesian hierarchical model with application to a study of schizophrenia. *Hum. Brain Mapp.* 29, 1092–1109. doi: 10.1002/hbm.20450
- Ham, S. L., and Kwak, N. (2012). “Boosted-PCA for binary classification problems,” in *IEEE International Symposium on Circuits and Systems (ISCAS)*, 1219–1222.
- Kang, J., Reich, B. J., and Staicu, A. (2016). Scalar-on-Image Regression via the Soft-Thresholded Gaussian Process. arXiv:1604.03192.
- Lee, E. Y., Sen, S., Eslinger, P. J., Wagner, D., Shaffer, M. L., Kong, L., et al. (2013). Early cortical gray matter loss and cognitive correlates in non-demented Parkinson’s patients. *Parkinsonism Relat. Disord.* 19, 1088–1093. doi: 10.1016/j.parkreldis.2013.07.018
- Li, J. P., Pan, P. L., Huang, R., and Shang, H. F. (2011). A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer’s disease. *Neurosci. Biobehav. Rev.* 36, 757–763. doi: 10.1016/j.neubiorev.2011.12.001
- Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., and Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage* 49, 2178–2189. doi: 10.1016/j.neuroimage.2009.10.072
- Nagano-Saito, A., Washimi, Y., Arahata, Y., Kachi, T., Lerch, J. P., Evans, A. C., et al. (2005). Cerebral atrophy and its relation to cognitive impairment in Parkinson disease. *Neurology* 64, 224–229. doi: 10.1212/01.WNL.0000149510.41793.50
- Reiss, P. T., Huo, L., Zhao, Y., Kelly, C., and Ogden, R. T. (2015). Wavelet-domain regression and predictive inference in psychiatric neuroimaging. *Ann. Appl. Stat.* 9:1076. doi: 10.1214/15-AOAS829
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23(Suppl. 1), S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Spencer, D. D., Robbins, R. J., Naftolin, F., Marek, K. L., Vollmer, T., Leranth, C., et al. (1992). Unilateral transplantation of human fetal mesencephalic tissue into the caudate nucleus of patients with Parkinson’s disease. *New Engl. J. Med.* 327, 1541–1548. doi: 10.1056/NEJM199211263272201
- Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., and Lunn, D. (1994/2003). *BUGS: Bayesian Inference Using Gibbs Sampling*. Cambridge: MRC Biostatistics Unit. Available online at: <http://www.mrc-bsu.cam.ac.uk/bugs/>
- Stonnington, C. M., Chu, C., Klöppel, S., Jack, C. R. Jr., Ashburner, J., Frackowiak, R. S. J., et al. (2010). Predicting clinical scores from magnetic resonance scans in Alzheimer’s disease. *NeuroImage* 51, 1405–1413. doi: 10.1016/j.neuroimage.2010.03.051
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B (Methodol.)* 58, 267–288.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Vaillancourt, D. E., Spraker, M. B., Prodoehl, J., Abraham, I., Corcos, D. M., Zhou, X. J., et al. (2009). High-resolution diffusion tensor imaging in the substantia nigra of de novo Parkinson’s disease. *Neurology* 72, 1378–1384. doi: 10.1212/01.wnl.0000340982.01727.6e
- Vehtari, A., and Lampinen, J. (2002). Bayesian model assessment and comparison using cross-validation predictive densities. *Neural Comput.* 14, 2439–2468. doi: 10.1162/08997660260293292
- Wang, X., Zhu, H., and Alzheimer’s disease Neuroimaging Initiative. (2017). Generalized scalar-on-image regression models via total variation. *J. Amer. Stat. Assoc.* 112, 1156–1168. doi: 10.1080/01621459.2016.1194846
- Zarei, M., Ibarretxe-Bilbao, N., Compta, Y., Hough, M., Junque, C., Bargallo, N., et al. (2013). Cortical thinning is associated with disease stages and dementia in Parkinson’s disease. *J. Neurol. Neurosurg. Psychiatry* 84, 875–881. doi: 10.1136/jnnp-2012-304126
- Zhang, L., Wang, M., Sterling, N., Lee, E., Eslinger, P., Wagner, D., et al. (2015). Cortical thinning and cognitive impairment in Parkinson’s disease without dementia. *IEEE/ACM Trans. Comput. Biol. Bioinformatics.* doi: 10.1109/TCBB.2015.2465951
- Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Xue, Bowman and Kang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Fully-Automated Subcortical and Ventricular Shape Generation Pipeline Preserving Smoothness and Anatomical Topology

Xiaoying Tang^{1,2,3,4*}, Yuan Luo^{1,4}, Zhibin Chen^{1,4}, Nianwei Huang², Hans J. Johnson⁵, Jane S. Paulsen⁵ and Michael I. Miller^{6,7,8}

¹ Sun Yat-sen University-Carnegie Mellon University Joint Institute of Engineering, Sun Yat-sen University, Guangzhou, China, ² Sun Yat-sen University-Carnegie Mellon University Shunde International Joint Research Institute, Shunde, China, ³ School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China, ⁴ Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, United States, ⁵ Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, United States, ⁶ Center for Imaging Science, Johns Hopkins University, Baltimore, MD, United States, ⁷ Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, United States, ⁸ Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Xi-Nian Zuo,
Institute of Psychology (CAS), China

Reviewed by:

Jie Shi,
Arizona State University, United States
Hongjian He,
Zhejiang University, China

*Correspondence:

Xiaoying Tang
tangxiaoy@mail.sysu.edu
tangxy@sustc.edu.cn

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 08 December 2016

Accepted: 25 April 2018

Published: 17 May 2018

Citation:

Tang X, Luo Y, Chen Z, Huang N, Johnson HJ, Paulsen JS and Miller MI (2018) A Fully-Automated Subcortical and Ventricular Shape Generation Pipeline Preserving Smoothness and Anatomical Topology. *Front. Neurosci.* 12:321. doi: 10.3389/fnins.2018.00321

In this paper, we present a fully-automated subcortical and ventricular shape generation pipeline that acts on structural magnetic resonance images (MRIs) of the human brain. Principally, the proposed pipeline consists of three steps: (1) automated structure segmentation using the diffeomorphic multi-atlas likelihood-fusion algorithm; (2) study-specific shape template creation based on the Delaunay triangulation; (3) deformation-based shape filtering using the large deformation diffeomorphic metric mapping for surfaces. The proposed pipeline is shown to provide high accuracy, sufficient smoothness, and accurate anatomical topology. Two datasets focused upon Huntington's disease (HD) were used for evaluating the performance of the proposed pipeline. The first of these contains a total of 16 MRI scans, each with a gold standard available, on which the proposed pipeline's outputs were observed to be highly accurate and smooth when compared with the gold standard. Visual examinations and outlier analyses on the second dataset, which contains a total of 1,445 MRI scans, revealed 100% success rates for the putamen, the thalamus, the globus pallidus, the amygdala, and the lateral ventricle in both hemispheres and rates no smaller than 97% for the bilateral hippocampus and caudate. Another independent dataset, consisting of 15 atlas images and 20 testing images, was also used to quantitatively evaluate the proposed pipeline, with high accuracy having been obtained. In short, the proposed pipeline is herein demonstrated to be effective, both quantitatively and qualitatively, using a large collection of MRI scans.

Keywords: subcortical structures, lateral ventricle, shape, surface filtering, large deformation diffeomorphic metric mapping, surface triangularization

INTRODUCTION

Analyzing the shape of subcortical and ventricular structures subjected to brain disorders is an area of ever growing importance, especially in the fields of neurodegenerative diseases such as Alzheimer's disease (Qiu et al., 2009b; Wang et al., 2011; Shi et al., 2013, 2015; Tang et al., 2014, 2015b; Miller et al., 2015), Huntington's disease (HD) (van den Bogaard et al., 2011; Younes et al., 2014; Faria et al., 2016), and Parkinson's disease (Sterling et al., 2013; Nemmi et al., 2015) as well as various neurodevelopmental disorders (Knickmeyer et al., 2008; Rimol et al., 2010; Seymour et al., 2017). The anatomical shapes of the structures of interest in those cases are usually represented using a mesh that can be created from the corresponding structural volumetric segmentation. In more detail, generating a segmentation-based shape representation of a specific structure of interest (such as the left hippocampus) consists of two steps: (1) segmenting that structure of interest from a structural magnetic resonance image (MRI), resulting in a 3D volumetric segmentation; (2) converting that volumetric segmentation into a smooth surface representing the structural segmentation's boundary (Levine et al., 2012).

The fully automated segmentation of subcortical and ventricular structures, based on structural MRIs, is a well-established field of research, with a variety of highly accurate algorithms having already been developed (Barra and Boire, 2001; Khan et al., 2008; Powell et al., 2008; Patenaude et al., 2011; Chakravarty et al., 2013; Tang et al., 2015c). As for the generation of surfaces, image-based meshing is typically employed, especially when creating computer models for computational fluid dynamics and finite element analysis (Young et al., 2008; Chen et al., 2013; Chernikov et al., 2013; Foteinos and Chrisochoides, 2013; Zhang, 2013). More recently, segmentation based meshing has also been applied to the medical imaging field, see Zhang (2013) for a general introduction. One of the most representative meshing techniques is the marching cubes algorithm, which has been incorporated into a number of commercial and non-commercial software packages. The marching cubes algorithm takes a 3D segmentation image as its input and outputs surface data in the form of a triangulated mesh, represented using vertices and faces.

