

TBSS++: A novel computational method for Tract-Based Spatial Statistics

Davood Karimi¹, Hamza Kebiri^{1,2}, and Ali Gholipour¹

¹ Computational Radiology Laboratory (CRL), Department of Radiology, Boston Children's Hospital, and Harvard Medical School, USA

² Department of Radiology, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

Abstract. Diffusion-weighted magnetic resonance imaging (dMRI) is widely used to assess the brain white matter. One of the most common computations in dMRI involves cross-subject tract-specific analysis, whereby dMRI-derived biomarkers are compared between cohorts of subjects. The accuracy and reliability of these studies hinges on the ability to compare precisely the same white matter tracts across subjects. This is an intricate and error-prone computation. Existing computational methods such as Tract-Based Spatial Statistics (TBSS) suffer from a host of shortcomings and limitations that can seriously undermine the validity of the results. We present a new computational framework that overcomes the limitations of existing methods via (i) accurate segmentation of the tracts, and (ii) precise registration of data from different subjects/scans. The registration is based on fiber orientation distributions. To further improve the alignment of cross-subject data, we create detailed atlases of white matter tracts. These atlases serve as an unbiased reference space where the data from all subjects is registered for comparison. Extensive evaluations show that, compared with TBSS, our proposed framework offers significantly higher reproducibility and robustness to data perturbations. Our method promises a drastic improvement in accuracy and reproducibility of cross-subject dMRI studies that are routinely used in neuroscience and medical research.

Keywords: Diffusion MRI · deep learning · brain white matter · tract-based spatial statistics · TBSS

1 Introduction

Diffusion of water in the brain tissue is influenced by tissue micro-structure. Diffusion MRI (dMRI) methods can characterize this process quantitatively [13, 5]. These methods are widely used in medicine and neuroscience. Thousands of studies have used dMRI parameters such as fractional anisotropy (FA) as biomarkers to assess normal brain development and to characterize the impacts of various neurological diseases [3, 24, 14].

Tract-specific studies analyze dMRI-computed biomarkers on specific white matter tracts [25]. They may consider the same subject(s) over time (longitu-

dinal studies) or compare different cohorts of subjects such as healthy and diseased (population studies). They are among the most common and most fruitful classes of dMRI studies because many neurodevelopmental deficits and neurological disorders are linked to damage to the tissue micro-structure on specific tracts [17, 23, 34]. The success of these studies depends on precise alignment of different brain scans to ensure that the same tracts are compared between scans/subjects. This is a complex and highly error-prone computation. Existing computational methods vary greatly in terms of spatial accuracy and the required manual work [25]. For example, voxel-based morphometry methods are simple, but they are incapable of accurate tract-specific comparisons [2]. There are also semi-automatic and automatic tractography-based methods that require performing tractography on each scan [16, 27].

Currently, Tract-Based Spatial Statistics (TBSS) [26] is the *de-facto* method for tract-specific dMRI analysis. It can analyze the entire brain and consists of five steps: (1) Compute FA images for each subject; (2) Register the FA images to a common space; (3) Compute the voxel-wise average of the registered FA images and create the skeleton of this mean FA; (4) Project subjects' FA images onto the skeleton; (5) Perform statistical analysis across subjects on the skeleton. Despite its popularity, TBSS suffers from important shortcomings that have been extensively reported. Here we briefly list some of them: (1) FA-based registration is inaccurate, and subsequent steps in TBSS fail to correct the registration errors [32]. (2) Presence of lesions reduce the accuracy of TBSS, which is a serious limitation because such pathology may be the main focus of the analysis [11]. (3) Results are sensitive to measurement noise, choice of FA template, and FA threshold [15]. (4) TBSS is prone to confusing adjacent tracts even under ideal and simulated settings [4]. (5) Its central assumption is that the locations of interest are the tract centers, defined as the voxels with the highest FA. This assumption may not be valid. (6) Computational steps such as skeletonization and skeleton projection can introduce systematic bias [7]. (7) TBSS is not suited for all tract shapes and may produce erroneous results on small tubular tracts [4]. (8) Although TBSS is an automatic method, it involves several settings that may influence the results. The impacts of these settings are not fully clear, and one should visually verify the results of different processing steps [15].

These limitations significantly compromise the validity of the results of cross-subject dMRI studies. It has been shown that tweaking the settings of the existing methods may lead to completely different conclusions [4, 12]. Alternative methods have been proposed to improve certain aspects of TBSS [31, 30]. However, these alternative methods have their own limitations and none of them has reached TBSS's popularity. Despite years of effort, existing methods suffer from serious shortcomings that are data-dependent and difficult to predict.