Combining what we have just outlined leads to an "automated volume segmentation + marching cubes based surface generation" pipeline for subcortical and ventricular structures. Such a procedure may well be vulnerable to noise induced by inaccurate segmentations, resulting in disconnected regions or holes within the surface (Qiu and Miller, 2008). In addition, it is plausible that the marching cubes algorithm is liable to miss thin subregions of a structure of interest such as the thin "bridge" connecting the inferior horn and the main body of the lateral ventricle (Qiu and Miller, 2008). In other words, the resulting surface may not have the correct anatomical topology. Furthermore, even for a structure of interest with a highly accurate segmentation and an "easy" topology (a relatively simple shape), it is likely that the marching cubes algorithm will not deliver surfaces of a sufficient smoothness. Indeed, it is a most challenging task to extract the structure of interest's surface

with high accuracy, correct anatomical topology, and sufficient smoothness in the same instance. To ensure a high degree of accuracy in the surface, a precise volumetric segmentation and a high fidelity in the surface with respect to the corresponding volumetric segmentation is required. Naturally, to ensure a correct anatomical topology, a surface generation approach that is devised around the notion of preserving the anatomical topology of the structure of interest is needed. Meanwhile, the classic filtering and smoothing approaches may not be sufficient to ensure the required smoothness without sacrificing the fidelity to the corresponding volumetric segmentation.

Alternatives to the aforementioned combination are certainly possible and there are numerous existing pipelines that can generate smooth subcortical structural shapes directly from MRIs. In contrast to a binary segmentation procedure for shape generation, those pipelines generally employ shape modeling for their segmentation purposes (Heimann and Meinzer, 2009; Patenaude et al., 2011). In other words, the structural shapes were not created from the binary segmentation, but directly from the dense MR images. The main limitation of these shape-modeling based approaches is the lack of flexibility in relation to individual components; one may desire the ability to utilize a more accurate segmentation algorithm or a more sophisticated meshing algorithm.

It is in the context of all of the above that we propose a fully-automated subcortical and ventricular shape generation pipeline which satisfies the demand for accuracy (both topological and otherwise) and smoothness in four steps: (1) automatically segment the subcortical and ventricular structures of interest using the raw structural MRI data acquired from a scanner; (2) create a study-specific template shape with the correct anatomical topology and sufficient surface smoothness; (3) create a triangulated mesh from each binary segmentation obtained in step (1) using the marching cubes algorithm; (4) filter and smooth the surfaces generated in step (3) in a deformation based manner.

To perform the initial segmentation, we employ a fully-automated segmentation pipeline, the diffeomorphic multi-atlas likelihood fusion (MALF) algorithm (Tang et al., 2013), the accuracy of which in segmenting subcortical and ventricular structures has been validated on a variety of MRI datasets (Tang et al., 2015c). Instead of applying the marching cubes algorithm directly, to generate a corresponding triangulated mesh from the segmentation of MALF with the desired properties, we rely on deformation based shape generation in the setting of large deformation diffeomorphic metric mapping (LDDMM) for surfaces (Vaillant and Glaunès, 2005). Given a pre-defined triangulated surface of a specific structure of interest, LDDMM is capable of preserving the topology and smoothness of that surface when registering it to a target surface. In other words, if we register a template surface with the correct anatomical topology and a high degree of smoothness to a target surface using LDDMM, the deformed surface is guaranteed to inherit that topology and smoothness from the template while being as similar as possible to the target surface. This is essentially due to the properties of diffeomorphic transformations and the capability of LDDMM to deliver the accurate diffeomorphisms needed for surface registration (Vaillant and Glaunès, 2005).

In this paper, we will first detail each of the above steps in the proposed pipeline. We then proceed to evaluate the proposed pipeline quantitatively and qualitatively using three MRI datasets. There are 16 structural MRIs in the first dataset, for each of which we manually segmented the subcortical and ventricular structures, with a view to quantitatively evaluating the performance of the proposed pipeline by comparison with the gold standard. Within the second dataset, there are a total of 1,445 structural MRIs, on which we qualitatively examine the surfaces delivered by the proposed pipeline. For the third dataset, there are 15 atlas structural MRIs and 20 testing structural MRIs, with the structures of interest being the subcortical structures that have been manually delineated. We also compared our results with those from a well-established pipeline that outputs smooth subcortical surfaces directly from dense MRIs, namely the FSL-FIRST pipeline (Patenaude et al., 2011). Three aspects were examined; the accuracy based on quantitative evaluation, the anatomy topology based on visual examination, and the smoothness based on quantitative assessment.

MATERIALS AND METHODS

PREDICT-HD

The first two datasets that feature in this work are both part of the PREDICT-HD study (<https://www.predict-hd.net/>) where all enrolled subjects were at risk of HD and had previously undergone elective predictive genetic testing. Subjects labeled as premanifest HD (pre-HD) are those who were found to be “gene expanded,” possessing a cytosine-adenine-guanine (CAG) ≥ 36 but not exhibiting the motor criteria consistent with a diagnosis of HD (The Huntington’s Disease Collaborative Research Group, 1993). A control group was defined as subjects who were deemed “non-gene expanded,” possessing a CAG ≤ 30 . Participants of PREDICT-HD were recruited from 32 sites across the United States, Canada, Europe, and Australia and underwent longitudinal study visits consisting of a neurological motor examination, cognitive assessment, brain MRI, psychiatric and functional assessment, and blood testing for genetic and biochemical analyses. Informed written consent was obtained from all subjects before participating in this study.

Subjects with pre-HD were further divided into three subgroups (“low-HD,” “mid-HD,” and “high-HD”) based on their CAP scores, a function of their CAG repeat length and current age given by $CAP = (\text{age at study entry}) \times (CAG - 33.66)$ (Zhang et al., 2011). The three subgroups are defined according to $CAP < 290$ (the low-HD group), $290 \leq CAP \leq 368$ (the mid-HD group), and $CAP > 368$ (the high-HD group).

Subjects

In the first dataset, there are a total of 16 subjects (3 males and 13 females, mean age = 42.1 ± 10.1 years), including 6 control subjects, 4 low-HD subjects, 3 mid-HD subjects, and 3 high-HD subjects. Only one scan of each subject was selected, resulting in a total of 16 MRI scans in the first dataset.

For the second dataset, there are a total of 169 control subjects, including 106 females (mean age at baseline = 48.3 ± 11.2 years) and 63 males (mean age at baseline = 48.6 ± 14.8 years). Within

the control group, 59 subjects had only 1 scan, 43 subjects had 2 scans, 27 subjects had 3 scans, 16 subjects had 4 scans, 15 subjects had 5 scans, 7 subjects had 6 scans, and 1 subject had 7 scans, resulting in a total of 414 MRI scans, with the average interval between two consecutive scans being 1.1 years. Within the low-HD group, there are a total of 113 subjects, including 85 females (mean age at baseline = 33.1 ± 9.1 years) and 28 males (mean age at baseline = 35.7 ± 10.8 years). In the low-HD group, 52 subjects had only 1 scan, 35 subjects had 2 scans, 12 subjects had 3 scans, 8 subjects had 4 scans, 3 subjects had 5 scans, 2 subjects had 6 scans, and 1 subject had 8 scans, resulting in a total of 225 MRI scans, with the average interval between two consecutive scans being 0.8 years. Within the mid-HD group, there are a total of 141 subjects, including 98 females (mean age at baseline = 42.1 ± 10.2 years) and 43 males (mean age at baseline = 42.4 ± 11.2 years). In the mid-HD group, 62 subjects had only 1 scan, 36 subjects had 2 scans, 14 subjects had 3 scans, 17 subjects had 4 scans, 5 subjects had 5 scans, 6 subjects had 6 scans, and 1 subject had 7 scans, resulting in a total of 312 MRI scans, with the average interval between two consecutive scans being 0.8 years. Within the high-HD group, there are a total of 227 subjects, including 136 females (mean age at baseline = 49.3 ± 10.9 years) and 91 males (mean age at baseline = 50.0 ± 11.1 years). In the high-HD group, 99 subjects had only 1 scan, 68 subjects had 2 scans, 26 subjects had 3 scans, 17 subjects had 4 scans, 8 subjects had 5 scans, 8 subjects had 6 scans, and 1 subject had 8 scans, resulting in a total of 477 MRI scans, with the average interval between two consecutive scans being 0.9 years. There are another 4 females (mean age at baseline = 44.6 ± 9.9 years) that were not identified as belonging to any group. Among those 4 subjects, 3 had been scanned once while the remainder had been scanned twice, resulting in a total of 5 MRI scans. There are another 12 MRI scans for which we could not identify their demographic and clinical information. However, given that the goal of this paper is to evaluate a surface generation pipeline rather than to compare groups of different clinical states, we retained all of the 1,445 scans from the second dataset for pipeline validation. A summary of this dataset is tabulated in **Table 1**.