Given the increasing popularity of dMRI and the growing size of datasets, there is an urgent need for accurate and reliable computational tools. The goal of this paper is to develop and validate a new computational framework, in large part based on deep neural networks, that overcomes some of the limitations of the existing methods. We show that the new methods enable unprecedented

accuracy in tract-specific cross-subject analysis of dMRI data. Therefore, they can have a significant impact as they dramatically enhance the reliability and reproducibility of longitudinal and population studies with dMRI.

2 Materials and methods

2.1 Data

We use two datasets in this work. (1) Human Connectome Project (HCP) dataset. We use 105 adult brains from the HCP dataset [28] and manual segmentations of 72 tracts on these brains available via [29]. The dMRI scans in this dataset include high angular resolution multi-shell ($b=1000, 2000$, and 3000) measurements. In all experiments with this data, we used 70 subjects for model development and the remaining 35 for test. (2) Calgary Preschool dataset [21, 20], which we only use for further validation. This dataset includes 396 dMRI scans from 120 children 2-8 years of age. Each scan in this dataset has 35 single-shell ($b=750$) measurements.

2.2 Methods

Figure 1 shows the proposed computational framework. It consists of three main components that we describe in the following three sub-sections.

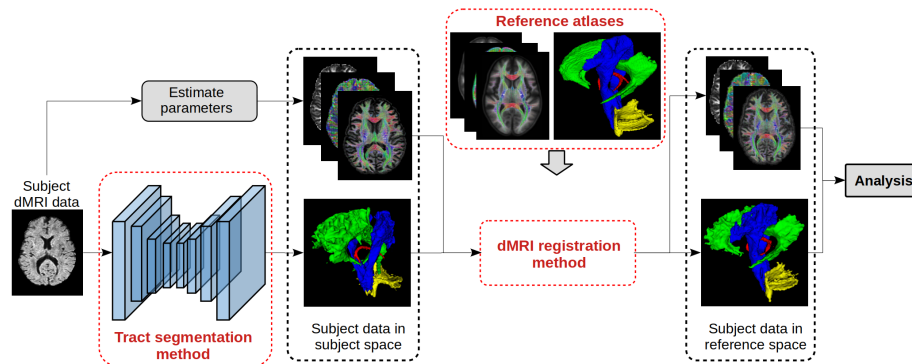


Fig. 1. A graphical summary of the proposed computational framework. To simplify the illustration, we have shown only one subject and only four tracts.

Tract segmentation method. Many cross-subject analysis methods, including TBSS, do not segment individual tracts, which causes two problems: (1) Tract-specific computation is not possible. In order to conduct a tract-specific

analysis, tracts will have to be segmented either manually or with an auxiliary method. (2) Ambiguity in the location and extent of the tracts can result in significant errors (e.g., mis-assignment of voxels) even for prominent tracts such as corpus callosum and cingulum [4]. Segmentation of individual tracts can overcome this fundamental limitation. To this end, we use a fully-convolutional network similar to nnU-Net [9] to segment the individual tracts. The segmentation method itself is not a novel aspect of this work. In fact, similar methods have been proposed in recent works [29].

Atlases of white matter tracts. We computed fiber orientation distribution images for the 70 training subjects using the MSMT-CSD method [10]. We aligned all 70 brains into a common space using an iterative succession of rigid, affine, and non-linear registrations computed based on fiber orientation distributions using the registration method proposed in [18]. We then used the composite deformation fields for each subject to align individual tract segmentations into the same space and used voxel-wise averaging and thresholding at 0.50 to compute tract atlases. We further applied the computed deformation fields to the diffusion MRI data of each subject, represented in spherical harmonics basis, to compute a corresponding atlas. These atlases would serve as an unbiased reference space where we could register the data from individual subjects for tract-specific analysis.

dMRI registration method. Accurate registration of brain dMRI data from different subjects is the most central computation in cross-subject studies. Existing methods mostly rely on registration based on fractional anisotropy (FA) images, which can be highly inaccurate. In our proposed framework, we require a flexible registration technique to register the subject brains to the reference atlas. Specifically, similar to our atlas computation described above, we use the method of [18] that is based on computing a nonlinear deformation map based on fiber orientation distributions.

To summarize: the combination of the tract segmentation method, the atlas, and the registration method constitute our proposed framework. The method works as follows (Figure 1): (i) Compute the parameter(s) of interest (e.g., FA) and segment tracts of interest for every subject; (ii) Align all subjects into the reference atlas space; (iii) Perform statistical analysis between desired sub-groups of the subjects to determine the differences in the parameters on each tract.