High resolution anatomical MR images of the first two datasets were used in this study. Given that the PREDICT-HD study was both multi-centered and longitudinal in nature, the image acquisition procedures were heterogeneous, including multiple vendors (GE, Phillips, and Siemens), different field strengths (1.5 Tesla and 3 Tesla), and more than 20 different MR acquisition protocols (due to issues with transmission and receiver hardware). Detailed scanning information for each of the 1,445 MR scans can be found in the Supplementary Material 1.

The third dataset used in this study includes 35 brain MRI scans from the OASIS project. The manual segmentations of these images were produced by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>) using the brainCOLOR labeling protocol. The data were applied in the 2012 MICCAI Multi-Atlas Labeling Challenge and are publicly accessible (https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_Page). In the challenge, 15 subjects were used as atlases and the remaining 20 images were used for testing. For this dataset, our structures of interest are the 12 subcortical regions.

TABLE 1 | A summary of the second dataset, consisting of 1,445 MRI scans.

	Control		Low-HD	
	Male (no = 63)	Female (no = 106)	Male (no = 28)	Female (no = 85)
Baseline age	48.6 ± 14.8 years	48.3 ± 11.2 years	35.7 ± 10.8 years	33.1 ± 9.1 years
No. of scans = 1	59		52	
No. of scans = 2	43		35	
no. of scans = 3	27		12	
No. of scans = 4	16		8	
No. of scans = 5	15		3	
No. of scans = 6	7		2	
No. of scans = 7	1		0	
No. of scans = 8	0		1	
Average inter-scan interval	1.1 years		0.8 years	
	Mid-HD		High-HD	
	Male (no = 43)	Female (no = 98)	Male (no = 91)	Female (no = 136)
Baseline age	42.4 ± 11.2 years	42.1 ± 10.2 years	50.0 ± 11.1 years	49.3 ± 10.9 years
No. of scans = 1	62		99	
No. of scans = 2	36		68	
No. of scans = 3	14		26	
No. of scans = 4	17		17	
No. of scans = 5	5		8	
No. of scans = 6	6		8	
No. of scans = 7	1		0	
No. of scans = 8	0		1	
Average inter-scan interval	0.8 years		0.9 years	

Automated Structure Segmentation

As shown in **Figure 1** (the work flow of the proposed pipeline), one can view this pipeline as having two major components; automated structure segmentation and surface filtering. The subcortical and ventricular structures, in both hemispheres, were extracted from each T1-weighted image using a fully-automated structure segmentation pipeline (Tang et al., 2015c) itself consisting of two steps, skull-stripping and brain structure segmentation. The underlying theoretical basis of this approach is multi-atlas likelihood-fusion (MALF) in the framework of a random deformable template model (Tang et al., 2013). This segmentation pipeline has been tested and validated on a number of datasets with relevance to various brain structures, particularly the subcortical and ventricular structures (Liang et al., 2015; Tang et al., 2015a).

In this study, the 16 T1-weighted images of the first dataset served as the atlases used in MALF to perform the automated structure segmentation for the first and the second datasets. Each structure of interest, such as the left hippocampus, was manually delineated in all 16 atlases by a team of neuroanatomists at Johns Hopkins University with more than 15 years' experience in manually tracing subcortical structures. Various sets of subcortical and ventricular atlases, used in our other studies, were all created by the same team and have proven their reliability (Tang et al., 2013, 2015c, 2016; Seymour et al., 2017). Intra- and inter-rater reliability of manual delineations by this team have

been quantified in earlier studies; intra-class correlation (ICC) statistics revealed high rates of intra- and interrater reliability (intra-rater ICC ranges between 0.96 and 0.98; inter-rater ICC ranges between 0.9 and 0.93) (Qiu et al., 2009a).

To evaluate the proposed pipeline's handling of the first dataset, we adopted a leave-one-out strategy; one atlas image was treated as the to-be-segmented image while the remainder served as the atlas set used in segmenting that excluded image. When evaluating the second dataset, we continued to use these 16 atlases for segmentation via MALF. For the third dataset, the 15 atlas images were used to segment the subcortical structures in each of the 20 testing images.

Surface Generation

With the binary segmentation of the structures of interest completed using the structure segmentation procedure discussed above, we proceeded to create a triangulated mesh contouring the boundary of the segmentation using the marching cubes algorithm. The marching cubes algorithm yields triangulated surfaces with a high fidelity to the segmentation. Thus, when the segmentation is lacking accuracy, the marching cubes algorithm will be incapable of correcting the mistakes incurred during the segmentation step. In addition, the resulting surface may well be insufficiently smooth for our purposes. To overcome these limitations, one potential approach is to register a template surface to a target surface (the raw structure surface created from

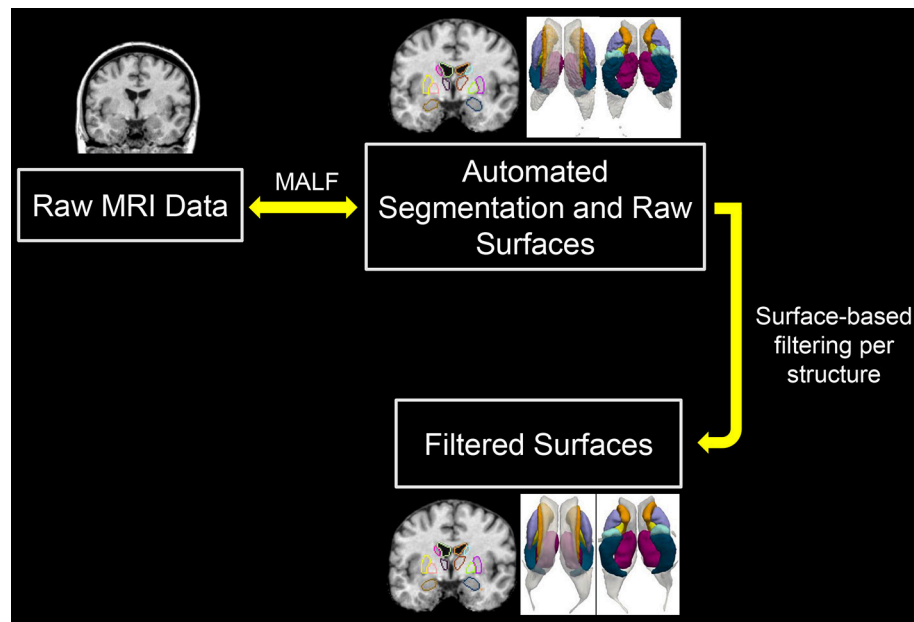


FIGURE 1 | Demonstration of the workflow of the proposed pipeline. MALF, multi-atlas likelihood fusion.

the marching cubes algorithm). The template surface is supposed to have correct anatomical topology and sufficient smoothness. The deformed template surfaces are therefore expected to have geometric characteristics identical to those of the target surfaces while possessing the topology and connectivity of the template surface.

In our pipeline, the template surface came from one of the 16 subjects in the first dataset. The 14 structures of interest for the selected subject were manually delineated with care taken to ensure both segmentation accuracy and boundary smoothness during the manual delineation. That specific subject was chosen based on three considerations: (1) the area of the subject's surface should be close to the mean area across all 16 surfaces from the manual segmentations; (2) the geometry and topology of the subject's surface should be correct based on visual examination; (3) the selected surface should be sufficiently smooth quantitatively and qualitatively.

In creating the template surface, instead of using the marching cubes algorithm, we adopted the Delaunay algorithm for triangulation (Lee and Schachter, 1980; Shewchuk, 2002) to guarantee further smoothness. We have noticed, however, that the Delaunay algorithm is much less stable than that of the marching cubes, even though it yields smoother results. This is our rationale for using marching cubes for the triangulation of the raw structure surfaces rather than the Delaunay algorithm.

With the template surface and target surfaces for each structure of interest created, we performed a rigid alignment of the surfaces and then proceeded to the LDDMM surface registration (Vaillant and Glaunès, 2005). Specifically, the template surface was rigidly aligned (rotation and translation) to the target surface, with the optimal rigid transformation between the vertex sets of the two surfaces obtained by minimizing a score

that combines registration and soft assignment. After that, the LDDMM surface registration was performed from the rigidly aligned template surface to the target surface. Details on the “rigid + LDDMM” surface registration pipeline can be found in our previous work (Tang et al., 2014). After obtaining all of the rigid and diffeomorphic transformations between the template surface and the target surfaces, we applied these transformations in turn to the template surface, generating a deformed template surface for each structure of interest in each subject MRI. This deformed template surface is the result of our proposed pipeline, a smooth surface of a subcortical and ventricular structure of interest in an individual MRI scan.

Evaluation Criteria

As we have the gold standard—manual segmentations—at our disposal for the first and the third datasets, we quantitatively computed the accuracy and reliability of the proposed pipeline through the use of the following evaluation metrics:

- Dice similarity coefficient (DSC)

$$DSC(A, B) = 2 \frac{V(A \cap B)}{V(A) + V(B)} \quad (1)$$

where $V(A)$ and $V(B)$ are the volumetric measurements of segmented images A and B . For example, A may represent the binary segmentation of the left hippocampus from manual delineation while B represents the corresponding automated segmentation from MALF.

- Absolute volume difference (AVD)

$$AVD(A, B) = \frac{|V(A) - V(B)|}{(V(A) + V(B)) / 2} \quad (2)$$

where $V(A)$ and $V(B)$ are again the volumetric measurements of segmented images A and B .

- Correlation coefficient

For the third quantitative comparison metric, we employed the Pearson product-moment correlation coefficient (PCC) between the volumetric measurements of the two segmentations in comparison, for example those of the manual segmentation and the MALF-derived automated segmentation.

In addition to evaluating the segmentation accuracy using the first and the third datasets, we also assessed the smoothness of the resulting surfaces quantitatively and qualitatively (through visual examination by several raters) using all three datasets. The smoothness of a surface was quantified using the following metric:

- Geometric Laplacian (GL)

$$GL(v) = v - \frac{\sum_{i \in n(v)} l_i^{-1} v_i}{\sum_{i \in n(v)} l_i^{-1}} \quad (3)$$

where $n(v)$ is the index set of the vertices v_i which are themselves the direct neighbors of v , and l_i is the Euclidean distance from v to v_i . $GL(v)$ represents a kind of measure of roughness: the higher it is, the rougher is the surface around v . The GL of a surface is computed as the sum of the norm of all vertex-wise GL vectors, namely $GL = \sum_v \|GL(v)\|_2$.

Group Comparisons

In our first experiment, we compared results from the proposed pipeline, in terms of both volumetric segmentations and triangulated surfaces, with those before filtering (obtained from MALF) using all three datasets. Their results were also compared to the gold standard of the first and the third datasets. In the first experiment, our structures of interest included all the 14 subcortical and lateral ventricle structures for the first two datasets and the 12 subcortical structures for the third dataset. In the second experiment, we performed a comparison with a state-of-the-art pipeline, FSL-FIRST, that outputs volumetric segmentations as well as smooth triangulated surfaces of subcortical structures as well. This experiment was conducted on the first dataset and analyzed the 12 subcortical structures only, as FSL-FIRST does not output lateral ventricle results. Student's t -tests were employed to evaluate the significance of a group difference in all settings.

RESULTS

The First Experiment

In Tables 2–4, we respectively detail the mean and standard deviations of the DSCs, the AVDs, and the PCCs for each of the 14 structures of interest of the first dataset when calculated under the three possible comparisons; the raw automated segmentations from MALF vs. the manual segmentations, the raw automated segmentations from MALF vs. the filtered automated segmentations, as well as the filtered automated segmentations vs. the manual ones. The corresponding results on

TABLE 2 | The average Dice overlap coefficients between every pairing of the three sets of segmentation results (manual segmentation, raw automated segmentation, and filtered automated segmentation) over the 16 MRI scans of the first group for each of the 14 subcortical and ventricle structures.

	Manual vs. Raw Auto	Raw Auto vs. Filtered Auto	Manual vs. Filtered Auto
Left caudate	0.914 ± 0.039	0.958 ± 0.008	0.913 ± 0.039
Right caudate	0.901 ± 0.028	0.957 ± 0.006	0.899 ± 0.027
Left pallidum	0.902 ± 0.023	0.957 ± 0.006	0.899 ± 0.024
Right pallidum	0.907 ± 0.018	0.957 ± 0.005	0.906 ± 0.022
Left putamen	0.928 ± 0.010	0.965 ± 0.005	0.928 ± 0.009
Right putamen	0.934 ± 0.012	0.966 ± 0.005	0.934 ± 0.009
Right thalamus	0.922 ± 0.011	0.969 ± 0.004	0.924 ± 0.011
Left thalamus	0.927 ± 0.009	0.970 ± 0.004	0.929 ± 0.008
Left amygdala	0.874 ± 0.017	0.943 ± 0.010	0.874 ± 0.020
Right amygdala	0.866 ± 0.025	0.946 ± 0.008	0.870 ± 0.025
Left hippocampus	0.917 ± 0.009	0.939 ± 0.006	0.909 ± 0.011
Right hippocampus	0.910 ± 0.013	0.943 ± 0.007	0.907 ± 0.014
Left ventricle	0.925 ± 0.023	0.891 ± 0.047	0.858 ± 0.058
Right ventricle	0.922 ± 0.027	0.902 ± 0.048	0.866 ± 0.059

TABLE 3 | The average absolute volume differences between every pairing of the three sets of segmentation results (manual segmentation, raw automated segmentation, and filtered automated segmentation) over the 16 MRI scans of the first group for each of the 14 subcortical and ventricle structures.

	Manual vs. Raw Auto	Raw Auto vs. Filtered Auto	Manual vs. Filtered Auto
Left caudate	0.076 ± 0.092	0.011 ± 0.005	0.078 ± 0.088
Right caudate	0.098 ± 0.080	0.010 ± 0.005	0.100 ± 0.079
Left pallidum	0.103 ± 0.075	0.016 ± 0.007	0.108 ± 0.079
Right pallidum	0.082 ± 0.061	0.018 ± 0.006	0.086 ± 0.071
Left putamen	0.037 ± 0.027	0.007 ± 0.005	0.035 ± 0.027
Right putamen	0.055 ± 0.026	0.008 ± 0.004	0.055 ± 0.023
Right thalamus	0.090 ± 0.038	0.002 ± 0.002	0.092 ± 0.039
Left thalamus	0.075 ± 0.035	0.003 ± 0.003	0.075 ± 0.036
Left amygdala	0.082 ± 0.041	0.018 ± 0.005	0.077 ± 0.045
Right amygdala	0.069 ± 0.072	0.015 ± 0.005	0.065 ± 0.072
Left hippocampus	0.064 ± 0.029	0.014 ± 0.006	0.076 ± 0.029
Right hippocampus	0.065 ± 0.031	0.014 ± 0.005	0.074 ± 0.033
Left ventricle	0.078 ± 0.053	0.006 ± 0.005	0.080 ± 0.055
Right ventricle	0.082 ± 0.057	0.010 ± 0.006	0.088 ± 0.061

the 12 subcortical structures of the third dataset are demonstrated in the Supplementary Material 2 (Table S1). Please note, the filtered automated segmentations were generated from the smoothly deformed surfaces via nearest neighbor assignment. As shown in the first column of each of the three tables, the raw automated segmentations obtained from MALF are highly accurate when compared to the gold standard. This illustrates the accuracy of the first step of our surface generation pipeline. For the second step, generating a smoothed version of the raw surface, we achieved a high fidelity, as is demonstrated in the second column in each of the three tables. Comparing the final

results, the filtered surface based segmentations, with the gold standard, the accuracy is again high (the third column of each of the three tables) and indeed similar to that of the raw accuracy.

Results on comparing the smoothness of the surfaces of those three approaches for the first and the third datasets are respectively demonstrated in **Table 5** and the Supplementary Material 2 (Table S2). Clearly, for each of the structures of interest, surfaces from the proposed pipeline are significantly smoother ($p < 10^{-10}$) than not only the raw automated results from MALF but also the manual results. In **Figure 2**, we present comparison results for the three methods (manual, raw automated, and filtered automated), in terms of segmentations that are superimposed on the structural MR image (for better visualization) and the corresponding surfaces, for one

representative subject. Evidently, the proposed method is capable of capturing thin regions of a structure of interest, such as in the lateral ventricle, and thus preserving the structure's anatomical topology. Furthermore, even when compared with the gold standard surfaces created from the marching cubes algorithm, the surfaces delivered by the proposed pipeline are much smoother.

In **Figure 3**, we illustrate the smoothness comparison results of both datasets before and after deformation based filtering, from which a significant increase in smoothness was observed for each structure in both datasets. In addition to smoothness, the segmentation accuracy of the second dataset were also visually examined independently by three experienced raters. We found that on the bilateral putamen, globus pallidus, amygdala, thalamus, and lateral ventricle, the proposed pipeline delivered sufficiently well-generated surfaces for all 1,445 scans. In other words, the failure rate for any of those 5 structures in both hemispheres is 0%. For the other subcortical structures the number of surfaces found to be flawed were as follows: 19 out of 1,445 surfaces of the left caudate (failure rate being 1.31%), 15 out of 1,445 surfaces of the right caudate (failure rate being 1.04%), 7 out of 1,445 surfaces of the left hippocampus (failure rate being 0.48%), and 33 out of 1,445 surfaces of the right hippocampus (failure rate being 2.28%). We also note that the 19 left caudate surfaces with flaws were generated from the scans of 16 subjects while the 15 right caudate surfaces came from 9 subjects, the 7 left hippocampus surfaces came from 4 subjects, and the 33 right hippocampus surfaces came from 14 subjects. Such observations suggest that a failure for the proposed pipeline is more likely to recur in longitudinal scans of the same subject than on the dataset as a whole. In **Figures 4, 5**, we present the outputs in representative failure cases for the caudate (both left and right) and the hippocampus (both left and right) respectively.

In addition to qualitative assessment, we also conducted outlier analysis based on each surface's GL value. To be specific, outliers were defined as those whose GL values were outside the range $[Q_1 - 1.5(Q_3 - Q_1), Q_1 + 1.5(Q_3 - Q_1)]$, where Q_1 and Q_3 respectively denote the 25 percentile and the 75 percentile

TABLE 4 | The Pearson product-moment correlation coefficients between every pairing of the three sets of segmentation results (manual segmentation, raw automated segmentation, and filtered automated segmentation) over the 16 MRI scans of the first group for each of the 14 subcortical and ventricle structures.