3 Results and Discussion

3.1 Assessment of the three components of the framework

Our main interest is on the effectiveness of the entire framework for performing tract-specific studies. Nonetheless, we first briefly explain the evaluation of the three components.

Assessment of the segmentation method. Our deep learning based segmentation method achieved a DSC of 0.826 ± 0.076 , which is comparable with the state of the art method [29] on the same dataset. However, unlike [29] that needs the data in all three shells, our method requires only the measurements in the $b=1000$ shell.

Assessment of atlases of white matter tract. Figure 2 shows several views from this atlas. All tracts were reconstructed in the atlas with high detail. In order to quantitatively assess the accuracy of the atlas, we divided the 70 training subjects into two groups, each with 35 subjects, generated two separate atlases, and computed the DSC between white matter tracts in the two atlas. The mean DSC across all tracts was 0.962, suggesting that the atlases could be used as an unbiased reference space for registering and comparing data from individual subjects.

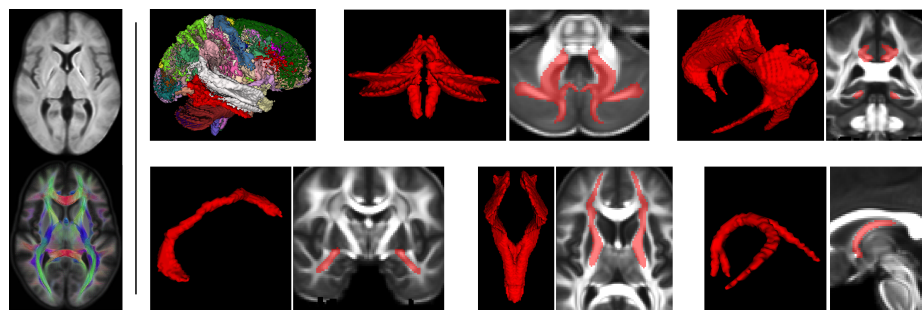


Fig. 2. LEFT: one slice from the dMRI data and fiber orientation distribution of the generated atlas. RIGHT (from top-left to bottom-right): a volume showing all 72 tracts, inferior cerebellar peduncle, cingulum, anterior commissure, frontopontine tract, and fornix.

Assessment of the registration method. The most common approach to quantitative evaluation of registration accuracy is in terms of the overlap of corresponding structures between moving and fixed images [6, 33]. Following this approach, we assessed registration accuracy in terms of DSC and average surface distance (ASD) between the warped subject tracts and atlas tracts. For this evaluation, we considered four of the tracts that spanned most of the brain volume: thalamo-prefrontal, striato-parietal, and anterior midbody and isthmus of corpus calosum. The registration accuracy in terms of DSC was 0.785 ± 0.074 , while the ASD was 0.775 ± 0.082 mm.

3.2 Assessment of the overall framework for tract-specific cross-subject analysis

In general, it is difficult to compare different computational frameworks for cross-subject studies. The aim of such studies is to compare two or more groups of subjects and identify the “true” differences between the groups. However, discovering such differences, which typically have complex biological underpinnings, is the *goal* of these studies and they are not known in advance. Therefore, for assessing computational methods, it is common to use synthetic data, simulations, and test-retest/reproducibility experiments [4, 26, 22, 1, 15]. Following similar evaluation strategies, we applied our new framework in several experiments and compared it with the official implementation of TBSS [8]. In these experiments, described below, to enable tract-specific comparison of our method with TBSS, we non-linearly registered the mean FA image from TBSS to the atlas FA image.

Reproducibility. Each dMRI scan from the HCP dataset has 60-90 measurements in the $b=1000$ shell. For each of the 35 test subjects from this dataset, we selected two independent subsets of 30 measurements from this shell. We applied our method and TBSS on these two measurement subsets and computed the average FA on each tract for each subject. Let us denote the absolute difference between the mean FA for tract i and subject j with $\Delta FA_{tract_i,j} = |FA_{tract_i,j}^{subset_2} - FA_{tract_i,j}^{subset_1}|$. Ideally, $\Delta FA_{tract_i,j}$ should be zero. Let us denote the average of $\Delta FA_{tract_i,j}$ across all subjects with ΔFA_{tract_i} . We computed ΔFA_{tract_i} on all tracts for our method and TBSS. On 35 out of 41 tracts, ΔFA_{tract_i} was smaller for our method than for TBSS. On 25 of these 35 tracts, the difference was statistically significant at $p = 0.01$ computed using paired t-tests. On the remaining 6 tracts, the difference was not significant. This result shows a higher reproducibility for our proposed method. We further repeated this experiment with only six measurements in each subset to simulate clinical scans. In this experiment, on all 41 tracts ΔFA_{tract_i} was smaller for our method than for TBSS; 22 of these differences were statistically significant at $p = 0.01$. We also plotted the profile of FA on the centerline of various tracts. The profiles computed with our method showed higher consistency and reproducibility.