	Manual vs. Raw Auto	Raw Auto vs. Filtered Auto	Manual vs. Filtered Auto
Left caudate	0.821	1.000	0.827
Right caudate	0.821	1.000	0.829
Left pallidum	0.737	0.999	0.741
Right pallidum	0.835	0.999	0.835
Left putamen	0.980	1.000	0.980
Right putamen	0.965	1.000	0.966
Right thalamus	0.899	1.000	0.898
Left thalamus	0.906	0.999	0.908
Left amygdala	0.633	0.999	0.641
Right amygdala	0.716	0.999	0.711
Left hippocampus	0.795	0.998	0.809
Right hippocampus	0.804	0.999	0.819
Left ventricle	0.996	1.000	0.996
Right ventricle	0.990	1.000	0.990

TABLE 5 | Smoothness quantification, as measured by the Geometric Laplacian, of the four sets of surface results [manual, raw automated (MALF), filtered automated (proposed), and FSL-FIRST] over the 16 MRI scans of the first group for the 12 subcortical structures.

	Manual	MALF	Proposed	FSL-FIRST
Left caudate	562.371 ± 70.610	619.822 ± 61.332	192.202 ± 8.967	187.216 ± 9.973
Right caudate	553.225 ± 81.988	599.452 ± 58.889	208.535 ± 7.896	193.338 ± 12.298
Left pallidum	249.926 ± 27.753	250.861 ± 30.475	87.136 ± 3.919	73.241 ± 4.416
Right pallidum	246.651 ± 20.956	254.202 ± 28.085	78.655 ± 3.506	67.427 ± 4.129
Left putamen	509.561 ± 61.449	560.402 ± 54.812	176.756 ± 6.163	123.851 ± 8.138
Right putamen	515.220 ± 57.479	568.109 ± 48.324	181.189 ± 6.611	126.411 ± 8.534
Right thalamus	634.443 ± 45.398	724.972 ± 65.240	197.801 ± 5.349	133.038 ± 5.835
Left thalamus	626.319 ± 46.549	708.264 ± 56.974	194.041 ± 7.095	132.420 ± 5.196
Left amygdala	232.511 ± 22.514	244.034 ± 18.729	84.181 ± 1.899	195.033 ± 13.753
Right amygdala	218.056 ± 22.433	229.562 ± 21.813	77.585 ± 2.445	167.832 ± 10.335
Left hippocampus	581.167 ± 27.421	607.406 ± 37.712	183.013 ± 4.637	77.366 ± 6.519
Right hippocampus	595.273 ± 32.929	619.702 ± 45.359	181.883 ± 4.959	80.187 ± 5.993

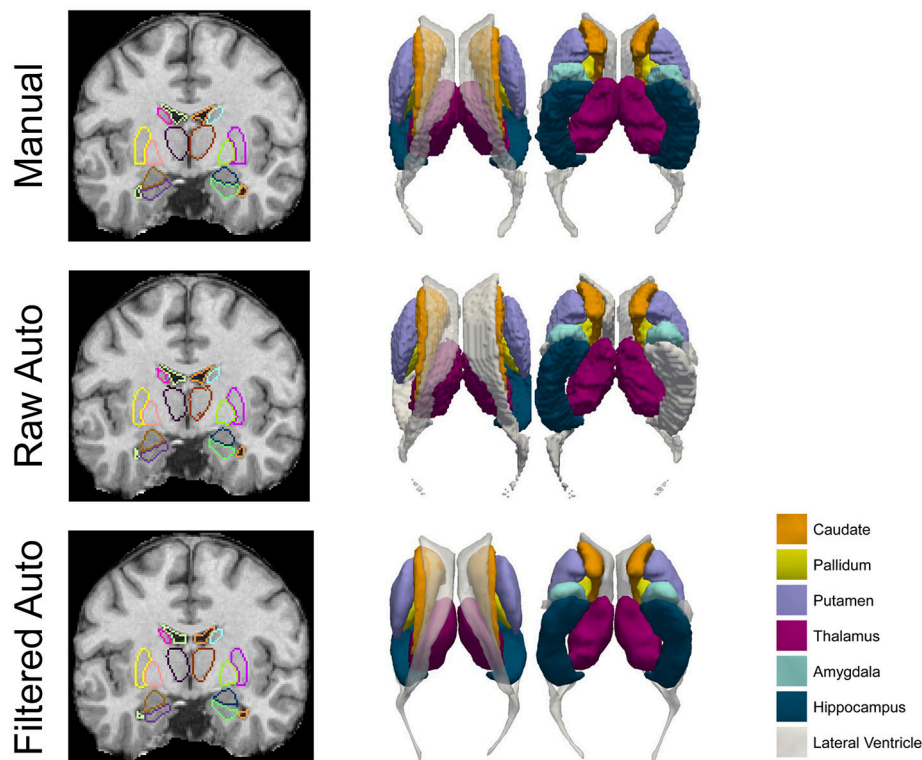


FIGURE 2 | A comparison of the manual results, the raw automated segmentation results, and the filtered automated results, for the 7 subcortical and ventricular structures (both left and right) of one representative subject. Both segmentations (left column) and the corresponding triangulated surfaces (right column) are presented.

of all structure-specific GL values. From this outlier analysis, we detected 15 outliers for the left caudate, 9 outliers for the right caudate, 6 outliers for the left hippocampus, and 26 outliers for the right hippocampus. These numbers agree well with our qualitative assessment results.

The Second Experiment

The mean values and standard deviations of GL for the 12 subcortical surfaces, delivered by FSL-FIRST, are also listed in **Table 5**, from which we observed a similar level of smoothness as results from the proposed pipeline, both being significantly smoother than those from the gold standard and MALF. Comparing between the proposed pipeline and FSL-FIRST, the bilateral amygdalar surfaces from the proposed pipeline are much smoother than those from FSL-FIRST whereas an opposite pattern was observed for the bilateral hippocampal surfaces. Overall, those two methods have similar performance in terms of surface smoothness. With regards to the segmentation accuracy, as quantified by the DSCs (**Table 6**), the AVDs (**Table 7**), and PCCs (**Table 8**), the proposed pipeline significantly outperformed FSL-FIRST.

DISCUSSION

In this paper, we have developed a fully-automated shape generation pipeline for subcortical and ventricular structures of

the human brain which preserves smoothness and anatomical topology in the surfaces. The performance of the pipeline has been validated on three datasets, both quantitatively and qualitatively. We found that, without sacrificing the accuracy, the resultant surfaces have high smoothness and correct anatomical topology. Based on visual examinations and outlier analyses on a large number of surfaces (1,445 in total for each structure), the pipeline has a very low rate of failure; to be specific, the failure rate is 0% for the putamen, the globus pallidus, the amygdala, the thalamus, and the lateral ventricle in both hemispheres, 1.31% for the left caudate, 1.04% for the right caudate, 0.48% for the left hippocampus, and 2.28% for the right hippocampus. As is exemplified in **Figures 4, 5**, the main cause of failure for the caudate and the hippocampus is segmentation inaccuracy incurred in the MALF based automated segmentation. Those two structures are both adjacent to the cerebrospinal fluid and it has been found that this makes them more susceptible to inaccuracy (Tang et al., 2013). Even for those two structures, the failure rates on the first and the third datasets are 0% while those on the second dataset are < 3% and we consider such results to be a strong indicator of the pipeline's capacity for high performance.

There are three main components in the pipeline: automated structure segmentation; creation of study-specific template shapes; and LDDMM-based shape filtering. For automated structure segmentation, we utilized a well-developed algorithm of our own group's creation, the diffeomorphic multi-atlas

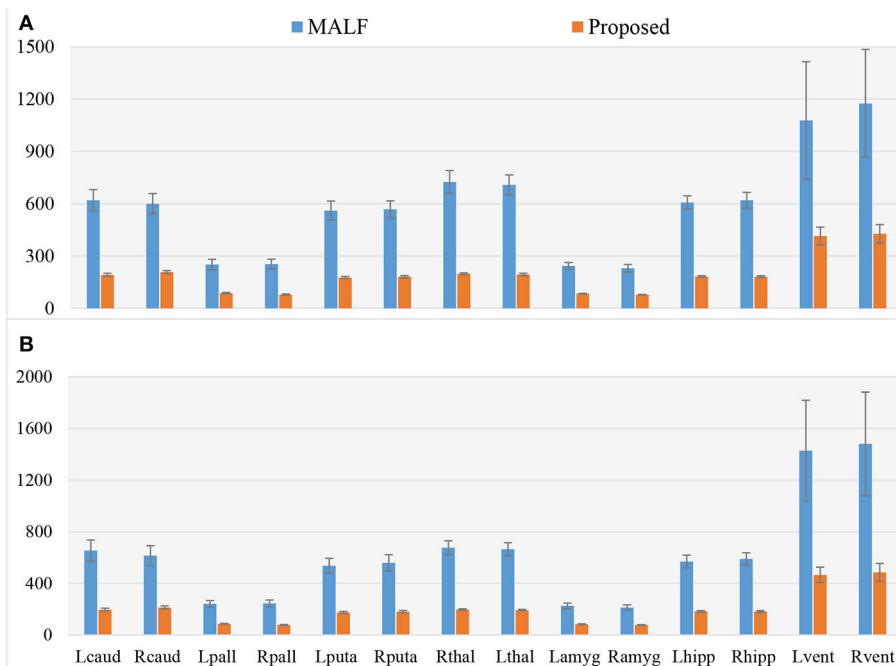


FIGURE 3 | A comparison of the smoothness, as assessed by the Geometric Laplacian, of surfaces from MALF (the raw automated segmentation results) and the proposed method (the filtered automated results), for the 7 subcortical and ventricular structures (both left and right) for both datasets. Lcaud, left caudate; Rcaud, right caudate; Lpall, left pallidum; Rpall, right pallidum; Lputa, left putamen; Rputa, right putamen; Lthal, Left Thalamus; Rthal, right thalamus; Lamyg, left amygdala; Ramyg, right amygdala; Lhipp, left hippocampus; Rhipp, right hippocampus; Lvent, left ventricle; Rvent, right ventricle. **(A,B)** Respectively denote the results for the first and the second dataset.

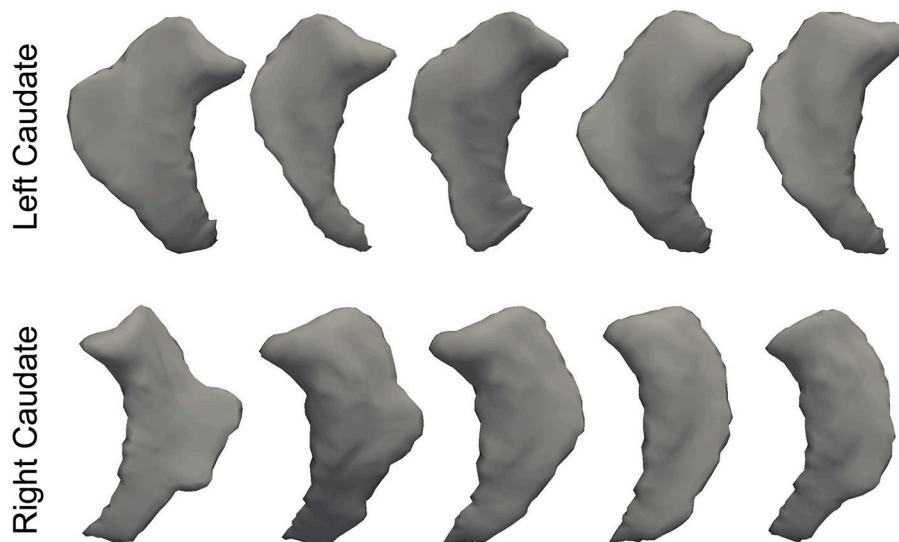


FIGURE 4 | Representative failure cases for the left caudate (**top**) and the right caudate (**bottom**) from the second dataset.

likelihood fusion. Using the first and the third datasets, which have the manual segmentations available, we again validated the performance of the MALF algorithm in terms of the automated segmentation of subcortical and ventricular structures. For this component, we can also use other automated structure

segmentation algorithms, as long as the accuracy is sufficient, such as FreeSurfer (Fischl et al., 2002) and FSL-FIRST (Patenaude et al., 2011). FreeSurfer based segmentations have also been used for surface generation in existing works (Qiu and Miller, 2008; Qiu et al., 2009b; Tang et al., 2014). In FSL-FIRST, the

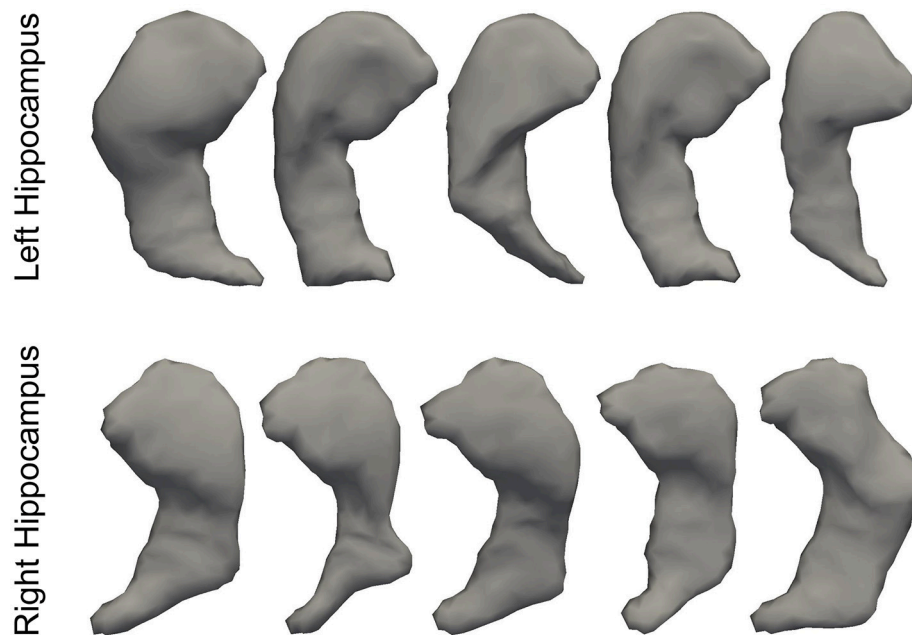


FIGURE 5 | Representative failure cases for the left hippocampus (**top**) and the right hippocampus (**bottom**) from the second dataset.

TABLE 6 | The average Dice overlap coefficients between the gold standard and segmentations from the proposed method as well as those between the gold standard and FSL-FIRST over the 16 MRI scans of the first group for the 12 subcortical structures alongside the corresponding p -values obtained from Student's t -tests.

	Proposed method	FSL-FIRST	p -value
Left caudate	0.913 ± 0.039	0.832 ± 0.025	$3.548\text{E-}06$
Right caudate	0.899 ± 0.027	0.834 ± 0.016	$2.727\text{E-}08$
Left pallidum	0.899 ± 0.024	0.818 ± 0.041	$1.658\text{E-}05$
Right pallidum	0.906 ± 0.022	0.797 ± 0.042	$7.068\text{E-}08$
Left putamen	0.928 ± 0.009	0.881 ± 0.023	$1.314\text{E-}06$
Right putamen	0.934 ± 0.009	0.882 ± 0.021	$1.286\text{E-}09$
Right thalamus	0.924 ± 0.011	0.898 ± 0.023	$2.849\text{E-}04$
Left thalamus	0.929 ± 0.008	0.902 ± 0.019	$2.397\text{E-}03$
Left amygdala	0.874 ± 0.020	0.779 ± 0.041	$3.057\text{E-}07$
Right amygdala	0.870 ± 0.025	0.776 ± 0.032	$2.818\text{E-}07$
Left hippocampus	0.909 ± 0.011	0.826 ± 0.024	$6.721\text{E-}09$
Right hippocampus	0.907 ± 0.014	0.833 ± 0.019	$9.484\text{E-}10$

TABLE 7 | The average absolute volume differences between the gold standard and segmentations from the proposed method as well as those between the gold standard and FSL-FIRST over the 16 MRI scans of the first group for the 12 subcortical structures alongside the corresponding p -values obtained from Student's t -tests.

	Proposed method	FSL-FIRST	p -value
Left caudate	0.078 ± 0.088	0.135 ± 0.042	$5.273\text{E-}02$
Right caudate	0.100 ± 0.079	0.099 ± 0.049	$9.674\text{E-}01$
Left pallidum	0.108 ± 0.079	0.179 ± 0.081	$4.673\text{E-}02$
Right pallidum	0.086 ± 0.071	0.244 ± 0.055	$1.359\text{E-}05$
Left putamen	0.035 ± 0.027	0.183 ± 0.059	$5.703\text{E-}09$
Right putamen	0.055 ± 0.023	0.147 ± 0.055	$1.153\text{E-}06$
Right thalamus	0.092 ± 0.039	0.083 ± 0.066	$7.772\text{E-}01$
Left thalamus	0.075 ± 0.036	0.061 ± 0.062	$1.786\text{E-}01$
Left amygdala	0.077 ± 0.045	0.132 ± 0.079	$4.229\text{E-}02$
Right amygdala	0.065 ± 0.072	0.205 ± 0.140	$1.486\text{E-}03$
Left hippocampus	0.076 ± 0.029	0.259 ± 0.087	$7.400\text{E-}07$
Right hippocampus	0.074 ± 0.033	0.188 ± 0.083	$2.403\text{E-}04$

segmentation of a subcortical structure of interest is actually obtained from its corresponding smooth surface. In other words, FSL-FIRST outputs both smooth surfaces and segmentations for subcortical structures. In that sense, it may be redundant to perform another round of surface generation based on segmentations from FSL-FIRST.

In this work, we did not compare the surface results from the proposed pipeline with those obtained from replacing our segmentation module with another one since that is essentially a comparison of various segmentation algorithms, which is not

the goal of this paper. With that being said, we did validate the segmentation accuracy of our pipeline using the gold standard of the first dataset, with the DSCs ranging between 0.87 and 0.93 (Table 2), the AVDs ranging between 0.04 and 0.1 (Table 3), and the PCCs ranging between 0.72 and 1 (Table 4), as well as the third dataset (see Table S1 in the Supplementary Material 2).