Robustness to data perturbation. For each of the 35 test subjects from the HCP dataset, we added random Rician noise with a signal-to-noise-ratio of 20dB to the dMRI measurements. We applied our method and TBSS on these noisy scans as well as on the original scans and computed the difference in mean FA for tract i and subject j as $\Delta FA_{tract_i,j} = |FA_{tract_i,j}^{noisy} - FA_{tract_i,j}^{original}|$. We denote the average of $\Delta FA_{tract_i,j}$ across all subjects with ΔFA_{tract_i} . On all 41 tracts, ΔFA_{tract_i} was smaller for our method than for TBSS. Paired t-tests showed that 28 of these differences were statistically significant at $p = 0.01$. This result shows that our method is more robust to data perturbations.

Analysis of the impact of gender and age with an independent dataset. We performed additional experiments on the Calgary Preschool dataset. With the help of a semi-manual tract segmentation technique [19], this dataset had been previously used to assess the impact of age and gender on the micro-

structural changes on 10 major white matter tracts [20]. We applied our proposed method and TBSS on this dataset to determine the impact of gender and age on FA and Mean Diffusivity (MD), and compared the results with [20]. This experiment showed that, compared with TBSS, the results of our method were consistently closer to the results of [20]. For example, our method showed a significantly ($p < 0.01$) higher FA on pyramidal tracts and higher MD on the superior longitudinal fasciculus for boys compared with girls, which was consistent with the results of [20]; TBSS did not reveal these differences. We also performed a reproducibility experiment with this dataset. This dataset included 75 subjects (30 female) aged 3 years and 73 subjects (34 female) aged 6 years. We divided these into {Subset 1: 37 (15 female) aged 3 & 37 (17 female) aged 6} and {Subset 2: 38 (15 female) aged 3 & 36 (17 female) aged 6}. We computed the change in FA and MD for each tract using the two subsets. Ideally, the change computed with the two subsets should be equal. In terms of FA for 35 of the tracts and in terms of MD for all 41 tracts the difference between the changes computed using the two subsets was smaller with our method than with TBSS. This experiment shows higher reproducibility for the proposed method.

4 Conclusions

The potential impact of the proposed method cannot be overstated. Tract-specific studies are increasingly used to answer critical research questions in medicine and neuroscience. However, the shortcomings of existing computational methods have seriously limited the accuracy and reproducibility of these studies. Our experiments showed that the method proposed in this paper offers superior reproducibility and robustness to data perturbations. Experiments with the Calgary Preschool dataset further showed that our method has high generalizability. Therefore, our proposed method can have a significant impact by enhancing the accuracy, reliability, and reproducibility of the results of tract-specific studies.

Acknowledgment

This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering, the National Institute of Neurological Disorders and Stroke, and Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (NIH) under award numbers R01HD110772, R01NS128281, R01NS106030, R01EB018988, R01EB031849, R01EB032366, and R01HD109395. This research was also partly supported by NVIDIA Corporation and utilized NVIDIA RTX A6000 and RTX A5000 GPUs. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NVIDIA. This work was also supported by the Swiss National Science Foundation (project 205321-182602). We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University