For the second step, the creation of study-specific template shapes, we applied the Delaunay algorithm (Lee and Schachter, 1980; Shewchuk, 2002) for triangulating a carefully-selected manual segmentation for each structure of interest. The reason

TABLE 8 | The Pearson product-moment correlation coefficients between the gold standard and segmentations from the proposed method as well as those between the gold standard and FSL-FIRST over the 16 MRI scans of the first group for the 12 subcortical structures alongside the corresponding *p*-values indicating the significance level of each correlation.

	Proposed method		FSL-FIRST	
	PCC	<i>p</i> -value	PCC	<i>p</i> -value
Left caudate	0.827	7.730E-05	0.953	1.129E-08
Right caudate	0.829	7.075E-05	0.948	2.528E-08
Left pallidum	0.741	1.031E-03	0.831	6.723E-05
Right pallidum	0.835	5.635E-05	0.921	4.022E-07
Left putamen	0.980	3.128E-11	0.946	3.024E-08
Right putamen	0.966	1.349E-09	0.965	1.475E-09
Right thalamus	0.908	1.160E-06	0.730	1.329E-03
Left thalamus	0.898	2.347E-06	0.706	2.237E-03
Left amygdala	0.641	7.484E-03	0.440	8.847E-02
Right amygdala	0.711	2.009E-03	0.097	7.217E-01
Left hippocampus	0.809	1.454E-04	0.568	2.179E-02
Right hippocampus	0.819	1.032E-04	0.576	1.962E-02

for using a manually created segmentation is 2-fold: firstly, a manual segmentation can guarantee correct anatomy and smoothness to some degree; secondly, we had previously generated the manual segmentations to serve as atlases in our automated structure segmentation phase, meaning no additional effort was required here. With that being said, we can also create a template shape based on an automated segmentation with sufficient accuracy, correct anatomy, and sufficient smoothness. The Delaunay algorithm is superior to the marching cubes algorithm in terms of smoothness of the resultant surfaces though it can fail in some cases, especially when the segmentation is flawed. Therefore, in this case, we were well-placed to generate the template shapes using the Delaunay algorithm since we could pay special attention to those surfaces. Meanwhile the marching cubes algorithm was better suited for the target segmentations.

In practice, there are two guiding rules in selecting the template surface: (1) the same definitions should be used in the automated segmentations of the target MRIs as in the segmentation of the template surface. For example, in this work, all automated segmentations of the first two datasets were obtained by using the atlases of the 16 subjects while the template surface was also obtained from this 16-subject pool. It may be inappropriate to use a template surface from a MALF-based segmentation definition to smooth an automated segmentation from FSL-FIRST; (2) It is better to select a template surface from the same study sample. In other words, it may be inappropriate to use a template surface from our HD study to smooth an automated segmentation from another study.

For the third step, LDDMM-based shape filtering, the key idea is to use a diffeomorphic transformation that can accurately deform the template shape to be very close to the target one while preserving the smoothness and topology of the

template shape. LDDMM-surface is a validated algorithm that has been shown to yield sophisticated diffeomorphisms that can accurately register a pair of surfaces (Vaillant and Glaunès, 2005). According to our experiments on all three datasets, the deformed results, based on LDDMM-surface matching, are very close to the raw data (the target segmentations for which we aim to create their corresponding surfaces) while preserving the topology and smoothness of the template shapes. The high fidelity of the resulting surfaces to the target segmentations is somewhat of a double-edge sword; on the one hand, it guarantees high accuracy while on the other, it causes sensitivity to the inaccuracy induced in the segmentation process. In other words, when the segmentations are noisy (like those from the second dataset that the pipeline failed on), the resulting surfaces will inherit the noise (inaccuracy) of the segmentations from MALF. A potential solution is to utilize a much more robust variant of the LDDMM-surface matching, such as the one proposed by Tward and colleagues (Tward et al., 2016). Investigation of more advanced surface matching algorithms that are capable of maintaining a high fidelity to the segmentation while being robust to noisy subregions of the segmentations will be one of our future efforts. Furthermore, there are wholly separate registration approaches that can be applied to deforming surfaces, such as the 14 methods compared in (Klein et al., 2009). We did not compare here the surfaces generated by using different surface deformation approaches as that goes beyond the scope of this paper; to formulate the proposed pipeline.

This work was strongly motivated by the ongoing search for simpler, more effective, and more flexible pipelines capable of generating subcortical and ventricular surfaces with high smoothness and correct anatomy. According to our comparison results with another popular pipeline that directly outputs binary segmentations and smooth triangulated surfaces, namely FSL-FIRST, the surface results from the proposed pipeline have a similar degree of smoothness as those from FSL-FIRST, whereas the proposed pipeline's segmentation accuracy is significantly higher than FSL-FIRST for almost each of the 12 subcortical structures, which agrees with our previous findings (Tang et al., 2015c). This again may suggest a superiority of the proposed pipeline, although we must be aware of the potential unfairness given that a specific structure's definition may differ significantly for atlases used in MALF and those in FSL-FIRST. Compared with existing pipelines, the main contribution of this work, aside from the pipeline performance, is to have provided a general framework that can be easily adopted or modified according to one's own purpose; for example, to replace MALF with another segmentation algorithm that one favors more or to choose a template surface that one considers to be more suitable for a specific study.

One potential limitation of the proposed pipeline is that it is difficult to be sure that no subtle disease-related features were lost during this surface generation process. A way to partially address this question is to compare the disease-related features (via group comparison to a control group) obtained by using a set of surfaces created

manually (to ensure accuracy) and those obtained by using a set of surfaces created from the proposed pipeline. However, given the lack of such a set of manually created surfaces involving both control and disease subjects, it is not possible to conduct such an experiment at this moment. We anticipate that as one of our future endeavors.

The statistical shape analysis of subcortical and ventricular structures of the human brain has become a topic of most considerable interest in contemporary research (Styner et al., 2003; Qiu and Miller, 2008; Qiu et al., 2010). We are confident that the proposed pipeline will further the development of this research field, especially in the investigations of HD.

AUTHOR CONTRIBUTIONS

XT and MM: Contributed to the design of the entire pipeline; YL, ZC, and NH: Contributed to the analysis and evaluation experiments; HJ and JP: Contributed to the data acquisition; XT:

Wrote the paper. All authors revised the manuscript critically for important intellectual content.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (NSFC 81501546) and the SYSU-CMU Shunde International Joint Research Institute Start-up Grant (20150306). We would like to thank Huilin Yang at Carnegie Mellon University for her efforts on the implementation of surface smoothness quantification. We would also like to thank the Johns Hopkins University team (Timothy Brown, Deana Crocetti, and Katarina A. Ament) on the manual delineation of the 16 atlases.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00321/full#supplementary-material>

REFERENCES

- Barra, V., and Boire, J. (2001). Automatic segmentation of subcortical brain structures in MR images using information fusion. *IEEE Trans. Med. Imaging* 20, 549–558. doi: 10.1109/42.932740
- Chakravarty, M. M., Steadman, P., van Eede, M. C., Calcott, R. D., Gu, V., Shaw, P., et al. (2013). Performing label-fusion-based segmentation using multiple automatically generated templates. *Hum. Brain Mapp.* 34, 2635–2654. doi: 10.1002/hbm.22092
- Chen, M., Tu, B., and Lu, B. (2013). “Surface triangular mesh and volume tetrahedral mesh generations for biomolecular modeling,” in *Image-Based Geometric Modeling and Mesh Generation*, ed Y. Zhang (Dordrecht: Springer), 85–106.
- Chernikov, A. N., Foteinos, P. A., Liu, Y., Audette, M., Enquobahrie, A., and Chrisochoides, N. P. (2013). “Tetrahedral image-to-mesh conversion approaches for surgery simulation and navigation,” in *Image-Based Geometric Modeling and Mesh Generation*, ed Y. Zhang (Dordrecht: Springer), 69–84.
- Faria, A. V., Ratnanather, J. T., Tward, D. J., Lee, D. S., van den Noort, F., Wu, D., et al. (2016). Linking white matter and deep gray matter alterations in premanifest huntington disease. *NeuroImage Clin.* 11, 450–460. doi: 10.1016/j.nicl.2016.02.014
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/S0896-6273(02)00569-X
- Foteinos, P. A., and Chrisochoides, N. P. (2013). “High-quality multi-tissue mesh generation for finite element analysis,” in *Image-Based Geometric Modeling and Mesh Generation*, ed Y. Zhang (Dordrecht: Springer), 159–169.
- Heimann, T., and Meinzer, H. (2009). Statistical shape models for 3D medical image segmentation: a review. *Med. Image Anal.* 13, 543–563. doi: 10.1016/j.media.2009.05.004
- Khan, A. R., Wang, L., and Beg, M. F. (2008). FreeSurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping. *Neuroimage* 41, 735–746. doi: 10.1016/j.neuroimage.2008.03.024
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46, 786–802. doi: 10.1016/j.neuroimage.2008.12.037
- Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., et al. (2008). A structural MRI study of human brain development from birth to 2 Years. *J. Neurosci.* 28, 12176–12182. doi: 10.1523/JNEUROSCI.3479-08.2008
- Lee, D., and Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *Int. J. Comput. Info. Sci.* 9, 219–242. doi: 10.1007/BF00977785
- Levine, J. A., Paulsen, R. R., and Zhang, Y. (2012). Mesh processing in medical-image analysis—a tutorial. *IEEE Comput. Graph. Appl.* 32, 22–28. doi: 10.1109/MCG.2012.91
- Liang, Z., He, X., Ceritoglu, C., Tang, X., Li, Y., Kuttan, K. S., et al. (2015). Evaluation of cross-protocol stability of a fully automated brain multi-atlas parcellation tool. *PLoS ONE* 10:e0133533. doi: 10.1371/journal.pone.0133533
- Miller, M. I., Younes, L., Ratnanather, J. T., Brown, T., Trinh, H., Lee, D. S., et al. (2015). Amygdalar atrophy in symptomatic Alzheimer’s disease based on diffeomorphicometry: the BIOCARD cohort. *Neurobiol. Aging* 36, S3–S10. doi: 10.1016/j.neurobiolaging.2014.06.032
- Nemmi, F., Sabatini, U., Rascol, O., and Péran, P. (2015). Parkinson’s disease and local atrophy in subcortical nuclei: insight from shape analysis. *Neurobiol. Aging* 36, 424–433. doi: 10.1016/j.neurobiolaging.2014.07.010
- Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 56, 907–922. doi: 10.1016/j.neuroimage.2011.02.046
- Powell, S., Magnotta, V. A., Johnson, H., Jammalamadaka, V. K., Pierson, R., and Andreasen, N. C. (2008). Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage* 39, 238–247. doi: 10.1016/j.neuroimage.2007.05.063
- Qiu, A., Brown, T., Fischl, B., Ma, J., and Miller, M. I. (2010). Atlas generation for subcortical and ventricular structures with its applications in shape analysis. *IEEE Trans. Image Process.* 19, 1539–1547. doi: 10.1109/TIP.2010.2042099
- Qiu, A., Crocetti, D., Adler, M., Mahone, E. M., Denckla, M. B., Miller, M. I., et al. (2009a). Basal ganglia volume and shape in children with attention deficit hyperactivity disorder. *Am. J. Psychiatr.* 166, 74–82. doi: 10.1176/appi.ajp.2008.08030426
- Qiu, A., Fennema-Notestine, C., Dale, A. M., Miller, M. I., and Alzheimer’s Disease Neuroimaging Initiative. (2009b). Regional shape abnormalities in mild cognitive impairment and Alzheimer’s disease. *Neuroimage* 45, 656–661. doi: 10.1016/j.neuroimage.2009.01.013
- Qiu, A., and Miller, M. I. (2008). Multi-structure network shape analysis via normal surface momentum maps. *Neuroimage* 42, 1430–1438. doi: 10.1016/j.neuroimage.2008.04.257