8 Davood Karimi, Hamza Kebiri, and Ali Gholipour

of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

References

1. Arkesteijn, G.A., Poot, D.H., Ikram, M.A., Niessen, W.J., Van Vliet, L.J., Vernooij, M.W., Vos, F.M.: Orientation prior and consistent model selection increase sensitivity of tract-based spatial statistics in crossing-fiber regions. *IEEE transactions on medical imaging* **39**(2), 308–319 (2019)
2. Ashburner, J., Friston, K.J.: Voxel-based morphometry—the methods. *Neuroimage* **11**(6), 805–821 (2000)
3. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (dti)-based white matter mapping in brain research: a review. *Journal of molecular neuroscience* **34**(1), 51–61 (2008)
4. Bach, M., et al.: Methodological considerations on tract-based spatial statistics (tbss). *Neuroimage* **100**, 358–369 (2014)
5. Basser, P.J., Mattiello, J., LeBihan, D.: Mr diffusion tensor spectroscopy and imaging. *Biophysical journal* **66**(1), 259–267 (1994)
6. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis* **57**, 226–236 (2019)
7. Edden, R.A., Jones, D.K.: Spatial and orientational heterogeneity in the statistical sensitivity of skeleton-based analyses of diffusion tensor mr imaging data. *Journal of neuroscience methods* **201**(1), 213–219 (2011)
8. FMRIB Analysis Group, O.U.: Tbss userguide - fslwiki. <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide>, accessed: 2023-01-14
9. Isensee, F., et al.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods* **18**(2), 203–211 (2021)
10. Jeurissen, B., Tournier, J.D., Dhollander, T., Connelly, A., Sijbers, J.: Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion mri data. *NeuroImage* **103**, 411–426 (2014)
11. Jones, D.K., Cercignani, M.: Twenty-five pitfalls in the analysis of diffusion mri data. *NMR in Biomedicine* **23**(7), 803–820 (2010)
12. Keihaninejad, S., et al.: The importance of group-wise registration in tract based spatial statistics study of neurodegeneration: a simulation study in alzheimer’s disease. *PloS one* **7**(11), e45996 (2012)
13. Le Bihan, D., et al.: Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **13**(4), 534–546 (2001)
14. Lebel, C., et al.: A review of diffusion mri of typical white matter development from early childhood to young adulthood. *NMR in Biomedicine* **32**(4), e3778 (2019)
15. Madhyastha, T., et al.: Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. *Human brain mapping* **35**(9), 4544–4555 (2014)
16. Pagani, E., et al.: A method for obtaining tract-specific diffusion tensor mri measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis. *Neuroimage* **26**(1), 258–265 (2005)
17. Pini, L., et al.: Brain atrophy in alzheimer’s disease and aging. *Ageing research reviews* **30**, 25–48 (2016)
18. Raffelt, D., Tournier, J.D., Frupp, J., Crozier, S., Connelly, A., Salvado, O.: Symmetric diffeomorphic registration of fibre orientation distributions. *Neuroimage* **56**(3), 1171–1180 (2011)
19. Reynolds, J., Grohs, M., Lebel, C.: White matter tractography guides. FigShare. Available at **10**, m9 (2019)

20. Reynolds, J.E., Grohs, M.N., Dewey, D., Lebel, C.: Global and regional white matter development in early childhood. *Neuroimage* **196**, 49–58 (2019)
21. Reynolds, J.E., Long, X., Paniukov, D., Bagshawe, M., Lebel, C.: Calgary preschool magnetic resonance imaging (mri) dataset. *Data in brief* **29**, 105224 (2020)
22. Schwarz, C.G., Reid, R.I., Gunter, J.L., Senjem, M.L., Przybelski, S.A., Zuk, S.M., Whitwell, J.L., Vemuri, P., Josephs, K.A., Kantarci, K., et al.: Improved dti registration allows voxel-based analysis that outperforms tract-based spatial statistics. *Neuroimage* **94**, 65–78 (2014)
23. Sexton, C.E., et al.: A meta-analysis of diffusion tensor imaging in mild cognitive impairment and alzheimer’s disease. *Neurobiology of aging* **32**(12), 2322–e5 (2011)
24. Shenton, M.E., et al.: A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. *Brain imaging and behavior* **6**(2), 137–192 (2012)
25. Smith, S.M., Kindlmann, G., Jbabdi, S.: Cross-subject comparison of local diffusion mri parameters. In: *Diffusion MRI*, pp. 209–239. Elsevier (2014)
26. Smith, S.M., et al.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**(4), 1487–1505 (2006)
27. Suarez, R.O., et al.: Automated delineation of white matter fiber tracts with a multiple region-of-interest approach. *Neuroimage* **59**(4), 3690–3700 (2012)
28. Van Essen, D.C., et al.: The wu-minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)
29. Wasserthal, J., Neher, P., Maier-Hein, K.H.: Tractseg-fast and accurate white matter tract segmentation. *NeuroImage* **183**, 239–253 (2018)
30. Yushkevich, P.A., et al.: Structure-specific statistical mapping of white matter tracts using the continuous medial representation. In: *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8. IEEE (2007)
31. Yushkevich, P.A., et al.: Structure-specific statistical mapping of white matter tracts. In: *Visualization and Processing of Tensor Fields*, pp. 83–112. Springer (2009)
32. Zalesky, A.: Moderating registration misalignment in voxelwise comparisons of dti data: a performance evaluation of skeleton projection. *Magnetic resonance imaging* **29**(1), 111–125 (2011)
33. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10600–10610 (2019)
34. Zhuang, L., et al.: White matter integrity in mild cognitive impairment: a tract-based spatial statistics study. *Neuroimage* **53**(1), 16–25 (2010)