- The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
- Rimol, L. M., Hartberg, C. B., Nesvåg, R., Fennema-Notestine, C., Hagler, D. J. Jr., Pung, C. J., et al. (2010). Cortical thickness and subcortical volumes in Schizophrenia and bipolar disorder. *Biol. Psychiatr.* 68, 41–50. doi: 10.1016/j.biopsych.2010.03.036
- Seymour, K. E., Tang, X., Crocetti, D., Mostofsky, S. H., Miller, M. I., and Rosch, K. S. (2017). Anomalous subcortical morphology in boys, but not girls, with ADHD compared to typically developing controls and correlates with emotion dysregulation. *Psychiatr. Res. Neuroimag.* 261, 20–28. doi: 10.1016/j.psychres.2017.01.002
- Shewchuk, J. R. (2002). Delaunay refinement algorithms for triangular mesh generation. *Comput. Geometr.* 22, 21–74. doi: 10.1016/S0925-7721(01)00047-5
- Shi, J., Stonnington, C. M., Thompson, P. M., Chen, K., Gutman, B., Reschke, C., et al. (2015). Studying ventricular abnormalities in mild cognitive impairment with hyperbolic Ricci flow and tensor-based morphometry. *Neuroimage* 104, 1–20. doi: 10.1016/j.neuroimage.2014.09.062
- Shi, J., Thompson, P. M., Gutman, B., Wang, Y., and Alzheimer's Disease Neuroimaging Initiative. (2013). Surface fluid registration of conformal representation: application to detect disease burden and genetic influence on hippocampus. *Neuroimage* 78, 111–134. doi: 10.1016/j.neuroimage.2013.04.018
- Sterling, N. W., Du, G., Lewis, M. M., Dimaio, C., Kong, L., Eslinger, P. J., et al. (2013). Striatal shape in parkinson's disease. *Neurobiol. Aging* 34, 2510–2516. doi: 10.1016/j.neurobiolaging.2013.05.017
- Styner, M., Gerig, G., Lieberman, J., Jones, D., and Weinberger, D. (2003). Statistical shape analysis of neuroanatomical structures based on medial models. *Med. Image Anal.* 7, 207–220. doi: 10.1016/S1361-8415(02)00110-X
- Tang, X., Kuttan, K., Ceritoglu, C., Mori, S., and Miller, M. I. (2015a). Simultaneous skull-stripping and lateral ventricle segmentation via fast multi-atlas likelihood fusion. *Proc. SPIE* 9413, 94130W–94130W-7. doi: 10.1117/12.2080830
- Tang, X., Holland, D., Dale, A. M., Younes, L., and Miller, M. I. (2015b). The diffeomorphometry of regional shape change rates and its relevance to cognitive deterioration in mild cognitive impairment and Alzheimer's disease. *Hum. Brain Mapp.* 36, 2093–2117. doi: 10.1002/hbm.22758
- Tang, X., Crocetti, D., Kuttan, K., Ceritoglu, C., Albert, M. S., Mori, S., et al. (2015c). Segmentation of brain magnetic resonance images based on multi-atlas likelihood fusion: testing using data with a broad range of anatomical and photometric profiles. *Front. Neurosci.* 9:61. doi: 10.3389/fnins.2015.00061
- Tang, X., Holland, D., Dale, A. M., Younes, L., and Miller, M. I. (2014). Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and Alzheimer's disease: detecting, quantifying, and predicting. *Hum. Brain Mapp.* 35, 3701–3725. doi: 10.1002/hbm.22431
- Tang, X., Oishi, K., Faria, A. V., Hillis, A. E., Albert, M. S., Mori, S., et al. (2013). Bayesian Parameter Estimation and Segmentation in the Multi-Atlas Random Orbit Model. *PLoS ONE* 8:e65591. doi: 10.1371/journal.pone.0065591
- Tang, X., Varma, V. R., Miller, M. I., and Carlson, M. C. (2016). Education is associated with sub-regions of the hippocampus and the amygdala vulnerable to neuropathologies of Alzheimer's disease. *Brain Struct. Funct.* 222, 1469–1479. doi: 10.1007/s00429-016-1287-9
- Tward, D., Miller, M., Trouve, A., and Younes, L. (2016). Parametric surface diffeomorphometry for low dimensional embeddings of dense segmentations and imagery. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1195–1208. doi: 10.1109/TPAMI.2016.2578317
- Vaillant, M., and Glaunès, J. (2005). "Surface matching via currents," in *Biennial International Conference on Information Processing in Medical Imaging*, Vol. 3565 (Glenwood Springs, CO), 381–392.
- van den Bogaard, S. J. A., Dumas, E. M., Ferrarini, L., Milles, J., van Buchem, M. A., van der Grond, J., et al. (2011). Shape analysis of subcortical nuclei in Huntington's disease, global versus local atrophy—Results from the TRACK-HD study. *J. Neurol. Sci.* 307, 60–68. doi: 10.1016/j.jns.2011.05.015
- Wang, Y., Song, Y., Rajagopalan, P., An, T., Liu, K., Chou, Y. Y., et al. (2011). Surface-based TBM boosts power to detect disease effects on the brain: an N=804 ADNI study. *Neuroimage* 56, 1993–2010. doi: 10.1016/j.neuroimage.2011.03.040
- Younes, L., Ratnanather, J. T., Brown, T., Aylward, E., Nopoulos, P., Johnson, H., et al. (2014). Regionally selective atrophy of subcortical structures in prodromal HD as revealed by statistical shape analysis. *Hum. Brain Mapp.* 35, 792–809. doi: 10.1002/hbm.22214
- Young, P. G., Beresford-West, T. B. H., Coward, S. R. L., Notarberardino, B., Walker, B., and Abdul-Aziz, A. (2008). An efficient approach to converting three-dimensional image data into highly accurate computational models. *Philos. Transact. A Math. Phys. Eng. Sci.* 366, 3155–3173. doi: 10.1098/rsta.2008.0090
- Zhang, Y. (2013). "Challenges and advances in image-based geometric modeling and mesh generation," in *Image-Based Geometric Modeling and Mesh Generation*, ed Y. Zhang (Dordrecht: Springer), 1–10.
- Zhang, Y., Long, J. D., Mills, J. A., Warner, J. H., Lu, W., Paulsen, J. S., et al. (2011). Indexing disease progression at study entry with individuals at-risk for Huntington disease. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 156B, 751–763. doi: 10.1002/ajmg.b.31232

Conflict of Interest Statement: MM owns an equal share in Anatomyworks LLC. The terms of this arrangement have been reviewed and approved by the Johns Hopkins University, in accordance with its conflict of interest policy.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Tang, Luo, Chen, Huang, Johnson, Paulsen and Miller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership