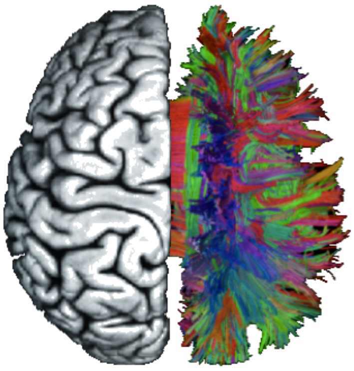
DEPARTMENT OF INFORMATION ENGINEERING AND COMPUTER SCIENCE
ICT International Doctoral School

Supervised Learning for White Matter Bundle Segmentation

PhD candidate:
Giulia Bertò



Advisor: Dr. Paolo Avesani[1,2]
Co-Advisor: Dr. Emanuele Olivetti[1,2]

[1]NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation (FBK), Trento, Italy
[2]Center for Mind and Brain Sciences (CIMeC), University of Trento, Italy

May 2020

The research described in this thesis was carried out at the
NeuroInformatics Laboratory (NILab), a collaboration
between Bruno Kessler Foundation (FBK) and
the University of Trento, Italy.

"The only source of knowledge is experience."

- ALBERT EINSTEIN

Abstract

Accurate delineation of anatomical structures in the white matter of the human brain is of paramount importance for multiple applications, such as neurosurgical planning, characterization of neurological disorders, and connectomic studies. Diffusion Magnetic Resonance Imaging (dMRI) techniques can provide, in-vivo, a mathematical representation of thousands of fibers composing such anatomical structures, in the form of 3D polylines called streamlines. Given this representation, a task of invaluable interest is known as **white matter bundle segmentation**, whose aim is to virtually group together streamlines sharing a similar pathway into anatomically meaningful structures, called white matter bundles. Obtaining a good and reliable bundle segmentation is however not trivial, mainly because of the intrinsic complexity of the data. Most of the current methods for bundle segmentation require extensive neuroanatomical knowledge, are time consuming, or are not able to adapt to different data settings. To overcome these limitations, the main goal of this thesis is to develop a new automatic method for accurate white matter bundle segmentation, by exploiting, combining and extending multiple up-to-date **supervised learning** techniques.

The main contribution of the project is the development of a novel streamline-based bundle segmentation method based on binary linear classification, which simultaneously combines information from atlases, bundle geometries, and connectivity patterns. We prove that the proposed method reaches unprecedented quality of segmentation, and that is robust to a multitude of diverse settings, such as when there are differences in bundle size, tracking algorithm, and/or quality of dMRI data. In addition, we show that some of the state-of-the-art bundle segmentation methods are deeply affected by a geometrical property of the shape of the bundles to be segmented, their fractal dimension.

Important factors involved in the task of streamline classification are: (i) the need for an effective streamline distance function and (ii) the definition of a proper feature space. To this end, we compare some of the most common streamline distance functions available in the literature and we provide some guidelines on their practical use for the task of supervised bundle segmentation. Moreover, we investigate the possibility to include, in a streamline-based segmentation method, additional information to the typically employed streamline distance measure. Specifically, we provide evidence that considering additional anatomical information regarding the cortical terminations of the streamlines and their proximity to specific Regions of Interest (ROIs) helps to improve the results of bundle segmentation.

Lastly, significant attention is paid to reproducibility in neuroscience. Following the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles, we have integrated our pipelines of analysis into an online open platform devoted to promoting reproducibility of scientific results and to facilitating knowledge discovery.

Keywords

White Matter Bundle Segmentation, Diffusion Magnetic Resonance Imaging (dMRI), Supervised Learning, Streamline Distances, Linear Classification, Brain Connectivity, Reproducibility for Neuroscience.

Publications

Journal Publications

- Bertò, G., Bullock, D., Astolfi, P., Hayashi, S., Zigiotto, L., Annicchiarico, L., Corsini, F., De Benedictis, A., Sarubbo, S., Pestilli, F., Avesani, P., Olivetti, E., *Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation.* [under review in NeuroImage]

- Pestilli, F. et al., *BrainLife, a free online platform for reproducible neuroscience.* [in preparation]

Conference Publications

- Olivetti, E., Bertò, G., Gori, P., Sharmin, N., Avesani, P., Jun. 2017. *Comparison of distances for supervised segmentation of white matter tractography.* In: 2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI 2017). IEEE, pp. 1–4.

- Bertò, G., Avesani, P., Pestilli, F., Bullock, D., Caron, B., Olivetti, E., Apr. 2019. *Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation.* In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 135–138.

- Astolfi, P., De Benedictis, A., Sarubbo, S., Bertò, G., Sona, D., Olivetti, E., Avesani, P., *A stem-based dissection of inferior fronto-occipital fasciculus with a deep learning model.* Accepted to the IEEE 17th International Symposium on Biomedical Imaging (ISBI 2020).

- Olivetti, E., Gori, P., Astolfi, P., Bertò, G., Avesani, P., *Nonlinear Alignment of Whole Tractograms with the Linear Assignment Problem.* Accepted to the 9th International Workshop on Biomedical Image Registration (WBIR 2020).

Conference Abstracts

- Bertò, G., Avesani, P., Pestilli, F., Bullock, D., Caron, B., Olivetti, E., Jun. 2019. *Segmentation of White Matter Bundles as Anatomically-Informed Multiple Linear Assignment Problems.* In Organization of Human Brain Mapping annual meeting (OHBM 2019).

- Olivetti, E., Gori, P., Astolfi, P., Bertò, G., Avesani, P., Jun. 2019. *Nonlinear Alignment of Whole Tractograms with the Linear Assignment Problem.* In Organization of Human Brain Mapping annual meeting (OHBM 2019).

- Olivetti, E., Bullock, D., Astolfi, P., Hayashi, S., Zigiotto, L., Annicchiarico, L., Corsini, F., De Benedictis, A., Sarubbo, S., Pestilli, F., Avesani, P., Bertò, G. *Classifyber: a linear classifier of single streamlines for white matter bundle segmentation.* In Organization of Human Brain Mapping annual meeting (OHBM 2020).

Main web applications

- Anatomically-informed multi-LAP: `https://doi.org/10.25663/brainlife.app.227`
- Classifyber: `https://doi.org/10.25663/brainlife.app.228`

Datasets

- HCP-canonical tractography dataset, a benchmark dataset of tractograms and well established white matter bundles: `https://doi.org/10.25663/brainlife.pub.3`

- HCP-minor bundle dataset, a dataset of white matter bundles connecting the human dorsal and posterior cortices: `https://doi.org/10.25663/brainlife.pub.11`

Acknowledgments

First of all, I wish to acknowledge all the people that have been by my side during this life changing experience, for their precious inspiration, guidance, and constant encouragement.

I would like to express my deep gratitude to my advisor Dr. Paolo Avesani for his excellent guidance and continuous support throughout my whole PhD journey. His academic advice and mentorship have been of unquestionable value towards the completion of this project. Thank you for having given me the opportunity to explore many different directions which significantly contributed to both my professional and personal growth.

I am extremely grateful to my co-advisor Dr. Emanuele Olivetti for having being by my side in every aspect of this project, from the early conceptualization to the final writing phase. Thank you very much for your essential research suggestions and your constant encouragement during all these years, for your invaluable help with data analysis, software management, and your writing tricks.

Both of them taught me that there is no success without failures, and to persist towards my goals. I really esteem you both, thank you for having pushed me beyond my limits to aspire to greater achievements.

A very special thanks to Prof. Franco Pestilli, who hosted me in his laboratory at Indiana University of Bloomington for my research internship. Thank you for your great suggestions and inspiration, for believing in me, and for letting me feel a bit more like at home in such a different and highly stimulating environment.

I would like to express my sincere gratitude to Prof. Eleftherios Garyfallidis and Prof. Laurent Petit, my reviewing committee members, for their time and precious feedback that contributed to greatly improve this manuscript.

I would like to thank also all the other collaborators I had the luck to work with during these years. I am particularly indebted to Dan Bullock, Doctor Silvio Sarubbo, Doctor Alessandro De Benedictis, and Luca Zigiotto, whose invaluable help in validating the datasets of this project significantly contributed to the added value of this research.

A very special thanks also to Soichi Hayashi for his impeccable support on software development and on the use of HPC systems.

I am also extremely thankful to Pietro, who shared with me not only the office, but

also most of the joys and sorrows of a PhD student. Thank you for your precious support and company while running experiments until late in the evening, and of course for our ritual of eating Chinese food while watching soccer matches.

A big thanks also to the other former and current PhD students whom I crossed the path with throughout the years, especially to Nusrat and Danilo, who gave me the strength to face the challenges at the very beginning of my PhD journey. Thanks also to Gabriele, for bearing me during the last past months in the office. I wish to thank as well all the other postdoctoral researchers, interns, and master students who spent some periods at NILab, and to all the "CIMeC people" I had the pleasure to meet and to share most of the days with. It has been a lot of fun organizing and taking part to our NILab "gourmet dinners" with you.

Moreover, a very special thanks to all my friends, the "old" ones and those met during these years, who have been very supportive and with whom I share wonderful memories and amazing experiences around the world. I am looking forward to more of these to come.

Last but not least, I am immensely grateful to my family, especially to my parents Giovanna and Maurizio, for their endless support and for always being present whenever I need it. Thank you very much for everything you do for me, and for giving me the courage to believe in myself and to pursue my dreams.

Contents

I Dissertation **1**

1 Introduction **3**
1.1 Motivation 3
1.2 Contributions 5
1.3 Thesis outline 8

2 Background **11**
2.1 Human brain basic anatomy 11
2.1.1 Brain tissues 12
2.1.2 White matter bundles 13
2.2 From the diffusion signal to the tractogram 16
2.2.1 Diffusion Magnetic Resonance Imaging (dMRI) 17
2.2.2 Signal reconstruction 19
2.2.3 Tractography algorithms and filtering 22
2.2.4 Tractography registration 27
2.3 White matter bundle segmentation 30
2.3.1 Validation 31
2.3.2 Applications 31
2.4 Basic principles of Machine Learning 33
2.4.1 Unsupervised vs Supervised Learning 34
2.4.2 Classification 35
2.4.3 Lazy learning methods 36

3 Related work **37**
3.1 Streamline distances and vectorial representation 37
3.1.1 Streamline distances 37
3.1.2 Streamline vectorial representation 39

3.2 White matter bundle segmentation techniques 42
3.2.1 Manual bundle segmentation approaches 42
3.2.2 Connectivity-based approaches 45
3.2.3 Streamline-based methods 47
3.2.4 Direct methods 51

4 Problem Statement 53
4.1 Basic notations and definitions 53
4.2 Research topics and questions 54
4.2.1 Streamline distances assessment 55
4.2.2 Beyond streamline geometry 56
4.2.3 Towards a robust streamline-based method for bundle segmentation 56
4.2.4 Summary of the main research questions 58

5 Methods 59
5.1 Streamline distance functions and dissimilarity representation 59
5.1.1 Streamline distance functions 59
5.1.2 Dissimilarity representation 61
5.2 Nearest Neighbor (NN) algorithm 61
5.3 Linear Assignment Problem (LAP) algorithm 62
5.4 Anatomically-Informed Linear Assignment Problem (LAP-anat) algorithm 63
5.4.1 Endpoint-based distance matrix 63
5.4.2 ROI-based distance matrix 64
5.5 Classifyber, a binary linear classification algorithm for streamlines 64
5.5.1 Feature Space 65
5.5.2 Training phase 69
5.5.3 Test phase 72
5.6 Evaluation metrics 74
5.6.1 Dice Similarity Coefficient (DSC) 74
5.6.2 Receiving Operating Characteristic (ROC) curve 74
5.7 Fractal Dimension (FD) 75

6 Materials 77
6.1 ROIs and parcellation atlases 77
6.1.1 MNI_JHU_tracts_ROIs_atlas 77
6.1.2 MNI152_ICBM2009c_reconstructed_atlas 78

6.2 Human Connectome Project (HCP) tractography datasets 79
6.2.1 HCP-EuDX dataset 80
6.2.2 HCP-IFOF dataset 80
6.2.3 HCP-ensemble dataset 81
6.3 Clinical tractography dataset 84

7 Experiments and Results 87
7.1 Comparison of streamline distances for bundle segmentation 87
7.1.1 Experimental design 88
7.1.2 Results 89
7.2 Anatomically-Informed Multiple Linear Assignment Problems (LAP-anat) for bundle segmentation 90
7.2.1 Experimental design 90
7.2.2 Results 91
7.3 Classifyber, a robust streamline-based linear classifier for bundle segmentation 93
7.3.1 Experimental design 93
7.3.2 Experiments and results on HCP-minor dataset 96
7.3.3 Experiments and results on HCP-IFOF dataset 97
7.3.4 Experiments and results on HCP-major dataset 99
7.3.5 Experiments and results on Clinical dataset 100
7.3.6 Experiments and results on fractal dimension 101
7.3.7 Size of the training set 103
7.3.8 Analysis of the computing time 103

8 Towards Reproducibility in Neuroscience 105
8.1 Reproducibility challenge 105
8.2 Promoting open sharing of scientific data and code 106
8.3 A reproducibility study: comparison of lazy learning methods for white matter bundle segmentation 108

9 Discussion and Conclusion 111
9.1 Discussion 111
9.1.1 Discussion about the comparison of streamline distance functions for supervised bundle segmentation 111

9.1.2 Discussion about the Anatomically-Informed Multiple Linear Assignment Problems (LAP-anat) for bundle segmentation 112
9.1.3 Discussion about Classifyber, a robust streamline-based linear classifier for bundle segmentation 113
9.1.4 Discussion about the comparison of lazy learning methods for bundle segmentation 117
9.2 Conclusion 118
9.3 Future work 119

II Papers 121

10 Comparison of Distances for Supervised Segmentation of White Matter Tractography 123

11 Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation 125

12 Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation 127

13 A reproducibility study: Comparison of lazy learning methods for white matter bundle segmentation 129

A Supplementary materials 131
A.1 Additional results related to LAP-anat 131
A.2 Preliminary results related to Classifyber 133

Bibliography 139

List of Tables

6.1 Clinical data of the 10 patients of the Clinical dataset. 85

7.1 Mean DSC voxel table across the 90 pairs of subjects, for each of the 9 different bundles, and each of the 8 different distance functions considered. 89
7.2 Computational time to compute 90000 pairwise streamline distances. . . . 89
7.3 Quantitative comparison over HCP-minor dataset: DSC (mean ± sd) across 25 target subjects for RecoBundles, TractSeg-retrained, LAP, LAP-anat and Classifyber. 96
7.4 Quantitative comparison over the HCP-IFOF dataset: DSC (mean ± sd) across 15 target subjects for RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber. 98
7.5 Quantitative comparison over HCP-major dataset: DSC (mean ± sd) across 21 target subjects for RecoBundles-atlas, TractSeg and Classifyber. 99
7.6 Quantitative comparison over the Clinical dataset: DSC (mean ± sd) across 7 target subjects for TractSeg, Classifyber-major, Classifyber-IFOF and Classifyber-LOSO. 101
7.7 Time in minutes required to train each method and to segment one IFOF for: RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber, when having 15 training examples. 104

8.1 Web links to the software repositories with the code developed to address the three research questions of the present thesis. 106
8.2 Link to the main BrainLife web applications we developed, implementing the two main methodological contributions of the present thesis. 107
8.3 Link to minor auxiliary BrainLife web applications we developed. 108
8.4 Link to the tractography datasets we share through the online platform BrainLife. 108

A.1 Quantitative comparison of LAP-anat, LAP-end, and LAP-roi over the HCP-canonical-large dataset. 132
A.2 Quantitative comparison of LAP-anat, LAP-end, and LAP-roi over the HCP-canonical-small dataset. 132
A.3 Quantitative comparison of LAP, ANTs-LAP, LAP-anat, and ANTs-LAP-anat over the HCP-canonical-large dataset. 133
A.4 Quantitative comparison of LAP, ANTs-LAP, LAP-anat, and ANTs-LAP-anat over the HCP-canonical-small dataset. 133

List of Figures

2.1 Illustration of a brain coronal slice and drawing of a neuron. 12
2.2 Historical drawings of long association white matter bundles. 13
2.3 Schematic representation of the pathways of four bundles connecting the posterior human cortex. 16
2.4 Visualization of the 7 parts composing the corpus callosum. 16
2.5 Water diffusivity in white and gray matter. 18
2.6 Illustration of two fiber populations within the same voxel, together with the diffusion tensor (DT) and the fiber orientation distribution function (fODF) reconstruction models. 22
2.7 A whole brain tractogram composed of hundreds of thousands of streamlines. 23
2.8 Reconstruction of the corticospinal tracts on a healthy subject using three different tractography pipelines. 26
2.9 Representation of the concept of segmentation of a bundle (the corticospinal tract) from a whole brain tractogram. 30
2.10 Examples of mean Fractional Anisotropy (FA) profiles of some white matter bundles. 33
2.11 Example of a learning machine. 34

3.1 Illustration of direct and flipped distances between two streamlines with the same number of points. 38
3.2 Toy example of streamline embedding. 39
3.3 Examples of ROI-based manual segmentation (virtual dissection) and ex-vivo dissection. 43
3.4 Automated Fiber Quantification (AFQ) workflow. 46
3.5 Linear Assignment Problem (LAP) workflow. 51
3.6 TractSeg workflow. 52

4.1 Supervised learning approach for white matter bundle segmentation. 54

4.2 Examples of different data settings. 57

5.1 Graphical intuition about how the individual distances are computed to form the distance matrix D, the endpoint-distance matrix E, and the ROI-based distance matrix R. 64
5.2 Vectorial representation of a streamline. 69
5.3 Standard sigmoid function that maps the relationship between dependent and independent variables. 71
5.4 Classifyber workflow. 73
5.5 Examples of white matter bundles with different fractal dimensions. 76
5.6 Linear regression analysis to compute the fractal dimension (FD) of the voxel mask of two white matter bundles with the box-counting method. . . 76

6.1 Terminal ROIs derived from the MNI152_ICBM2009c_reconstructed_atlas in one random HCP subject. 78
6.2 Streamline count of the 12 bundles selected in the HCP-canonical dataset in one random HCP subject. 82
6.3 Three bundles of the dataset HCP-canonical-large with their associated ROIs that define their path before they diverge towards the cortex. 82
6.4 Three bundles of the dataset HCP-canonical-small with their associated ROIs that define their path before they diverge towards the cortex. 82
6.5 Examples of four bundles of the HCP-minor dataset with their associated terminal ROIs. 83
6.6 Streamline count of the 8 minor bundles of the HCP-minor dataset in one random HCP subject. 83
6.7 Two examples of expert-based segmented Left_AF and Left_IFOF in two random patients of the Clinical dataset. 85

7.1 Mean DSC and standard deviation of the mean obtained across 30 subjects with the original method (LAP) an the proposed method (LAP-anat). . . . 92
7.2 Comparative paradigmatic example of a segmented uncinate fasciculus (UF) predicted by LAP and LAP-anat with the ground truth bundle. . . . 92
7.3 Quantitative comparison over HCP-minor dataset: mean DSC across 25 target subjects for RecoBundles, TractSeg-retrained, LAP, LAP-anat and Classifyber. 97

7.4 Qualitative comparison of segmented bundles in one target subject of the HCP-minor dataset predicted by RecoBundles, TractSeg-retrained, LAP, LAP-anat, and Classifyber with the expert-based segmentation. 97
7.5 Qualitative comparison of segmented bundles in one target subject of the HCP-IFOF dataset predicted by RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber with the expert-based segmented bundle. 98
7.6 Quantitative comparison over HCP-major dataset: mean DSC across 21 target subjects for RecoBundles-atlas, TractSeg and Classifyber. 100
7.7 Qualitative comparison of segmented bundles in one target subject of the HCP-major dataset predicted by RecoBundles-atlas, TractSeg, and Classifyber with the expert-based segmented bundle. 100
7.8 DSC vs FD across all methods for all the predicted bundles of the experiments in the work of Classifyber. 102
7.9 DSC (mean ± sd of the mean) across 25 test subjects of the HCP-minor dataset when varying the number of examples, from 1 to 60, for Classifyber. 103

8.1 Graphical example of how the web applications Anatomically-informed multi-LAP and Classifyber appear on the BrainLife website. 107
8.2 Examples of averaged Receiver Operating Characteristic (ROC) curve related to the LAP-based segmentation method and NN-based segmentation method when segmenting the right Cingulum bundle. 109

A.1 Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification on subjects registered with ACPC alignment, SLR registration, and ANTs registration. 134
A.2 Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification considering 7 different combinations of features. 135
A.3 Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification considering 4 different combinations of classification algorithms and solvers. . . . 137

Part I

Dissertation

Chapter 1

Introduction

1.1 Motivation

Technological development and innovation have had an incredible impact in the medical field, contributing to an undeniable improvement in patient care and to an overall increase in quality of life. Among the most outstanding technologies in medicine, neuroimaging techniques have become of invaluable help to understand the functioning of the human brain, in both pathological and healthy conditions.

Diffusion Magnetic Resonance Imaging (dMRI) is the only technique that can provide information about the orientation of the neuronal axons of the human brain in-vivo [Le Bihan et al., 1986, Basser et al., 1994]. This technique, being sensitive to the diffusion of the water molecules in the brain, allows us to study the local orientation of these axons within a voxel, the imaging volume unit. Local measurements can then be integrated over a longer spatial scale, tracing neuronal pathways and connections to build a model of the structural connectivity of the brain, known as the *tractogram* [Mori et al., 1999]. A tractogram is thus a collection of mathematical objects called *streamlines*, each of them representing thousands of neuronal axons. Given a tractogram, a task of particular interest is the so-called *white matter bundle segmentation*, whose aim is to group together streamlines sharing a similar path into anatomically meaningful structures, known as *white matter bundles*.

Accurate segmentation of white matter bundles in the human brain is of paramount importance for multiple applications, such as in neurosurgery for surgical planning, in neurology for diagnosis and monitoring of brain diseases, and in cognitive neuroscience for population studies. Obtaining a good and reliable bundle segmentation is however not trivial, mainly because of the intrinsic complexity of the data. In particular, the borders of the bundles are not always well-defined, many bundles overlap with others, and their shape and size have a large intra- and inter-subject variability. This variability can also be due, for example, to the adoption of different models to reconstruct the diffusion signal within a voxel, or to the use of different tracking algorithms to build the tractogram. How to effectively address the task of bundle segmentation in order to obtain accurate white matter bundles in the human brain is still an open problem and is the primary focus of this thesis.

Historically, white matter bundle segmentation was performed manually by expert neuroanatomists [Catani et al., 2002, Wakana et al., 2007]. In the last few decades, to overcome the limitations of *manual* approaches, which are very time consuming and difficult to reproduce, several *automatic* bundle segmentation techniques have been developed. Promising automatic methods for bundle segmentation have proven to be streamline-based methods based on *supervised learning*, such as those in [Garyfallidis et al., 2018, Sharmin et al., 2018]. These methods are able to learn how to segment a bundle of interest in a new subject by establishing *streamline* correspondences with examples of the same bundle available in other subjects. Alternatively, in order to apply advanced machine learning models, in [Wasserthal et al., 2018a] they have recently proposed a method to segment bundles by leveraging their volumetric representation as *voxel* masks, i.e. without the need for streamlines. In this way, they showed that their method can reach a high quality of bundle segmentation by using convolutional neural networks (CNNs), which are known to be the best machine learning models for grid-based domains like voxels.

Nevertheless, these state-of-the-art supervised methods suffer from some important drawbacks, such as dependency on multiple parameters, scalability issues, and inability to adapt to different data settings. The quality of segmentation may be strongly affected by some properties of the bundles, by the algorithm adopted to build the tractogram, and by the quality of the dMRI data. For example, most of the automatic methods can obtain accurate bundle segmentations when dealing with well characterized major bundles, but they are less effective in segmenting small and short bundles, which are notoriously harder to segment [Guevara et al., 2017]. However, small bundles have recently received increased attention and have proven to be of great help in understanding how information flows in the human brain [Bullock et al., 2019, Sani et al., 2019].

To overcome these limitations, the main goal of this thesis is to develop a robust automatic method for accurate white matter bundle segmentation, by exploiting, combining and extending multiple up-to-date supervised learning techniques. We believe that maintaining a high quality of bundle segmentation across different settings is of primary importance for subsequent applications.

1.2 Contributions

The main contribution of the present thesis is the development of a novel supervised method for white matter bundle segmentation which is able to adapt to multiple data settings. To achieve this goal, we propose *Classifyber*, a linear classifier that predicts whether or not a single streamline belongs to the bundle of interest. Classifyber combines the current knowledge in bundle segmentation, exploiting both the similarity between streamlines, typical of streamline-based methods, and the anatomical information from Regions of Interest (ROIs), typical of ROI-based and connectivity-based methods. Moreover, the proposed method is fast to compute and requires minimal resources. We demonstrate that Classifyber outperforms other state-of-the-art supervised methods with a large set of experiments across multiple different datasets. These datasets differ in terms of bundle size (large or small), tracking algorithm (probabilistic or deterministic), and dMRI data quality (research or clinical). Important factors involved in the task of streamline classification are (i) the need for an effective *streamline distance function* and (ii) the definition of a proper *feature space*.

Several streamline distance functions have been proposed in the literature, but, to the best of our knowledge, no comparison has be made in order to determine which one is the most effective for supervised bundle segmentation. We compare some of the most common streamline distance functions available in the literature and we provide some guidelines on their practical use for the task of supervised segmentation.

Current methods for bundle segmentation consider either only geometrical properties of the streamlines, or only prior knowledge about the relative anatomical position of a bundle. We investigate the possibility to take into account, in a streamline-based bundle segmentation method, information about both geometrical properties of the streamlines and their vicinity to anatomical ROIs. To this end, in addition to the typically employed streamline distance measure, we define other two anatomically-informed distance functions: the *endpoint-based* distance, which measures the average distance between the endpoints of two streamlines, and the *ROI-based* distance, which measures the distance between a streamline and an ROI. We prove that including this additional anatomical information into a streamline-based method helps to improve the results of bundle segmentation, especially when dealing with small bundles. These two distances are used to build a novel specifically-designed feature space for Classifyber.

We extensively compare a number of state-of-the-art methods for bundle segmentation across tractograms and expert-based segmented bundles of hundreds of subjects composing four different tractography datasets, three of which are novel. Three of the datasets were derived from the Human Connectome Project (HCP) [Van Essen et al., 2013], while the fourth dataset is an in-house clinical dataset of patients affected by brain tumors. Interestingly, we provide evidence that some of these segmentation methods are deeply affected by a geometrical property of the shape of the bundles to be segmented: their fractal dimension (FD). We show that these methods fail in particular to segment bundles that are more wrinkled and have a less rounded shape.

In addition, significant attention is paid to reproducibility in neuroscience. Typically, this requires a considerable effort towards code and data sharing initiatives through specific websites and repositories, a practice not so common until a few years ago, but that to-date is becoming standard within the neuroscientific community [Garyfallidis et al., 2014, Avesani et al., 2019]. For example, free and open source software like *DIPY* [Garyfallidis et al., 2014] and *MRtrix* [Tournier et al., 2012] were essential for all the activities of this thesis. For this reason, following the FAIR (Findable, Accessible, Interoperable and Reusable) Data Principles [Wilkinson et al., 2016], we provide the source code we developed within this thesis as open source software. Moreover, we have integrated our pipelines of analysis, composed of code, data, and environment of execution, into an online open platform called *BrainLife*, which is devoted to promoting reproducibility of scientific results.

Lastly, we perform a reproducibility study with the aim of reproducing some of the findings contained in the recent work of [Sharmin et al., 2018]. The main claim of the work is that "LAP [Linear Assignment Problem]-based segmentation [...] is substantially more accurate than the NN [Nearest Neighbor] strategy". The objective of our study is to independently assess, by replicating the same empirical study, whether this claim holds even in the case of diverse experimental conditions, such as for example a different preprocessing pipeline and tracking algorithm. Our results demonstrate that the original claim can be generalized, thus bringing added value to the original scientific findings.

To summarize, our major contributions are that we:

- propose a new streamline-based bundle segmentation method, with a novel specifically-designed feature space, which is robust to different data settings, specifically that can deal with either (i) large/small bundles, (ii) probabilistic/deterministic tractography, and (iii) research/clinical dMRI data quality

- combine, in a streamline-based bundle segmentation method, the similarity between streamlines, typical of streamline-based methods, with the anatomical information from ROIs, typical of ROI-based and connectivity-based methods
- publish open (i) code, (ii) data, and (iii) environment of execution as web applications to promote openness and facilitate reproducibility in neuroscience

Additionally, other minor contributions of this work are that we:

- compare several streamline distance functions available in the literature and their effect on supervised white matter bundle segmentation
- extensively compare a number of state-of-the-art methods for bundle segmentation across multiple tractography datasets with different characteristics
- show that some bundle segmentation methods are deeply affected by a geometrical property of the shape of the bundles, their fractal dimension
- replicate some of the scientific findings contained in the recent work of [Sharmin et al., 2018] with different experimental conditions, confirming that the original claim can be generalized

1.3 Thesis outline

This thesis is divided into two parts: **Part I**, the Dissertation, and **Part II**, the Papers.

Part I describes the research activities carried out within this thesis. A few results were omitted since the ultimate goal was to present the main scientific findings of the project. For further details and results, refer to Part II, which comprises the main scientific papers published, under review and in preparation resulting from our work.

Part I is structured as follows:

In Chapter 2, we present an overview of the *background* knowledge needed to develop the main methodological contribution of the present thesis. In Section 2.1 is described the human brain basic anatomy, with a particular focus on the definition and categorization of the white matter bundles subjects of study of the present thesis. In Section 2.2.1, we present the basic principles of diffusion in the brain and the pipeline to build a whole brain tractogram starting from raw diffusion data. We also briefly describe some tractography applications like filtering and registration. Section 2.3 is devoted to the problem of white matter bundle segmentation, which is the main focus of the present thesis. Lastly, Section 2.4 explains the basic principles of Machine Learning, and in particular the key concept of supervised learning.

In Chapter 3, *related work* about the research topics of the present thesis are reported. In Section 3.1, first, we report related work about streamline distance functions and their comparison in the literature, and then, we present state-of-the-art streamline vectorial representation techniques. Section 3.2 is aimed at revising state-of-the-art techniques commonly employed to address bundle segmentation, with a particular focus on those adopted or compared in the experimental part of the thesis. Additionally, we will discuss advantages and disadvantages of each of them, a kind of analysis that was fundamental to make grow our research interest and to drive our methodological research on this topic.

In Chapter 4, we describe the *open problems* we aim to address with this thesis. First, in Section 4.1, we provide some basic notations and definitions in order to formulate the problem of supervised white matter bundle segmentation. Then, in Section 4.2, we present the main research topics of the thesis and we summarize the three main research questions we want to answer with **Q1**, **Q2** and **Q3**.

In Chapter 5, we present the *methods* used and developed to address the research questions of this thesis. Sections 5.1, 5.2 and 5.3 describe key concepts and algorithms available in the literature that were auxiliary to develop the main methodological contributions of the thesis, which are presented in detail in Sections 5.4 and 5.5. Section 5.6 reports the evaluation metrics used to quantify the bundle segmentation performances, while in Section 5.7 we introduce the notion of fractal dimension (FD) of a bundle, which will be used to discuss some of the experimental results.

Chapter 6 is devoted to all the data sources and *materials* used for the experiments of this thesis. In Section 6.1, we report the Regions of Interest (ROIs) atlas and the parcellation atlas exploited to derive the waypoint ROIs and the terminal ROIs related to the bundles of interest. In Section 6.2, we describe the tractography datasets derived from the Human Connectome Project (HCP), while Section 6.3 describes an in-house tractography clinical dataset acquired at Santa Chiara's Hospital of Trento, Italy.

In Chapter 7, we report the *experiments* conducted and the *results* obtained in order to address each of the three main research questions of this thesis. Specifically, Section 7.1 addresses the research questions **Q1**: "Which streamline distance function available in the literature is the most effective for supervised bundle segmentation?". Section 7.2 addresses the research questions **Q2**: "Why not including prior anatomical information in a streamline-based bundle segmentation method?". Section 7.3 addresses the research questions **Q3**: "Can we develop a novel robust streamline-based bundle segmentation method which is able to adapt to multiple data settings?".

Chapter 8 is devoted to *reproducibility* for Neuroscience. Section 8.1 briefly illustrates the general reproducibility challenge within the scientific community. In Section 8.2, we describe our contributions to promote openness and facilitate reproducibility in neuroscience. Lastly, in Section 8.3, we report a reproducibility study conducted by us about comparison of lazy learning methods for white matter bundle segmentation.

Finally, Chapter 9 concludes Part I of the present thesis. In Section 9.1, we *discuss* the results of the research activities that have been presented in the previous Chapters. Section 9.2 provides the *conclusions* of this thesis, and finally, some future activities that can arise from the work presented in this thesis are collected in Section 9.3.

Part II is structured as follows:

Chapter 10 contains our contribution to address the research question **Q1**. The work, titled "*Comparison of Distances for Supervised Segmentation of White Matter Tractography*", is published in: 2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, pp. 1–4. This work is the results of an internal activity of the NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation (FBK), Trento, Italy.

Chapter 11 contains our contribution to address the research question **Q2**. The work, titled "*Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation*", is published in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 135–138. This work is the results of a collaboration with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA.

Chapter 12 contains our contribution to address the research question **Q3**. The work, titled "*Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation*", is under review in the NeuroImage scientific journal. This work is the results of a collaboration with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA, the Neurosurgery Unit, Department of Neuroscience and Neurorehabilitation, Bambino Gesù Children's Hospital, IRCCS, Roma, Italy, and the Division of Neurosurgery, Structural and Functional Connectivity Lab, S. Chiara Hospital, Trento, Italy.

Chapter 13 contains a reproducibility study titled "*A reproducibility study: Comparison of lazy learning methods for white matter bundle segmentation*", which is in preparation and is part of a larger contribution to be submitted to a scientific journal. This work is the results of a collaboration with the Department of Psychological and Brain Sciences, Indiana University, Bloomington, USA.

Appendix A concludes this thesis and comprises some supplementary materials. Specifically, Section A.1 describes some additional results related to the work presented in Chapter 11, while Section A.2 describes some preliminary results related to the contribution presented in Chapter 12.

Chapter 2

Background

In this Chapter, we present an overview of the background knowledge needed to develop the main methodological contribution of the present thesis. In Section 2.1 is described the human brain basic anatomy, with a particular focus on the definition and categorization of the white matter bundles subjects of study of the present thesis. In Section 2.2.1, we present the basic principles of diffusion in the brain and the pipeline to build a whole brain tractogram starting from raw diffusion data. We also briefly describe some tractography applications like filtering and registration. Section 2.3 is devoted to the problem of white matter bundle segmentation, which is the main focus of the present thesis. Lastly, Section 2.4 explains the basic principles of Machine Learning, and in particular the key concept of supervised learning.

2.1 Human brain basic anatomy

The brain is one of the most complex organs of the human body. It is responsible for regulating all human activities, which are made possible by the conduction of electrical impulses that transmit information from one body region to another. At the macroscopic level, the brain is composed of different elements, such as blood, cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM). The CSF is a colorless fluid located in the ventricular system inside and around the brain, and it provides to the brain a basic immunological and mechanical protection [Khan, 2014]. The gray and white matter are instead the two brain tissues, which in the living brain appear in pinkish-gray and in white color respectively, see Figure 2.1.

2.1.1 Brain tissues

The core component of the brain are the neurons, which regulate all the communication and neural processing. A single neuron comprises a cell body, an axon and several dentrites, see Figure 2.1.

The *gray matter* consists of dense concentrations of cell bodies and dendrites of these neurons, and it is where all the brain processing occurs [Porro-Muñoz et al., 2015]. The cerebral cortex, which constitutes the outermost layer of the brain, is the part of the gray matter involved in cognitive functions. Its folded surface creates grooves (called sulci) and ridges (called gyri).

The *white matter* is composed of billions of myelinated neuronal axons that connect different regions of the brain and occupies 50% of the brain volume in adults [Martino and De Lucas, 2014]. These axons, which transmit the nerve impulses between the neurons, constitute a "wiring system", also known as *anatomical connectivity* or *structural connectivity* [Sporns et al., 2005]. The anatomical connectivity varies across individuals and can change for example due to aging or mental disorders, such as dementia or Alzheimer's disease [Porro-Muñoz et al., 2015].

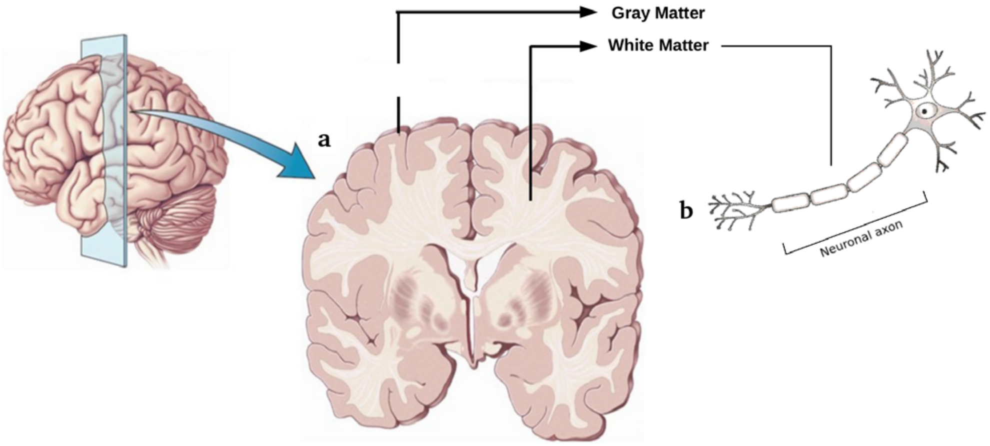


Figure 2.1: (a) Illustration of a brain coronal slice where it is visible the subdivision into gray and white matter. (b) Drawing of a neuron: the neuronal axon, surrounded by an insulating myelin sheath, is part of the white matter, while the cell body and the dendrites are part of the gray matter. Image adapted with permission from [Sharmin, 2017].

White matter myelinated axons can be distributed diffusely or can be densely packed into organized *white matter bundles*, also called fascicles or tracts, whose structure and architecture are subjects of study of the present thesis.

2.1.2 White matter bundles

White matter bundles are anatomically meaningful structures composing the brain white matter. A single bundle can be seen as a collection of white matter axons that smoothly travel together and link brain regions with similar functional roles [Bullock et al., 2019]. Earliest studies of human white matter were conducted in the 19th century, during which the brain anatomy could be investigated only after death. Pioneering neuroanatomical studies, such as those of the anatomists Paul Broca (1824-1880), Theodor Hermann Meynert (1833-1892), Carl Wernicke (1848-1905), and Jules Dejerine (1849-1917) had been of great importance to build the foundations of modern neuroanatomy. The anatomist Theodor Hermann Meynert [Meynert, 1885] proposed to arrange the bundles into three main categories, which are widely recognized still today: (i) the *association bundles*, which connect different cortical regions within the same hemisphere, (ii) the *commissural bundles*, which interconnect brain regions of the two hemispheres, and (iii) the *projection bundles*, which connect the cerebral cortex with subcortical centers and the spinal cord.

Association bundles

The association bundles connect different cortical regions within the same hemisphere. A conventional further subdivision differentiate between *long association bundles* and *short association bundles*.

Long association bundles are located in deeper areas of the white matter and connect different lobes within the same hemisphere. Their anterior-posterior orientation was originally determined by post-mortem anatomical studies and camera lucida drawings, such as those of [Edinger, 1885] and [Gray, 1918], see for example Figure 2.2.

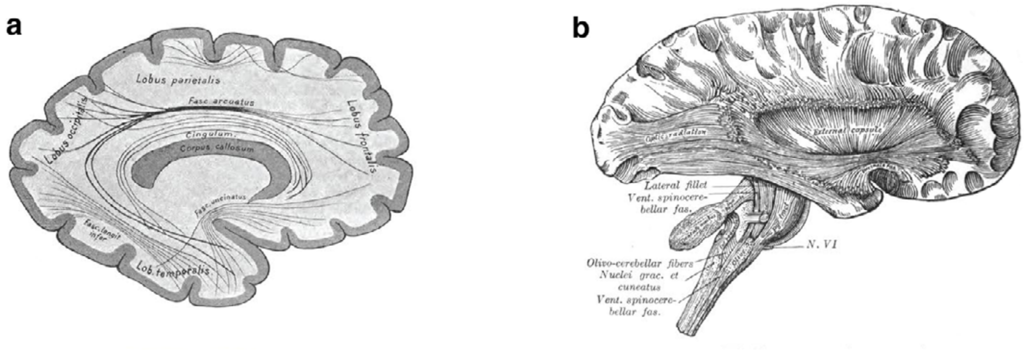


Figure 2.2: Historical drawings of long association white matter bundles of (a) Edinger [Edinger, 1885] and (b) Gray [Gray, 1918]. Image adapted with permission from [Bullock et al., 2019].

Existence and location of long association bundles are well established in the literature since they seem to share many common anatomical features across individuals [Oishi et al., 2008, Guevara et al., 2012, Wandell, 2016, Wu et al., 2016b, Pestilli, 2018, Bullock et al., 2019]. Here, we briefly describe some of the most known long association bundles conforming the definition of classical post-mortem studies and later confirmed by [Catani and De Schotten, 2008] (some of these bundles are depicted in Figures 6.3, 6.4 and 6.7):

- the *arcuate fasciculus* (AF) is a lateral bundle connecting the perisylvian cortex of the frontal, parietal, and temporal lobe. The left AF is usually involved in language, while the right AF in some aspects of language and in visuospatial processing.
- the *inferior fronto-occipital fasciculus* (IFOF) is a ventral bundle that links the occipital lobe and the orbitofrontal cortex. Its precise anatomical definition is to-date unclear and debated [Sarubbo et al., 2013, Forkel et al., 2014, Wu et al., 2016a], as well as its function. However, it probably plays a role in reading, attention and visual processing.
- the *inferior longitudinal fasciculus* (ILF) is a ventral bundle connecting the occipital and temporal lobes. It participates in face recognition, reading, visual perception, visual memory and other functions related to language.
- the *uncinate fasciculus* (UF) is a ventral bundle that links the anterior temporal lobe with the medial and lateral orbitofrontal cortex. It is probably related to memory, emotion processing and language.
- the *cingulum bundle* (CG) is a medial bundle that runs within the cingulated gyrus around the corpus callosum. It is sometimes divided into cingulum cingulate (CGC) and cingulum hippocampus (CGH) [Wakana et al., 2007], and plays a role in attention, emotions and memory.

Short association bundles, also called *U-fibers*, connect adjacent gyri and lie in what is called superficial white matter (SWM), which is the space between the deep white matter (DWM) and the cortex. Due to their location, number and functions, these bundles are not well characterized in the literature [Oishi et al., 2008, Guevara et al., 2011, Wu et al., 2016b].

An additional group of associative bundles that has been recently receiving increased attention comprises those bundles connecting the posterior human cortex

[Wu et al., 2016b, Bullock et al., 2019, Sani et al., 2019]. Also in this case, due to their relatively smaller size and shorter length with respect to long associative bundles, these bundles have been less studied in the past [Bullock et al., 2019]. Over the past few years, the functions attributed to them has expanded from pure language perception to more complex processes [Wu et al., 2016b]. For example, they have been found to play a role in attention [Sani et al., 2019] and in transferring information from superior to inferior horizontal bundles [Bullock et al., 2019]. Below, we list some of these bundles connecting dorsal to ventral cortical regions, following the anatomical definition in [Bullock et al., 2019] (see Figures 2.3 and 6.5 for their illustrations):

- the *posterior arcuate fasciculus* (pArc) is a vertically oriented bundle that connects the inferior parietal with the middle and inferior temporal lobes.
- the *temporo-parietal connection to the superior parietal lobule* (TP-SPL) is a dorso-ventrally oriented bundle connecting the parietal and temporal lobes.
- *the middle longitudinal fasciculus–superior parietal lobule component* (MdLF-SPL) is an obliquely oriented bundle connecting the parietal and latero-temporal lobes.
- the *middle longitudinal fasciculus–superior angular gyrus component* (MdLF-Ang) is, like the MdLF-SPL, an obliquely oriented bundle connecting the parietal and latero-temporal lobes.

Commisural bundles

The commisural bundles interconnect brain regions of the two hemispheres. The largest bundle of this category is the *corpus callosum* (CC), which is also the the largest bundle of the human brain. It connects the left and right hemispheres and is generally subdivided into four parts: the genu, the rostrum, the body and the splenium. An alternative subdivision considers the CC composed of seven parts [Witelson, 1989], see Figure 2.4. The CC allows the transfer of information from one hemisphere to the other and plays a role in several cognitive, perceptual and motor function [Catani and De Schotten, 2008].

Projection bundles

The projection bundles connect the cerebral cortex with the diencephalon, the corpus striatum, the brain stem, and the spinal cord [Mancall and Brock, 2011]. The most known bundles belonging to this category are the corticospinal tract (CST) and the thalamic radiations (TR) (see Figure 6.3 for their illustrations). The CST is a vertically oriented

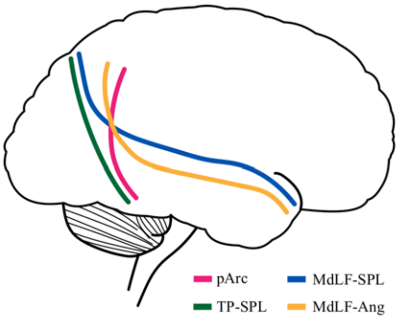


Figure 2.3: Schematic representation of the pathways of four bundles connecting the posterior human cortex: the pArc (pink), the TP-SPL (green), the MdLF-SPL (blue) and the MdLF-Ang (yellow). Image inspired from [Bullock et al., 2019].

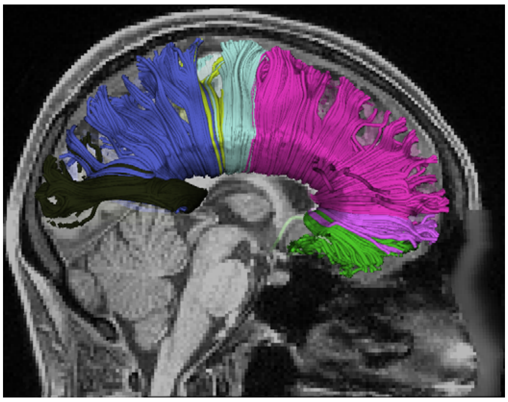


Figure 2.4: Visualization of the 7 parts composing the corpus callosum based on the subdivision of [Witelson, 1989], depicted with different colors.

bundle connecting the cerebral cortex to the spinal cord, and it is mainly involved in movement control. The TR are composed of thalamo-cortical and cortico-thalamic fibers, forming a continuous fan. These fibers are generally divided into four subgroups: inferior, anterior, superior and posterior radiation. Functionally, they transmit motor or sensory information from the thalamus to different regions of the cerebral cortex through relay neurons [George and M Das, 2019].

2.2 From the diffusion signal to the tractogram

Technological development and innovation have had an incredible impact in the medical field, contributing to an undeniable improvement in patient care and to an overall increase in quality of life. In the last few decades, thanks to a continuous increase in computational power, together with significant advances in the pattern recognition and digital image processing fields, cutting-edge technologies like modern imaging techniques, robotic surgery, and augmented reality have been established as essential tools for a better clinical practice, from prevention and earlier diagnosis to improved surgical procedure and medical treatments. Among the most outstanding technologies in medicine, neuroimaging techniques have become of invaluable help to understand the functioning of the human brain, in both pathological and healthy conditions.

Besides Computed Tomography (CT) and Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI) is of particular relevance because it provides a wide range of image contrasts without exposing people to harmful radiation. In addition to structural MRI measurements, several advanced MRI techniques sensitive to oxygenation [Ogawa et al., 1990], perfusion [Buxton et al., 1998], or blood flow [Dumoulin and Turski, 2007] have been developed. Moreover, there are specialized MRI modalities which able to derive information about the architecture of the white matter tissue, a research topic that, as mentioned above, has become of paramount importance in neuroscience. Among those techniques, one of the most powerful is called *diffusion MRI* (dMRI) [Le Bihan et al., 1986, Basser et al., 1994], which exploits the principle of diffusion of water molecules in the brain.

In what follows, we briefly describe the basic principles of diffusion in the brain[1] and the pipeline to build a whole brain tractogram starting from raw diffusion data.

2.2.1 Diffusion Magnetic Resonance Imaging (dMRI)

As all the other molecules in a fluid, water molecules are subjected to a constant thermally driven motion, called Brownian motion [Brown, 1828]. In an unrestricted homogeneous medium, their trajectories are strongly modified by their mechanical collisions and can be described by a "random walk". This phenomenon is known as the *diffusion phenomenon*, a physical process indispensable for the normal functioning of living systems.

Mathematically, the Brownian motion in 3D space of a particle in a fluid can be described by the following equation [Einstein, 1905]:

$$< r^2 >= 6Dt \tag{2.1}$$

where $< r^2 >$ is the root mean square displacement in any direction after a time t, and D is the *diffusion coefficient* measured in mm^2/s, which is dependent from the size of the molecules, the viscosity and the temperature of the medium.

In an unrestricted homogeneous medium, the probability distribution function of the motion $p(\mathbf{r}, t)$ of any particle moving from the origin to the position r after a diffusion time t is given by the following Gaussian distribution:

$$p(\mathbf{r}, t) = \frac{1}{\sqrt{(4\pi t)^3 D}} e^{-\frac{\mathbf{r}^T\mathbf{r}}{4tD}} \tag{2.2}$$

where $p(\mathbf{r}, t)$ is also known as the *diffusion propagator*. This Gaussian property is true only in the case of free diffusion, i.e. without obstacles restricting the diffusion process [Descoteaux, 2015], while it is not preserved in heterogeneous tissues where biological obstacles hinder or restrict the motion of molecules [Le Bihan et al., 1993].

---

[1] A review on the topic can be found for example in [O'Donnell and Westin, 2011].

In the human brain, the diffusion process of the water molecules may substantially vary across the biological tissues due their microscopic heterogeneity [Pandit, 2013]. Except in the CSF and to some extent in the gray matter, in which the molecules are relatively free to move without having a preferred direction, in the white matter their movement is restricted in certain directions due to the presence of cell membranes, axons, macro-molecules, and intracellular organelles. In the case of free diffusion, the diffusion is described as being *isotropic*, while in the opposite case as *anisotropic* [Pandit, 2013], see Figure 2.5. Specifically, in the white matter, the water molecules tend to diffuse more along the principal direction of the axons than in the perpendicular direction, and thus the diffusion coefficient appears to be lower with respect to that in an unrestricted medium [Descoteaux, 2015]. To take into account that this effect is related to the underlying tissue properties, [Le Bihan et al., 1986] introduced the term *apparent diffusion coefficient* (ADC).

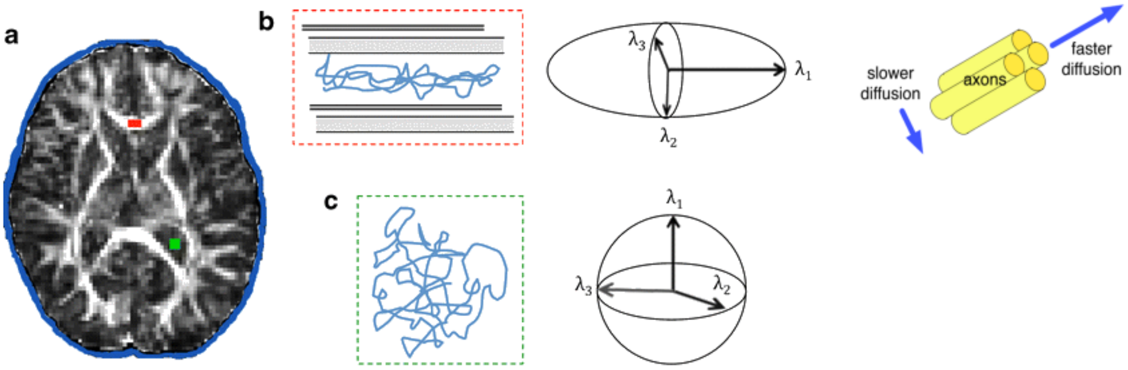


Figure 2.5: Water diffusivity in white and gray matter. (a) An axial slice with highlighted a voxel of the white matter (red) and a voxel of the gray matter (green). (b) Schematic illustration of the anisotropic diffusion of water molecules in the white matter voxel, which is hindered by obstacles and can be represented by an ellipsoid. Water diffusivity is faster along the principal direction of the axons. (c) Schematic illustration of the isotropic diffusion of water molecules in the gray matter voxel, which can be represented by a sphere. Images taken from [Pandit, 2013] and from [O'Donnell and Westin, 2011].

Diffusion Magnetic Resonance Imagin (dMRI) is a magnetic resonance technique that takes advantage of these different tissue properties to indirectly provide information about the microstructure of the human brain. To measure the diffusion, additional magnetic field gradients are used to create an image that is sensitive to the motion along a particular direction. For stationary molecules, the effect of the gradient pulses cancels out, while for molecule that diffuse parallel to the direction of the gradient, it causes a random phase shift that produces a signal loss [O'Donnell and Westin, 2011].

If S_0 is the signal intensity in the absence of a diffusion gradient and S_b is the decreased signal, in the presence of Gaussian diffusion, the diffusion coefficient D can be estimated by using the Stejskal-Tanner equation [Basser, 1995]:

$$S_b = S_0 e^{-bD} \tag{2.3}$$

The term b is known as the b-value [Le Bihan, 1991], and depends on the strength, the duration, and the temporal interval between the two diffusion gradients. It is measured in s/mm^2 and is defined as:

$$b = (\gamma \delta G)^2 \Delta t \tag{2.4}$$

where γ is a physical constant known as the gyromagnetic ratio[2], δ is the duration of the gradient in milliseconds, G is its amplitude and Δt is the time interval between the two gradients.

2.2.2 Signal reconstruction

By acquiring a large number of images for different gradient strengths and directions, it is possible to estimate the 3D diffusion probability density function (PDF) in each voxel. This procedure is commonly known as *reconstruction* since it aims to reconstruct the diffusion signal within each voxel, thus providing information about the main direction of diffusion [Assemlal et al., 2011]. Several different reconstruction methods have been proposed, each of them characterized by a different choice of signal model and parameter setting. In the following sections, we briefly present some of the most common diffusion models.

Diffusion Tensor Imaging (DTI)

In the case of isotropic (Gaussian) diffusion, the signal S_b decays exponentially with b following equation 2.3. In this circumstance, D is a scalar, which is equal to the ADC and can be computed as:

$$ADC = \frac{log(S_0/S_b)}{b} \tag{2.5}$$

In the case of anisotropic diffusion, i.e. when the diffusion is restricted in some of the

---

[2] The gyromagnetic ratio of the ^{1}H is $\gamma = 42.58$ MHz/T.

directions, the simplest representation of the 3D diffusion is the *diffusion tensor* (DT) model [Basser et al., 1994]. The diffusion tensor $\mathbf{D}$ is no longer a scalar, but a 3x3 positive definite symmetric matrix that can be described by an ellipsoid (instead of a sphere) with 3 principal axes:

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix} \tag{2.6}$$

where the diagonal elements correspond to the displacement along the orthogonal axes of the scanner and the off-diagonal elements to the correlation between the displacements along those axes [Pandit, 2013]. Since $\mathbf{D}$ is symmetrical, by acquiring at least six diffusion weighted directions and one non-diffusion weighted volume, its six unknown parameters can be computed with an extension of equation 2.3 [Descoteaux, 2015]. The directional components along the three main directions of the ellipsoid are represented by the eigenvectors ν_1, ν_2 and ν_3, while the scalar components by the eigenvalues λ_1, λ_2 and λ_3. The longest axis of the ellipsoid corresponds to the direction along which the diffusion is the highest, and it is associated with λ_1 [Pandit, 2013]. The fractional anisotropy (FA) is a normalized index that represents the degree to which this ellipsoid is anisotropic. The FA can take values form 0, which means completely isotropic, to 1, which means completely anisotropic, and can be calculated using the following formula [Pierpaoli et al., 1996]:

$$FA = \sqrt{\frac{1}{2}} \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}} \tag{2.7}$$

Although being widely used from both neuroscientists and clinicians, the DT model has some important limitations. The main one is the assumption of the Gaussian diffusion model, which allows to describe only a single-fiber population per voxel [Descoteaux, 2015]. In voxels containing multiple-fiber populations, the DT model is violated because the diffusion cannot be modeled by a Gaussian distribution. This situation is very common because, as [Jeurissen et al., 2013] estimated, approximately the 90% of the white matter voxels contain multiple-fiber populations. The DT model thus represents an oversimplification of the actual underlying microstructure, leading to an erroneous orientation estimation that can cause tensor-derived measures, such as the FA, to be unreliable or misleading [Jones, 2008]. As a consequence, it is necessary to go beyond the DT model, and exploit new reconstruction techniques that are able to describe non-Gaussian diffusion signals [Descoteaux, 2015]. To this end, high angular resolution diffusion imaging (HARDI) techniques have been developed.

High angular resolution diffusion imaging (HARDI)

HARDI techniques were firstly developed to overcome the limitations of the DT model and solve the crossing fibers issue[3], see Figure 2.6. To account for more complex fiber architectures, several different alternatives to the DT model have been proposed, which can be divided into two groups: *model-based* and *model-free* methods [Descoteaux, 2015].

Model-based approaches fit the measured diffusion signal with a predefined model and estimates, for each voxel, a distribution of the fiber orientation (rather than a single one), called fiber orientation distribution function (fODF) [Dell'Acqua et al., 2007], see Figure 2.6. One of the simplest approaches of this category is called ball and stick model [Behrens et al., 2007], which assumes that molecules inside a voxel belong either to a hindered component (modelled by a stick) or to a free component (modelled by a ball). Other common model-based techniques are spherical deconvolution (SD) methods [Dell'Acqua et al., 2007, Tournier et al., 2004, Tournier et al., 2007], which assume that the diffusion signal can be described by the sum of the individual signals originating from each of the fiber populations of the sample. The different peaks of the estimated fODF can provide information about the orientation and about the weight of the fiber populations in each voxel. Variants of SD methods are constrained spherical deconvolution (CSD) methods [Dell'Acqua et al., 2007, Tournier et al., 2007], which impose a non-negativity constraint.

On the other hand, *model-free* approaches estimate the three-dimensional displacement probability profile by using a diffusion orientation distribution function (dODF) [Wedeen et al., 2005]. Estimating the dODF does not need any assumption about its form since it is derived directly from the diffusion characteristics of the fibers. Examples of these techniques include q-ball imaging (QBI) [Tuch, 2004] and diffusion spectrum imaging (DSI) [Wedeen et al., 2005].

However, also HARDI techniques have some drawbacks. One of these is the fact that, with respect to the simple DT model, both more data and sophisticated optimization schemes are required. For this reason, a higher number of diffusion-weighted images must be collected, which results in an overall increase of the acquisition time [Descoteaux, 2015]. A second limitation is that certain types of fiber configurations still cannot be discriminated. Within a voxel, fibers can curve, fan, cross, kiss, or merge,

---

[3] A review on the topic can be found for example in [Descoteaux, 2015].

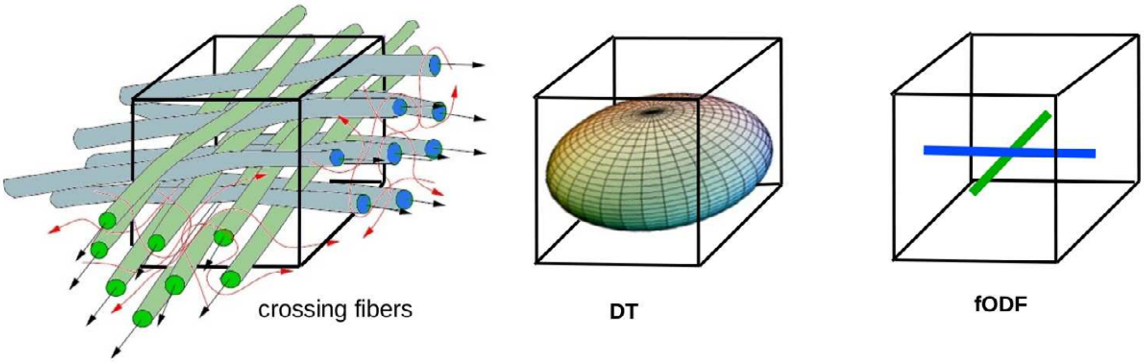


Figure 2.6: Illustration of two fiber populations within the same voxel, crossing with a 90° angle (left). The DT model cannot discriminate the two populations, providing a diffusion model with an undefined principal direction (middle). On the contrary, the fODF identifies the two spikes corresponding to the main orientation of the two fiber populations (right). Image taken from [Guevara Alvez, 2011].

and still generate a similar ODF. Therefore, many different configurations can produce the same fODF and dODF profile, which are subjected to misleading interpretations [Savadjiev et al., 2008, Jbabdi and Johansen-Berg, 2011].

2.2.3 Tractography algorithms and filtering

After the reconstruction of the diffusion signal within each voxel, one of the major subsequent applications is the 3D reconstruction of white matter fiber pathways. The computerized process of virtually reconstructing the fiber pathways is called *tractography*, or *tracking*. Tractography algorithms integrate the diffusion signal within each voxel through continuous lines known as *streamlines*. Streamlines are thus 3D polylines that approximate the trajectory of thousands of neuronal axons of the white matter. The entire set of streamlines is called a *tractogram*, see Figure 2.7 for an example.

Tractography algorithms

Tractography algorithms are based on two main assumptions: (i) the principal diffusion direction aligns with the main orientation of the fibers present within a voxel, and (ii) the continuity from voxel to voxel is anatomically legitimated [Jbabdi and Johansen-Berg, 2011]. On the basis of these assumptions, fiber pathways are reconstructed following the direction of maximum diffusion from a given voxel into a neighboring voxel, conform to some user-defined heuristics. The way tracking is performed depends on multiple parameters and on the kind of algorithm adopted. Tractography algorithms can be mainly categorized into *deterministic* and *probabilistic* tracking, and into

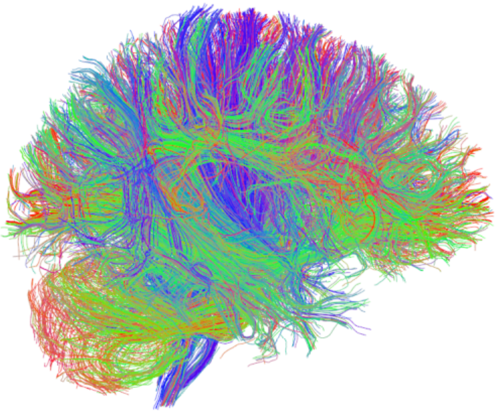


Figure 2.7: A whole brain tractogram composed of hundreds of thousands of streamlines. The color-coding corresponds to the fiber orientation (red: left–right, green: dorsal–ventral, blue: cranial–caudal).

local and *global* tracking. In what follows, we report some of the most common tractography algorithms available in the literature[4], including those adopted to prepare the material for the experimental part of this thesis.

Deterministic Tracking The first tractography algorithms were purely deterministic, in the sense that they consider only the main direction of diffusion to integrate the fiber pathways and propagate the streamlines by making a series of local optimal decisions. In general, deterministic algorithms stop propagating when a low anisotropy region, such as CSF or GM, is found, or when the angle between two consecutive steps is too wide and produces unrealistic trajectories[5].

Earlier deterministic algorithms based on the DT model include the Fiber Assignment by Continuous Tracking (FACT) [Mori et al., 1999] and the approach in [Basser et al., 2000], which uses Runge-Kutta integration. More recently, with the advent of HARDI techniques, advanced deterministic algorithms have been developed, which are able to deal with multiple orientations within a voxel. Examples of this category are the algorithm in [Descoteaux et al., 2009] and the streamline tractography based on spherical deconvolution (SDSTREAM) of [Tournier et al., 2012]. A more general and robust tractography algorithm, which combines strategies of both classical and modern approaches, was proposed in [Garyfallidis, 2013]. This algorithm, called Euler Delta Crossing (EuDX), can have as input both model-free or model-based reconstruction, and can deal with any number of crossing fibers.

---

[4] A review on the topic can be found for example in [Lazar, 2010].

[5] Common stopping criteria are FA < 0.2 and curvature $> 45°$.

A common limitation of deterministic algorithms is that the selection of the main direction of diffusion does not account for the uncertainty of its estimation process. This makes deterministic algorithms be very sensitive to the estimated main direction of diffusion and highly noise-dependent [Behrens et al., 2003].

Probabilistic Tracking Probabilistic algorithms were proposed to take into account the uncertainty of the fiber orientation estimation [Behrens et al., 2003]. For each streamline, each propagation step follows a random perturbation of the fODF estimated within that voxel. These perturbations are generated with functions that characterize the uncertainty of the fiber orientation estimation [Garyfallidis, 2013].

Probabilistic algorithms mainly differ in the manner this uncertainty is assessed. For example, in [Behrens et al., 2003] and in [Behrens et al., 2007], a Markov Chain Monte Carlo (MCMC) approach is applied for sampling from the orientation distribution. Another probabilistic algorithm to describe crossing and splitting fibers based on the fODF reconstructed from QBI was proposed in [Descoteaux et al., 2009]. However, results are not very reliable in regions with complex fiber configuration [Wassermann et al., 2010]. One of the most common probabilistic algorithm is proposed in [Tournier et al., 2010], and is based on the second-order Integration over Fiber Orientation Distributions (iFOD2). This algorithm allows a better tracking in regions with crossing fibers and generate less spurious streamlines.

Although probabilistic algorithms are more robust to noise than deterministic algorithms, a major issue is that the resulting tractograms present a large number of false positive streamlines. A tractography algorithm is said to produce false positives (FP) when it creates anatomically implausible connections, and false negatives (FN) when it fails to reconstruct pathways that truly exist.

Ensemble Tracking Every tracking algorithm requires the user to select multiple parameters, such as curvature threshold and step size. However, the optimal choice of these parameters may be different across white matter regions, individuals or acquisition settings [Takemura et al., 2016]. To overcome this issue, [Takemura et al., 2016] proposed to use ensemble methods to limit algorithm and parameter dependencies. Their tracking method, called *ensemble tracking*, produces an optimized tractogram by systematically combining candidate streamlines from multiple algorithms (deterministic and probabilistic) and systematically varying parameters (stopping criterion and curvature threshold). The resulting tractogram contains both short- and long-range streamlines, whereas single-parameter tractograms are biased towards one or the other.

Global Tracking The algorithms described above belong to the category of local tracking because, to determine the fiber trajectories, they only exploit local diffusion information. On the contrary, global tracking algorithms reconstruct the most probable trajectories in a way that path configurations are consistent with the underlying diffusion data at the global level [Lazar, 2010]. Global parameters are optimized at the path level according to some optimization criteria, for example by minimizing a cost function. Some examples of global tracking can be found in [O'Donnell et al., 2002, Jbabdi et al., 2007, Reisert et al., 2011, Christiaens et al., 2015]. Although global algorithms are less affected by local effects like noise and modeling errors, they are generally computationally intensive and require further validation [Lazar, 2010].

Which is the best tractography pipeline? It is well known that the tractography pipeline, i.e. the process of data acquisition followed by model reconstruction and tracking, is highly dependent on the subsequent applications and on the desired level of sensitivity/specificity [Thomas et al., 2014, Pestilli et al., 2014, Caiafa and Pestilli, 2017]. As an example, we can think of a clinical setting in which long acquisition times are not always feasible, thus restricting the possible choices of model reconstruction. An additional example is given by neurosurgical applications, in which it is important to perform optimal lesion excision while preserving the patient's neurological functions [Farquharson et al., 2013]. In these cases, to avoid unintentional dissection of crucial fibers during surgery, a tractography technique with high sensitivity may be preferred. On the contrary, if the goal is to reduce the possibility to identify spurious connections, a more conservative tractography method may be desirable to reduce false positives [Thomas et al., 2014].

Figure 2.8 shows the difference of two corticospinal tracts reconstructed with a different reconstruction model (DTI or CSD) and with a different tracking algorithm (deterministic or probabilistic). In general, while deterministic tractography is prone to a higher number of false negative streamlines, probabilistic tractography tends to generate more false positive connections.

Tractography filtering

As mentioned above, tractography is the only technique that can investigate the white matter structure in-vivo. Nevertheless, most of the state-of-the art algorithms create pathways into anatomically implausible regions and produce a substantial amount of false positive connections [Pestilli et al., 2014, Daducci et al., 2015, Maier-Hein et al., 2017, Jeurissen et al., 2019]. Spurious streamlines are due to multiple reasons, such as noise,

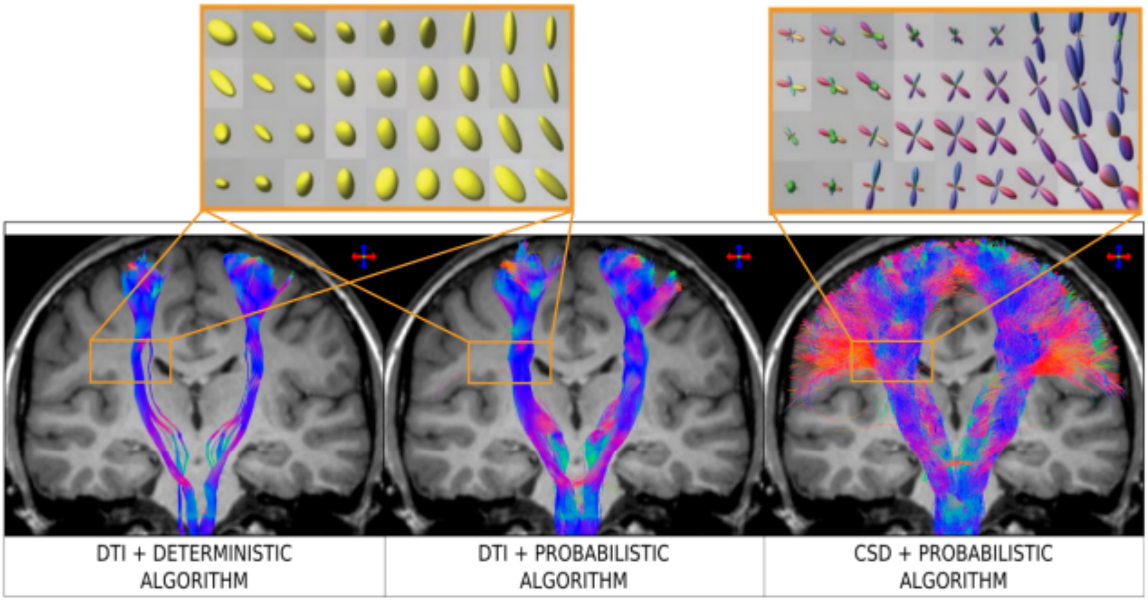


Figure 2.8: Reconstruction of the corticospinal tracts on a healthy subject using three different tractography pipelines (60 directions, b-value = 3000 s/mm^2). Coronal view with tractography results from Diffusion Tensor Imaging (DTI) reconstruction followed by a deterministic tracking algorithm (left), DTI reconstruction followed by a probabilistic tracking algorithm (middle), and constrained spherical deconvolution (CSD) followed by a probabilistic tracking algorithm (right). Image adapted from [Calamante, 2019] (CC BY license).

error propagation issues, reconstruction inaccuracy and lack of anatomical ground truth information.

To prevent streamlines to reach anatomically implausible regions, a class of algorithms that leverage anatomical priors have been developed, such as the Anatomically-Constrained Tractography (ACT) [Smith et al., 2012] and the Particle Filtering Tractography (PFT) [Girard et al., 2014].

Alternatively, given a tractogram, a post-processing step that can be performed to remove spurious streamlines is called *tractography filtering*. Although this process is still an open problem (mainly due to the absence of established criteria to filter streamlines), some promising methods to tackle the issue of false positives have been proposed. A group of these methods aims to filter tractograms through fiber evaluation, in the sense that they give each streamline a weight based on how well it explains the underlying diffusion signal. Streamlines with lower weights are considered invalid and then removed from the tractogram. Common examples of these methods are Spherical-deconvolution Filtering of Tractograms (SIFT) [Smith et al., 2013], Linear Fascicle Evaluation (LiFE) [Pestilli et al., 2014], and Convex Optimization Modeling for Microstructure Informed

Tractography (COMMIT) [Daducci et al., 2015].

Another recent tool to automatically filter out anatomically uninterpretable streamlines (AUS) was proposed in [Petit et al., 2019] and is called ExTractor. A streamline is considered as AUS if: (i) its length is lower that 20 mm, (ii) it makes a loop with an angle of 360° or more, (iii) it has a spurious shape, or (iv) it is broken with at least one endpoint within the DWM or the ventricles. In [Petit et al., 2019] they showed that ExTractor detected almost half of the total streamlines of tractograms as being anatomically non plausible.

2.2.4 Tractography registration

The task of *tractography registration*, or *normalization*, aims to spatially align tractograms of different subjects into a common neuroanatomical space to ensure spatial correspondence of anatomical structures across subjects. Spatially aligned tractograms are important to multiple purposes, such as to conduct population studies, to characterize differences in white matter morphology due to aging, or to build new atlases [O'Donnell et al., 2012, Greene et al., 2017].

Tractography registration is a particular case of medical image registration. In general, image registration is achieved by applying a geometric transformation or deformation to volumetric images. In the case of streamlines, registration techniques can be applied to both diffusion images (before tractography) or directly to streamlines [Guo et al., 2016]. In the first case, we refer to *image-based registration*, while in the second case to *streamline-based registration.*

In what follows, we report some of the most common registration algorithms available in the literature[6], including those adopted in experimental part of this thesis.

Image-based registration

Image-based registration usually begins with the alignment of diffusion-weighted images to a higher resolution T1 anatomical image. The anatomical image is then registered to a reference[7] atlas (template), such as the Montreal Neurological Institute (MNI) atlas, by using registration techniques that are generally optimized over a similarity metric, such as mutual information, cross-correlation, or mean squared difference [Greene et al., 2017]. Afterwards, there are two ways to create spatially normalized tractograms: the first is to register the underlying diffusion-weighted images to the template and then perform

---

[6] A review on the topic can be found for example in [Klein et al., 2009].

[7] The reference space can also be a subject's space.

streamline tractography, the second is to build the tractogram in the native diffusion space and then warp the streamlines to the template space [Greene et al., 2017].

Registration algorithms can be classified into *linear* and *non-linear* methods. Linear registration is a global transformation, while non-linear registration is characterized by a higher degree of elasticity that can model local deformations [Zhang et al., 2019].

Image-based linear registration Linear registration typically involves a 12-parametric affine transformation (rotation, translation, scaling, and shearing on x, y, and z axes) [Zhang et al., 2019]. The registration can be computed on scalar images, such as T1-weighted images, fractional anisotropy (FA) images, and non-weighted (b_0) images, or on diffusion tensors. One of the most commonly adopted linear registration algorithm for intra- and inter-modal brain image registration is the FMRIB's Linear Image Registration Tool (FLIRT) [Jenkinson et al., 2002], while a method based on tensor orientation is instead proposed in [Alexander et al., 2001].

Linear methods, despite having a low computational complexity, can only partially accommodate the differences between tractograms, leaving many discrepancies at the local level [Olivetti et al., 2016].

Image-based non-linear registration To-date, non-linear algorithms are the most accepted techniques within the neuroimaging community. In this cases, a deformation field is estimated and applied locally within each voxel. Deformation fields can be extracted from scalar images, diffusion tensors or fiber orientation distribution functions (fODFs). Most of the non-linear methods require however an initial linear registration to match the images globally before the non-linear one [Zhang et al., 2019].

Examples of registration methods that reorient the fOFDs are given in [Raffelt et al., 2011, Raffelt et al., 2012]. However, the strategy of reorienting the fODFs before streamline tractography has been demonstrated to be outperformed by that of applying the deformation directly to the streamline coordinates [Greene et al., 2017].

Well-known non-linear registration methods based on scalar images are the FMRIB's Non-linear Image Registration Tool (FNIRT) [Andersson et al., 2007] and the Advanced Normalization Tools (ANTs) [Avants et al., 2008]. ANTs is based on a symmetric normalization algorithm (SyN) that maximizes the cross-correlation within the space of diffeomorphic maps. In an extensive evaluation of 14 non-linear registration algorithms, SyN ranked first [Klein et al., 2009].

Full-brain image-based registration techniques may however fail in some situations, e.g. when they have to deal with severely injured brain images in which there are large areas with low contrast and intensity (due for examples to a brain tumor) [Garyfallidis et al., 2015].

Streamline-based registration

Streamline-based registration techniques aim to find a transformation or deformation by minimizing a given loss function computed only from streamlines. One important advantage of this kind of methods is that streamlines integrate information across many voxels, thus improving the alignment of white matter structures [Garyfallidis et al., 2015]. As for image-based methods, we can divide streamline-based registration techniques into *linear* and *non-linear*.

Streamline-based linear registration A recent linear method to directly transform tractograms without the intermediate step of volumetric image registration is called Streamline Linear Registration (SLR) [Garyfallidis et al., 2015]. This method, which minimizes a loss function based on distances between groups of streamlines, provided evidence to be robust to noise, missing data and important anatomical differences [Garyfallidis et al., 2015].

Nonetheless, as mentioned above, linear tractography registration is often inadequate for aligning brain structures since it cannot reconcile local differences [Klein et al., 2009, Olivetti et al., 2016].

Streamline-based non-linear registration Non-linear streamline-based methods like those proposed in [Wassermann et al., 2011, Durrleman et al., 2011], are mainly focused on bundle alignment and do not easily address whole brain tractograms [Olivetti et al., 2016].

Recently, [Olivetti et al., 2016] presented a new way to register whole brain tractograms, which is based on the idea of streamline correspondence and is formulated as a Graph Matching (GM) problem. In this case, the main limitation is that the algorithm used for the optimization only provides an approximate solution.

Streamline-based registration methods have in common one basic building block: the definition of a streamline-streamline distance function. In general, streamline distances are needed to quantify the relative displacements between streamlines and to establish their degree of similarity. The idea is that streamlines with a similar shape lie at small distances, while streamlines with a less similar shape lie at greater distances. Streamline distances are a key element not only for tractography registration, but also to multiple other applications, such as clustering and bundle segmentation. We will provide further details about streamline distance functions in Section 3.1 and in Section 5.1.1.

2.3 White matter bundle segmentation

The previous Sections of this Chapter were devoted to the tractography pipeline to build a tractogram starting from raw diffusion data. It is worth noting that in all the contributions of the present thesis, the tractograms are considered our input data. None of our contributions was meant to ameliorate any of the steps of the tractography pipeline, nor of tractography filtering, nor of tractography registration. The main focus of the present thesis concerns a common task that can be performed once having a tractogram, which is called *white matter bundle segmentation*[8].

In many applications, streamlines from a whole brain tractogram are not immediately useful for scientific or clinical purposes [O'Donnell and Westin, 2007]. The task of white matter bundle segmentation aims is to virtually group together fibers sharing a similar pathway into anatomically meaningful structures, called white matter bundles or tracts. In this context, we refer to a bundle as a group of streamlines representing the trajectories of hundreds of thousands of neuronal axons. Figure 2.9 shows an illustration of the concept of segmentation of a bundle starting from a whole brain tractogram.

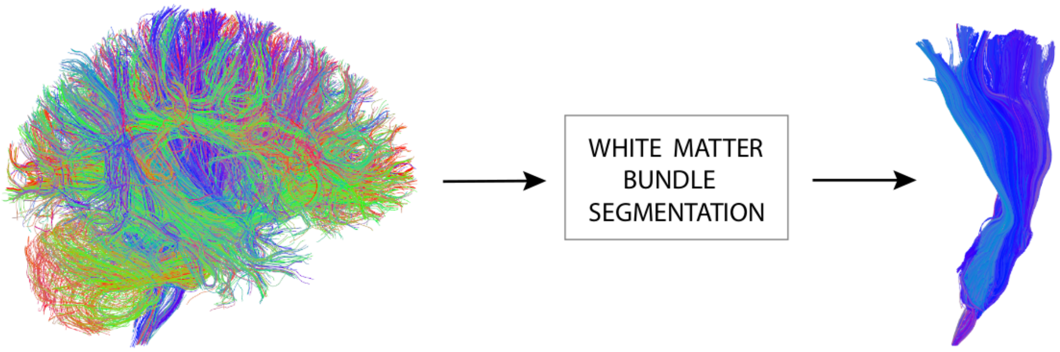


Figure 2.9: Representation of the concept of segmentation of a bundle (the corticospinal tract) from a whole brain tractogram. Streamlines composing the bundle are a subset of those composing the tractogram.

In Section 2.1.2, we described the anatomy of some kinds of white matter bundles, included those adopted in the experimental part of the present thesis. This Section is

---

[8] As we will discuss in Chapther 3, the bundle segmentation task can be performed also from other input data, such as a specimen or raw diffusion data.

instead devoted to some concepts that are strictly related to the task of virtual bundle segmentation, i.e. validation and applications.

2.3.1 Validation

Due to the high dependency of streamline tractography on the quality of data, reconstruction models, and user-defined parameters, it is important that tractograms and bundles undergo through a validation process. Throughout the last few decades, many attempts have been performed to assess the accuracy of tractography techniques. Examples of these methods involve animal studies, *post-mortem* human studies, physically or numerically simulated phantoms and computational techniques.

Post-mortem dissection based on the Klingler's method [Klingler et al., 1935], although it permitted to validate the pathways of several white matter bundles [Sarubbo et al., 2015, De Benedictis et al., 2016], it is limited by the fact that crossing fibers are not recoverable and that it is only applicable to ex-vivo data.

A computational strategy to validate fibers in-vivo can be found for example in [Pestilli et al., 2014]. This method can attest the strength of evidence supporting the existence of set of fibers with a virtual lesion approach. A major issue of this class of methods is the lack of a golden standard to compare the results with [Maier-Hein et al., 2017].

As an alternative, physical [Poupon et al., 2008, Bach et al., 2014] and synthetic [Close et al., 2009, Neher et al., 2014, Caruyer, 2014] phantoms have been proposed. Synthetic phantoms are simulation tools to define artificial white matter fibers and to generate corresponding diffusion-weighted images. The tool proposed in [Neher et al., 2014], called Fiberfox, was used to generate the ground truth dataset for the ISMRM 2015 Tractography Challenge (`tractometer.org/ismrm_2015_challenge`), in which results were evaluated using the evaluation tool called Tractometer of [Côté et al., 2013]. However, building realistic and complex phantoms is challenging, and thus only basic configurations of fibers can be validated in this way [Maier-Hein et al., 2017].

All the above mentioned methods suffer from common a major limitation: the lack of reliable anatomical ground truth information [Maier-Hein et al., 2017]. For this reason, tractography validation still remains an open problem.

2.3.2 Applications

Segmented white matter bundles are of primary importance for a large number of applications, both in the scientific field and in the medical field. In what follows, we briefly present three of the most common ones.

Neurosurgical planning

Nowadays, the adoption of tractography techniques has become a standard practice also within clinical settings. Segmented bundles in the area of interest can provide further insights into the anatomical structure under examination that classical MRI scans cannot supply [De Benedictis et al., 2018]. For example, segmented bundles may serve in neurosurgical planning to drive decisions about resection boundaries in case of tumors or epilepsy. It is of primary importance in fact maximizing resection while minimizing the damage to the patient's neurological function [Wu et al., 2007, Farquharson et al., 2013].

Tractometry

From segmented white matter bundles it is possible to extract quantitative measures computed at the voxel level, such as the Fractional Anisotropy (FA). This approach, usually known as *tractometry*, can provide a comprehensive assessment of bundle-specific microstructural measurements [Bells et al., 2011, Jones and Nilsson, 2015]. Such measurements are particularly useful, e.g., in group studies, to assess developmental changes of the white matter, or to correlate tractography results with clinical or behavioral scores [Forkel et al., 2014].

Examples of tractometry studies are given in [Colby et al., 2012, Yeatman et al., 2012, Yeatman et al., 2018], in which they proposed to represent a bundle by a vector of diffusion measurements like FA values (FA profile) computed along its main direction. In [Yeatman et al., 2012], FA profiles of various bundles are compared across different subject populations to assess developmental changes, see Figure 2.10 for an example.

Brain connectomics

The term *brain connectomics* refers to the study of the brain connectivity, or connectome. Segmented bundles can be of great help in understanding both the structural and the functional connectivity of the human brain [Sporns et al., 2005, Bullmore and Sporns, 2009]. For example, in [Bullmore and Sporns, 2009], observations of bundles connecting components of the default network and a direct comparisons of structural and functional connectivity in the same group of subjects revealed that structural connections highly predict functional connections.

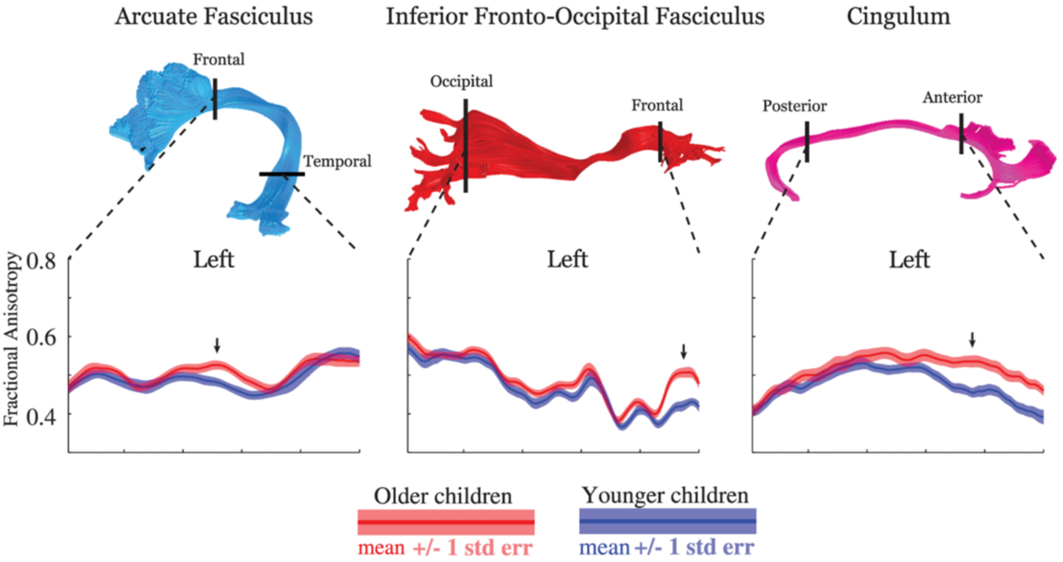


Figure 2.10: Mean Fractional Anisotropy (FA) profiles of the left arcuate fasciculus, inferior fronto-occipital fasciculus and cingulum bundle in younger children (blue), and older typically developing children (red). Arrows indicate the area of the FA profile showing the highest group difference. Image adapted from [Yeatman et al., 2012] (CC BY license).

2.4 Basic principles of Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI), the field driving the development of machines able to perform intelligent and cognitive tasks simulating humans. Specifically, ML investigates how machines (computers) can improve actions, predictions and decisions based on data and experience [Vu et al., 2018]. Other branches of AI include for example reasoning and knowledge representation, natural language processing, and perception.

The term *learning* refers to the process of inferring a general rule by observing examples. To this end, particular examples related to a specific task are given to a machine. Its goal is then to derive a general rule that can both explain the examples it has already seen and generalize to new unseen examples [von Luxburg and Schoelkopf, 2008]. This general rule is typically inferred from *features* of the example input data. Features are transformations of input data that facilitate a learning algorithm to return correct outcomes on new examples, see Figure 2.11. Choosing informative and independent features is of primary importance towards the correct functioning of learning algorithms [Stevens and Antigua, 2019].

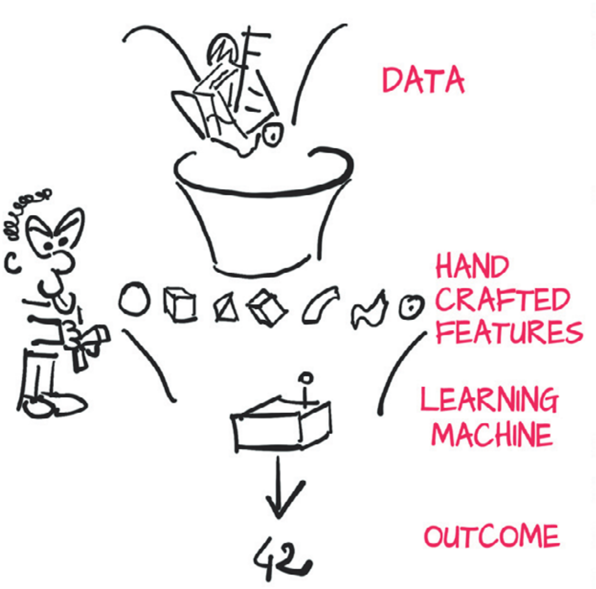


Figure 2.11: A learning machine takes as input features of the input data and returns as output an outcome that fulfills the inferred general rule. Image taken from [Stevens and Antigua, 2019].

2.4.1 Unsupervised vs Supervised Learning

Learning problems can generally be divided into two categories: *supervised learning* and *unsupervised learning*. These two categories differ with respect to their inputs, outputs, and final objective [Vu et al., 2018].

Unsupervised learning techniques aim to discover a 'structure' in the given data without any further information about the kind of output [von Luxburg and Schoelkopf, 2008]. Hidden structures in the data can be found by taking advantage of similarity or redundancy among objects. Classical examples of this category are given by clustering and Principal Component Analysis (PCA). Clearly, a main challenge of this kind of algorithms is to understand the features that can better discriminate between classes [Vu et al., 2018].

On the other hand, *supervised learning* techniques aim to find a relation between the given data and the kind of possible output using some prior knowledge [von Luxburg and Schoelkopf, 2008]. The goal of supervised methods is to categorize unknown (unlabeled) instances, leveraging multiple examples previously labeled, i.e. whose class is known. In this scenario, data are usually split into a *training set* and a *test set*. The training set is composed of examples to learn from in order to develop a model. When the model predictions reach the desired level of accuracy, the model is test on the remaining set, known as test set. Common examples of supervised algorithms include regression and classification.

In the present thesis, all the methodological contributions that will be presented in Chapter 5 fall in the category of **supervised learning**.

2.4.2 Classification

Classification is one of the most known supervised learning task that aims to categorize new observations into predefined classes[9].

In this case, we deal with two kind of spaces: the input space X (feature space) and the output space Y (label space). To accomplish the classification task, the learning algorithm is given some examples representing pairs of objects with the corresponding class label, i.e. $(x_1, y_1), \ldots, (x_n, y_n)$. The aim is then to establish a mapping $f : X \mapsto Y$ that minimizes the number of errors, i.e. that makes the number of misclassified objects as small as possible. The mapping $f : X \mapsto Y$ is called a *classifier* [von Luxburg and Schoelkopf, 2008].

To minimize as much as possible the error between predicted and measured outputs, it is necessary to define an error measure. Such measure, known as *loss function* or *cost function*, computes a single numerical value that the learning process attempts to minimize. Ideally, to obtain a high model accuracy, this value should be as low as possible [Stevens and Antigua, 2019].

Linear classification

Linear classification algorithms make a classification decision based on the value of a linear combination of the features. Common linear classification algorithms are Linear Discriminant Analysis, Logistic Regression, and Support Vector Machines (SVMs) with linear kernels, see for example [Hastie et al., 2017, Bishop, 2013].

Linear models are the simplest reasonable models to adopt to fit data. In the absence of further knowledge, linear models can be a successful choice [Stevens and Antigua, 2019]. Morever, they are extremely well-known and easy to understand, they require minimal resources and time to be trained, and they can be interpreted.

Non-linear classification

As opposed to linear classification algorithms, non-linear classification algorithms make a classification decision based on the value of a non-linear combination of the features. Common non-linear classification algorithms are k-Nearest Neighbor (k-NN), Support Vector

---

[9] If the number of classes is two, the classification is known as binary classification.

Machines (SVMs) with non-linear kernels, and convolutional neural networks (CNN), see for example [Hastie et al., 2017, Bishop, 2013].

Although non-linear models are in general a better choice when the classes are non-linearly separable, they usually require a long training time and can hardly be interpreted.

2.4.3 Lazy learning methods

Lazy learning methods are ML methods in which the processing of training data is delayed until a query is made to the system [Aha, 2013]. On the contrary, *eager* methods process training data before receiving queries.

Lazy learning methods are often employed in situations in which the dataset is continuously updated with new entries, as for online recommendation systems. When a query is made, the learning algorithm searches for those examples that are considered the most relevant to answer the query according to some similarity measure. These examples are used to derive a local description of the system and finally to answer the query. The answer and any intermediate results are then discarded and, for each of the following queries, the full procedure has to start again. For this reason, these kind of methods do not really have a training phase.

Examples of lazy learning algorithms include k-Nearest Neighbor (k-NN) and optimization through the Linear Assignment Problem (LAP).

Chapter 3

Related work

In this Chapter, related work about the research topics of the present thesis are reported. In Section 3.1, first, we report related work about streamline distance functions and their comparison in the literature, and then, we present state-of-the-art streamline vectorial representation techniques. Section 3.2 is aimed at revising state-of-the-art techniques commonly employed to address bundle segmentation, with a particular focus on those adopted or compared in the experimental part of the thesis. Additionally, we will discuss advantages and disadvantages of each of them, a kind of analysis that was fundamental to make grow our research interest and to drive our methodological research on this topic.

3.1 Streamline distances and vectorial representation

First, in Section 3.1.1, we report related work about streamline distance functions and their comparison in the literature (the mathematical formulation to obtain the most frequently used distance functions is given in Section 5.1.1). Then, in Section 3.1.2, we present state-of-the-art streamline vectorial representation techniques.

3.1.1 Streamline distances

As already mentioned in Section 2.2.4, bundle segmentation algorithms need a similarity measure to establish a degree of similarity between objects. To quantify shape similarity and location closeness between a pair of streamlines, the similarity measure (or more precisely *dis*similarity measure) usually adopted is the streamline-streamline (pairwise) distance function. When two streamlines are similar in shape and close in location, their distance is small, while when their shapes are different or their displacement is large, their distance is large [Ding et al., 2003].

Numerous streamline distance functions have been proposed in the literature, see for example [Alt and Godau, 1995, Ding et al., 2003, Corouge et al., 2004, Zhang et al., 2008, Garyfallidis et al., 2012, Charon and Trouvé, 2013]. The most common distance functions are applied on streamlines parametrized as sequences of 3D points[1].

In [Corouge et al., 2004], they proposed three different distance functions: the closest point distance, the mean of closest distances (d_{MC}), and the Hausdorff distance. Additionally, they used shape-based distances based on the geometric characteristics of the streamlines, such as length, center of mass, and second order moments. In [Zhang et al., 2008], they implemented two variations of the mean of closest distances, namely the shorter mean of closest distances (d_{SC}) and the longer mean of closest distances (d_{LC}). Among the distances just mentioned, one of the most commonly adopted is the Hausdorff distance. However, despite being a useful metric to detect outliers [Corouge et al., 2004], it does not account for the sequential nature of streamlines and it is slow to compute [Garyfallidis et al., 2012]. For these reasons, [Garyfallidis et al., 2012] defined a novel distance function called minimum average direct flip distance (d_{MDF}), which takes the minimum between the direct and the flipped distances. This distance is very fast to compute and it is a metric on the space of streamlines. However, it requires streamlines to be resampled at the same number of points, see Figure 3.1 for a graphical intuition. More recently, other streamline distances have been proposed, such as the Point Density Model (d_{PDM}) [Siless et al., 2013] and the varifolds distance ($d_{varifolds}$) [Charon and Trouvé, 2013].

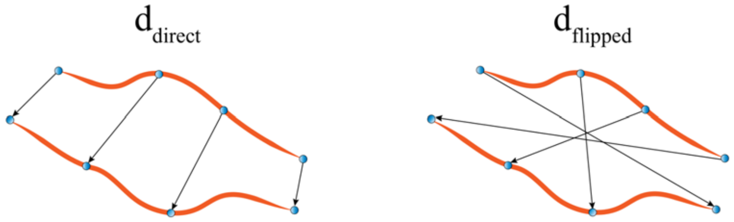


Figure 3.1: Illustration of direct and flipped distances between two streamlines with the same number of points. The direct distance (d_{direct}) is computed by averaging the individual Euclidean direct distances, while the flipped distance ($d_{flipped}$) by averaging the individual Euclidean flipped distances.

Although each streamline distance has its distinct technical motivation, little has been said to guide the selection of a distance for a specific task. To the best of our knowledge, comparison and evaluation of distances were performed only in the case of *unsupervised* bundle segmentation, i.e. by means of clustering. In [Moberts et al., 2005], four different

---

[1] Other streamline parametrizations exist, such as B-splines [Corouge et al., 2006] or Fourier descriptors [Batchelor et al., 2006].

distances were compared to see the impact on various indexes for clustering of streamlines, concluding that the mean of closest distances performs better than the Hausdorff distance, the closest point distance, and the end point distance. In [Siless et al., 2013], they compared the k-means clustering with three different distances, i.e. the d_{PDM} distance, the Hausdorff distance, and the Undirected Euclidean (EU) distance, suggesting that the d_{PDM} distance should be preferred over the others. In [Alexandroni et al., 2016], they evaluated seven distance metrics, attesting that the cosine distance gives the best performance among the clustering based methods.

On the other hand, investigations about the most suitable streamline distance function to perform *supervised* bundle segmentation have not been conducted yet.

3.1.2 Streamline vectorial representation

To work efficiently, most classification algorithms require data to be in the form of vectors, which is not the case of streamlines, since they are 3D polylines composed of different number of points. To this end, several embedding techniques have been adopted to transform streamlines into multidimensional vectors. In the embedding space, each streamline is represented as a point, and thus points close to each other correspond to streamlines following similar pathways [O'Donnell and Westin, 2007], see Figure 3.2 for a toy example.

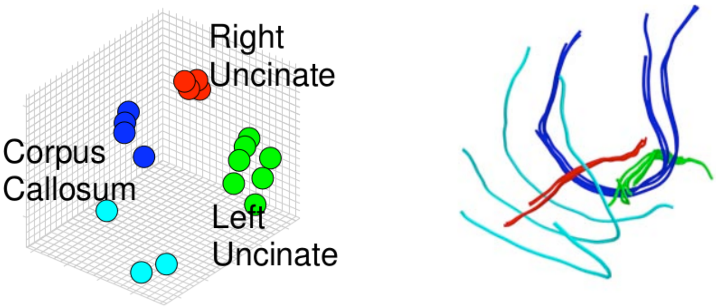


Figure 3.2: Toy example of streamline embedding. Streamlines in their original space (right) are embedded as multidimensional points in the embedding space (left). Points close to each other (same color) correspond to streamlines belonging to the same kind of bundle. Image taken from [O'Donnell and Westin, 2007].

Common embedding techniques for streamlines are: (i) streamline 'flattening', (ii) shape descriptor vector, (iii) 9D descriptor vector, (iv) spectral embedding, and (v) dissimilarity representation.

Streamline ’flattening’

The easiest and fastest way to represent streamlines into vectors is by ’flattening’ their 3D coordinates. First, all streamlines are resampled at p equally spaced points. Then, each streamline is transformed into a vector of length $3p$ by appending the sequence of its 3D coordinates, as it was done for example in [Mayer et al., 2011].

In principle, resampling to a specific number of points may cause loss of information. Moreover, another major issue is given by the streamline orientation. A streamline with p points can equivalently have its points ordered from 1 to p or be flipped with order from p to 1. This embedding technique will produce a different vectorial representation depending on the order of the points, which is not desirable as the streamline in the two cases is identical.

Shape descriptor vector

An alternative way to describe a streamline as a vector is to use shape descriptors as in [Ding et al., 2003, Corouge et al., 2004, Batchelor et al., 2006]. Their approaches make use of fundamental geometric invariants, such as curvature and torsion, which are derived from the Frenet-Serret framework, see Equation 5.21. After a resampling of the streamline at the same number of points, curvature and torsion are computed along each point of the streamline.

Curvature and torsion are rigid body invariants, i.e. they do not change under rotation and translation, and due to their differential nature, they are very noise-dependent. To provide meaningful results, this embedding strategy thus requires some smoothing of the data, and moreover, as streamline ’flattening’, it is sensitive to streamline orientation [Batchelor et al., 2006].

9D descriptor vector

To preserve in the mapping of a streamline to a vector important properties like position, shape and connectivity, [Brun et al., 2004] introduced a 9D descriptor vector. This vector is composed of the mean vector (3 values) and the lower triangular part of the root-squared covariance matrix (6 values) of the points of the streamline.

For applications in which mean and covariance are not enough to discriminate between streamlines, higher order central moments can be added to the vector. Moreover, if connectivity is more important than shape, a higher weight can be given to the streamline endpoints in the computation of the mean vector and covariance matrix [Brun et al., 2004].

Spectral embedding

Spectral embedding of streamlines is an embedding technique firstly proposed in [O'Donnell and Westin, 2007] for clustering, and then adopted in [Vercruysse et al., 2014] for supervised learning. First, streamlines of multiple subjects are registered into a common coordinate system and a pairwise distance matrix is obtained by computing the mean closest point distances between all pairs of streamlines. Then, the distance matrix is converted to an affinity matrix via a Gaussian kernel. Due to the large size of the affinity matrix, they used an approximation called the Nystrom method [Fowlkes et al., 2004], which requires only a subset of all the streamlines. Finally, the spectral embedding vectors are computed using the eigenvectors of the normalized affinity matrix, ordered in descending order according to the eigenvalue [O'Donnell and Westin, 2007].

The overall computation can be very time demanding if a large number of streamlines and subjects are used.

Dissimilarity representation

An alternative way to embed a streamline into a vectorial space is to compute its dissimilarity representation, as originally presented in [Olivetti et al., 2012], and later employed also in [Porro-Muñoz et al., 2015, Sharmin et al., 2016, Sharmin et al., 2018]. The general concept of dissimilarity representation of objects was initially proposed for pattern recognition problems, see for example [Pekalska and Duin, 2005]. This embedding technique describes each streamline with a set of distances from other streamlines, known as *landmarks*, or *prototypes*[2]. Landmarks are usually taken from a whole brain tractogram with specific landmark selection policies.

The accuracy and effectiveness of this type of Euclidean embedding is demonstrated in [Olivetti et al., 2012], where they provided evidence that the geometrical information of streamlines is preserved. Moreover, they showed that the selection policy called Subset Farthest First (SFF) outperforms other policies of selection, and that a high accuracy can be reached using only a few tens of landmarks (approximately 20 - 25). However, as for spectral embedding, the procedure can be computationally intensive if a large number of streamlines and landmarks are adopted.

It is worth noting that the choice of the embedding technique is task-related: one can be more suitable for clustering, while another for bundle segmentation.

---

[2] A more formal definition of the dissimilarity representation is provided in Section 5.1.2.

3.2 White matter bundle segmentation techniques

A substantial body of literature is devoted to white matter bundle segmentation techniques. This Section is aimed at revising state-of-the-art techniques commonly employed to address bundle segmentation, with a particular focus on those adopted or compared in the experimental part of the thesis. Additionally, we will discuss advantages and disadvantages of each of them, a kind of analysis that was fundamental to make grow our research interest and to drive our methodological research on this topic.

Historically, white matter bundle segmentation was performed manually by expert neuroanatomists. In Section 3.2.1 we review the most common manual segmentation approaches. More recently, to overcome the limitations of manual approaches, several automatic bundle segmentation techniques have been developed. These methods can be grouped into three main categories: (i) connectivity-based methods, (ii) streamline-based methods, and (iii) direct methods, which are presented in Sections 3.2.2, 3.2.3 and 3.2.4, respectively.

3.2.1 Manual bundle segmentation approaches

We categorize manual segmentation approaches into three groups: (i) ex-vivo, (ii) Region Of Interest (ROI)-based, and (iii) bundle-based approaches.

Ex-vivo approach

Earliest studies of human white matter dissection[3] were conducted in the 19th century, before the advent of neuroimaging techniques. In that period, the human brain anatomy could be investigated only after death, i.e. *ex-vivo*. Despite the impressive quality of the topographical representations of the pioneering neuroanatomical studies of that time, a comprehensive description of the human brain white matter architecture was still lacking, mainly due to methodological issues in dealing with specimens [Zemmoura et al., 2016].

Recently, renew interest has been expressed towards micro-dissection studies, see for example [Sarubbo et al., 2015, Sarubbo et al., 2016, De Benedictis et al., 2016, De Benedictis et al., 2018]. These studies performed brain dissection with the Klingler's freezing-defrosting technique, which was firstly introduced by the seminal work of Joseph Klingler (1888-1963) [Klingler et al., 1935]. This technique consists of fixing the specimen in a 10% formalin solution for 40 days and then freeze it for 30 days at -80°C. The wa-

---

[3] The term dissection refers to a special anatomical technique to reveal white matter subcortical organization of human brain cadavers (specimens).

ter crystallization induced by freezing disrupts the structure of the gray matter, making possible to peel off the cortex from the brain surface without damaging the underlying subcortical white matter organization. Additionally, the freezing process induces a clear separation between white matter fibers, facilitating the dissection by progressive fiber removal usually with the help of wooden spatulas [Zemmoura et al., 2016]. The right side of Figure 3.3 shows an example of the result of ex-vivo dissection of an arcuate fascicle.

Although micro-dissection studies have greatly advanced the understanding of human brain white matter architecture, the procedure is complex and has to be performed by expert neuroanatomists. Moreover, due to its destructive nature, it allows the expert to analyze only very few bundles at the time, and obtaining a digital model of the dissection is not trivial [De Benedictis et al., 2018].

ROI-based approach

The development of tractography techniques and specific software tools have allowed practitioners to mimic ex-vivo dissection through ROI-based (in-vivo) approaches, see Figure 3.3. For this reason, the ROI-based approach is also called virtual dissection.

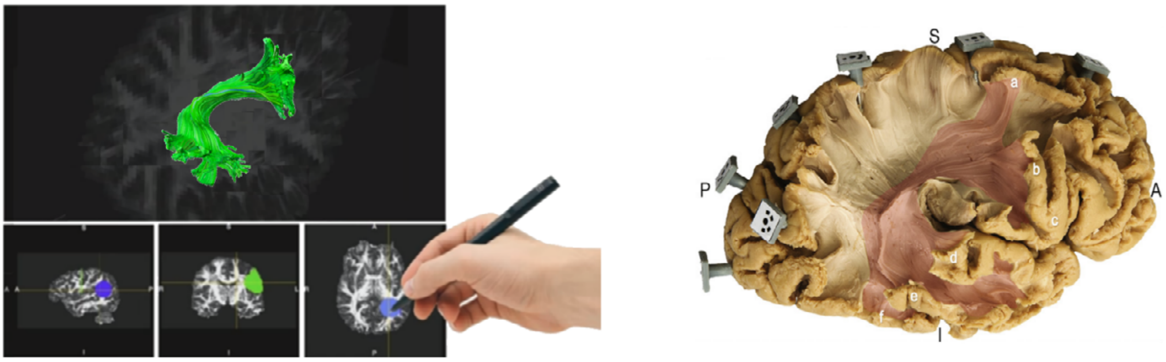


Figure 3.3: ROI-based manual segmentation (virtual dissection) of an arcuate fascicle with the support of a software tool (left) to mimic ex-vivo dissection (right). Image taken from [De Benedictis et al., 2018].

Early works about virtual dissection of well-characterized bundle on tractography data were proposed in [Catani et al., 2002, Mori et al., 2005, Wakana et al., 2007, Thiebaut de Schotten et al., 2011]. The general approach consists of defining one or more ROIs that define the bundle of interest, typically with the aid of software tools, see left side of Figure 3.3 for an example. Common software tools usually adopted are TrackVis (`http://trackvis.org`) [Wang et al., 2007] and MRtrix (`http://mrtrix.org`) [Tournier et al., 2012]. First, the ROIs are manually drawn in the brain space by a trained expert, usually by leveraging anatomical information contained in structural images. These ROIs can be regions where the bundle starts, ends or passes through. Then, only streamlines of the tractogram that pass through them are selected to form the segmented bundle. In the case of multi-ROIs approaches, the ROIs can also be used to exclude streamlines that do not belong to the bundle [Wakana et al., 2007].

In [Wakana et al., 2007], they proposed to segment 20 major bundles using two planar waypoint ROIs per bundle that delineate their trajectories before they diverge towards the cortex. This collection of ROIs, drawn on a group-average DTI dataset in MNI space, is available at `https://github.com/vistalab/vistasoft/tree/master/mrDiffusion/templates/MNI_JHU_tracts_ROIs`.

More recently, a novel ROI-based approach to virtually dissect the inferior fronto-occipital fascicle and the uncinate fascicle was proposed in [Hau et al., 2016, Hau et al., 2017]. This approach is known as *stem-based* approach since, as an anatomical reference to identify a specific bundle, it makes use of a particular ROI call the *stem*. First, inspired from ex-vivo dissection techniques, the tractogram is prepared by selectively removing streamlines not belonging to long association pathways using a combinations of rules and template ROIs. Then, the stem, which is the point of passage where all fibers pass through, is drawn on a single planar slice. Finally, the desired bundle is virtually extracted by retaining only the streamlines passing through the stem, mimicking classical ex-vivo dissection.

ROI-based approaches have some important limitations. First, if there are streamlines that are broken in one or more points of their pathway due to an incorrect reconstruction, they will be completely filtered out, causing an underestimation of the bundle size. Second, the manual interaction of the user (with the aid of software tools) with sets of streamlines that are densely packed together is challenging and indirect [Porro-Muñoz et al., 2015]. Third, the segmentation is inherently limited by the quality of registration between the tractogram and the underlying structural image.

Bundle-based approach

The bundle-based approach proposed in [Porro-Muñoz et al., 2015] for manual segmentation is based on direct interaction of the user with the streamlines, as opposed to the indirect interaction of ROI-based method. The user can interact with a summary of the whole brain tractogram, which is obtained by clustering streamlines into a predefined number of clusters (usually in the order of tens or one hundred). The task of manual segmentation is then achieved with an iterative process in which the user alternates a phase of clusters selection to a phase of re-clustering to incrementally reveal details of the bundle of interest. This process is repeated until the remaining streamlines faithfully represent the bundle of interest. This algorithm is implemented in a software tool for visualization, exploration and segmentation of tractography data called Tractome (`http://tractome.org`) [Porro-Muñoz et al., 2015].

Major limitations in common to all the manual segmentation approaches are that they require experts with extensive neuroanatomical knowledge, and that the process is very time consuming. Moreover, manual approaches are severely affected by human subjectivity and are hardly reproducible.

To overcome these limitations, throughout the last few decades a substantial amount of automatic techniques for bundle segmentation have been developed. Generally, performing a task automatically has a series of benefits, including less dependency to biases and human errors, and more objective, consistent and reproducible results [Vu et al., 2018].

3.2.2 Connectivity-based approaches

Connectivity-based approaches, which can be seen as the automated version of ROI-based manual approaches, aim to extract white matter bundles by filtering the entire set of streamlines with inclusion/exclusion ROIs that the bundle is assumed to pass/not to pass through, on the basis of some pre-defined rules, see for example [Catani and De Schotten, 2008, Oishi et al., 2008, Zhang et al., 2010, Yeatman et al., 2012, Wassermann et al., 2016]. ROIs can be both in the white matter or in the gray matter (cortical ROIs) and frequently come from atlases or parcellations that have to be registered into the individual subject's space. Two of the most common techniques belonging to this group are Automatic Fiber Quantification (AFQ) [Yeatman et al., 2012] and White Matter Query Language (WMQL) [Wassermann et al., 2013, Wassermann et al., 2016].

The AFQ algorithm of [Yeatman et al., 2012] can segment 20 major white matter bundles with a procedure composed of the following three steps, which are illustrated in Figure 3.4. 1. A candidate bundle is identified with the waypoint ROI-based approach described in [Wakana et al., 2007], in which streamlines are assigned to a particular bundle if they pass through two waypoint ROIs that define its trajectory[4]. 2. A bundle refinement is performed by comparing each candidate streamline to a fiber probability map, in which at each voxel is assigned a probability to belong to certain bundles [Hua et al., 2008]. 3. A final bundle cleaning is accomplished by removing streamlines that are more than

---

[4] Each waypoint ROI was warped into the individual subject's space with a non-linear registration.

4 standard deviations above the mean streamline length or that deviate more than 5 standard deviations from the core of the bundle.

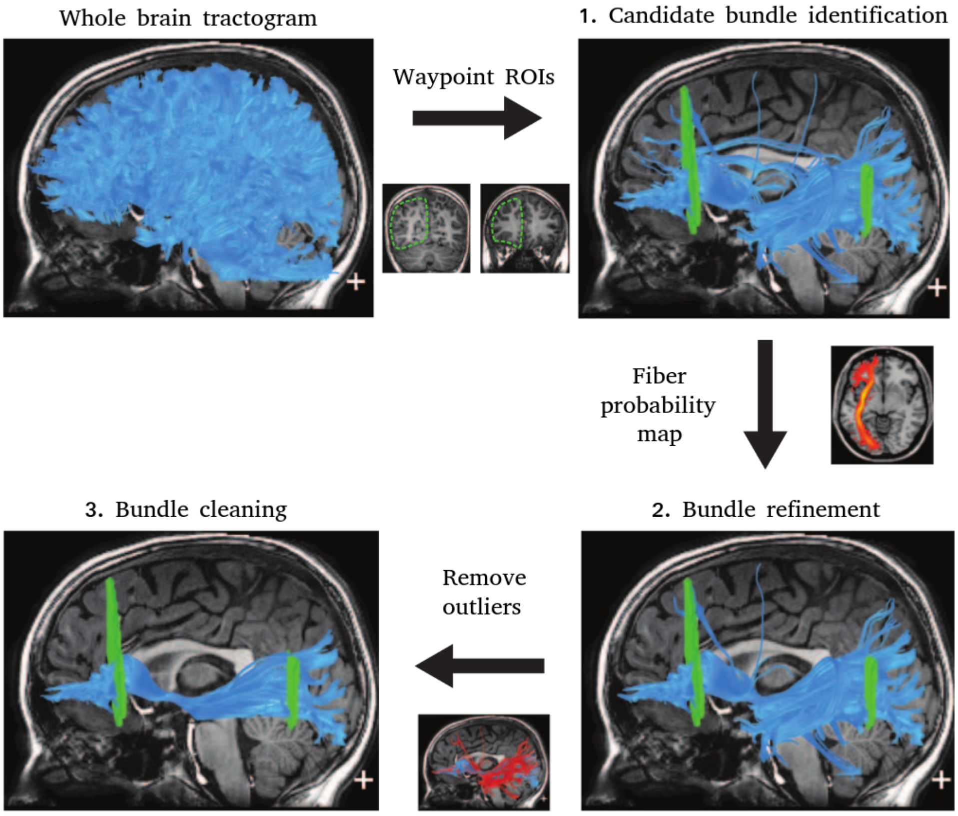


Figure 3.4: Automated Fiber Quantification (AFQ) workflow to segment the left inferior fronto-occipital fasciculus. 1. Candidate bundle identification with the waypoint ROI-based approach of [Wakana et al., 2007]. 2. Bundle refinement with the fiber probability map. 3. Bundle cleaning to remove outliers. Image adapted from [Yeatman et al., 2012] (CC BY license).

The WMQL technique of [Wassermann et al., 2013, Wassermann et al., 2016] formally describes a set of white matter bundles and automatically segment them from tractography data. The query language permits to build a dictionary of definitions, which consists of adjacent gray and white matter ROIs, and rules for spatial relations, to describe the anatomy of white matter bundles. This enables the encoding of anatomical white matter knowledge and coherent labeling of bundles across subjects.

A more recent connectivity-based approach similar to WMQL is presented in [Bullock et al., 2019], which uses multiple cortical ROIs per bundle to derive the segmentation. Typically, this begins with confining the segmentation to the type of bundle it is (i.e. a fronto-parietal bundle). To better refine the segmentation, this approach also incorporates a number of additional inclusion or exclusion criteria based on other anatomical features, such as subcortical structures.

Connectivity-based techniques like WMQL have the advantage to be very versatile, in the sense that the rules can be customized by the user. However, all connectivity-based approaches are inherently limited by the anatomical variability of the subjects, by the quality of the atlas or parcellation, and by the process of registration. Moreover, the heuristic rules derived from a small sample of the data may not generalize well on data with diverse characteristics.

3.2.3 Streamline-based methods

Streamline-based methods perform bundle segmentation by grouping together streamlines according to some similarity measure. A further subdivision consists in discerning between (i) clustering methods, (ii) atlas-based methods, (iii) model-based methods, (iv) classification-based methods, and (iv) correspondence-based methods.

Clustering methods

Clustering methods for white matter bundle segmentation, such as those in [Brun et al., 2004, Moberts et al., 2005, O'Donnell et al., 2006, Zhang et al., 2008, Siless et al., 2016, Siless et al., 2018], perform whole brain segmentation without prior knowledge about the anatomy of the bundles and without leveraging examples of expert-based segmented bundles. The clustering result depends on the choice of two fundamental elements: the similarity measure and the clustering method [O'Donnell et al., 2013].

Typical similarity measures between streamlines are pairwise distance functions [Moberts et al., 2005, Siless et al., 2013] or Euclidean distances in a multidimensional embedding space [Brun et al., 2004, O'Donnell et al., 2006]. Common strategies include spectral clustering [Brun et al., 2004, O'Donnell et al., 2006] or hierarchical clustering [Moberts et al., 2005, Zhang et al., 2008, Siless et al., 2018]. All these methods usually require the computation of large distance matrices and thus suffer from high computational complexity.

In [Garyfallidis et al., 2012], they proposed a completely unsupervised technique, called QuickBundles, which efficiently clusters streamlines according to a distance similarity measure. They however assumed that a fully unsupervised technique is not able to create anatomically meaningful bundles without considering extensive prior anatomical knowledge. In fact, meaningful bundles differ consistently both in shape and in length, and no single threshold is suitable to cluster a full dataset [Garyfallidis et al., 2012]. As a consequence, they suggest adopting QuickBundles rather for dimensionality reduction and tractography simplification. Most of the clustering methods explicitly targeted to bundle segmentation are in fact semi-automatic, since the clustering is usually followed by expert manual labeling.

The main challenges of clustering techniques for bundle segmentation are threefold: (i) handling large datasets, (ii) choosing the number of clusters or the threshold that decides when to stop merging or splitting clusters, and (ii) incorporating anatomical knowledge into the clustering result.

Atlas-based methods

Atlas-based methods, such as those in [Maddah et al., 2005, O'Donnell and Westin, 2007, Zhang et al., 2010, Guevara et al., 2012, Tunç et al., 2014, Labra et al., 2016], rely on the information stored in atlases. Atlases can either be composed of multiple anatomical ROIs [Maddah et al., 2005, Zhang et al., 2010] or by streamline clusters [O'Donnell and Westin, 2007, Guevara et al., 2012, Labra et al., 2016], which are typically the result of a clustering strategy on multiple tractography dataset followed by expert anatomical labeling. Usually, given a tractogram, all the streamlines are first, projected to the atlas, and then labeled as the most similar structure found in the atlas [Maddah et al., 2005].

Common limitations of atlas-based methods are that they are affected by the quality of the atlas, by the accuracy of the projection of the streamlines to the atlas, and by the number of subjects used to construct the atlas [Guevara et al., 2012].

Model-based methods

Model-based methods for bundle segmentation aim to learn relations between streamlines and bundles by assuming prior structures of the data, given in the form of models. Several different prior models have been proposed in the literature, such as beta mixture model [Clayden et al., 2007], gamma mixture model [Maddah et al., 2008], Gaussian process of tract probability map [Wassermann et al., 2010], cubic spline with a fixed number of control points [Yendiki et al., 2011], and hierarchical Dirichlet process mixture model [Wang et al., 2011].

Model-based methods are however not very popular, mainly because usually the models can fit only a limited number of white matter bundles.

Classification-based methods

Classification-based methods are supervised methods that segment specific bundles of interest in new subjects by leveraging one or multiple examples of that segmented bundle in other subjects. Specifically, the core of these methods is to create from a set of examples a classifier that is able to classify new given samples in one or more classes, see for example [Olivetti and Avesani, 2011, Mayer et al., 2011, Vercruysse et al., 2014, Heker et al., 2016]. Classification-based methods mainly differ with respect to their input features and to the classification algorithm adopted.

For example, in [Mayer et al., 2011, Heker et al., 2016], they adopted an AdaBoost framework over data embedded with the streamline 'flattening' strategy [Mayer et al., 2011] or using tailored physical, statistical and geometrical features [Heker et al., 2016]. In [Vercruysse et al., 2014], they used the spectral embedding technique to build the feature space to train Support Vector Machine (SVM) and random forest classifiers. In [Olivetti and Avesani, 2011], first the streamlines were embedded into an Euclidean space through the dissimilarity representation, and then a linear SVM classifier was chosen for the final single streamline classification.

Although mainly due to their flexibility these methods have demonstrated to be promising to solve the bundle segmentation problem, they suffer from several limitations, such as the dependency on the number of subjects used for training, the imbalance of positive and negative classes, and low computational efficiency. Additionally, another significant disadvantage is their dependency on the quality of input data. To work properly, classification algorithms should ideally learn from accurate examples validated by experts, which represent the ground truth. Using example bundles not validated [Olivetti and Avesani, 2011, Vercruysse et al., 2014], or segmented manually by unexperienced hands [Heker et al., 2016], drastically reduces the classification performances.

Correspondence-based methods

Recently, several methods based on the idea of finding a correspondence between streamlines across different tractograms have been proposed, see for example [Yoo et al., 2015, Olivetti et al., 2016, Labra et al., 2016, Sharmin et al., 2018, Garyfallidis et al., 2018]. A key element of these approaches is the definition of a similarity measure, which in general is a pairwise distance function between pairs of streamlines. As classification-based methods, correspondence-based methods for bundle segmentation leverage information from one or more example bundles segmented a priori and opportunely registered in a common anatomical space.

Given one streamline of the example bundle, the simplest solution to find the corresponding streamline in the target subject is to consider its closest one, i.e. the one with the shortest distance [Yoo et al., 2015, Garyfallidis et al., 2018]. This algorithm, usually referred to as the nearest neighbor (NN) algorithm, creates a many-to-one correspondence, i.e. multiple streamlines of the example bundle can correspond to the same streamline in the target tractogram. However, when the two tractograms exhibit a local systematic displacement at the bundle level, the many-to-one correspondence may become predominant, leading to an unrealistic thinning of the segmented bundle and to an inaccurate segmentation [Sharmin et al., 2016].

In [Garyfallidis et al., 2018], they proposed a supervised segmentation method, called RecoBundles, that takes as input one example bundle (model bundle) that is used to estimate the corresponding bundle in a new tractogram. The method first, leverages state-of-the-art streamline-based registration and clustering algorithm to align and prepare the tractogram for segmentation. Then, it segments the bundle of interest with a variation of the NN algorithm, in the sense that streamlines within a certain distance threshold from the model bundle are kept to form the segmented bundle. RecoBundles has proven to be efficient in dealing with large datasets, and to be robust to missing data and to bundles with missing components [Garyfallidis et al., 2018]. Nevertheless, tuning the threshold values for different kind of bundles and datasets to obtain an accurate segmentation is not straightforward.

To partly address the issue of thin bundles obtained with the NN approach, in [Sharmin et al., 2016, Sharmin et al., 2018] they proposed to find streamline correspondence through the solution of a Linear Assignment Problem (LAP). Bundle segmentation is indeed treated as an assignment problem, which is one of the fundamental combinatorial optimization problem. As opposed to the NN strategy, this method introduces a further constraint of a one-to-one correspondence between the streamlines of the example and those of the target tractogram. This means that a streamline of the target tractogram can be the corresponding streamline only of one streamline of the example bundle, and viceversa. The supervised method presented in [Sharmin et al., 2018] is the extension of that proposed in [Sharmin et al., 2016] to the case of multiple example bundles. If N example bundles are available, first, the correspondence is computed individually in order to obtain N candidate bundles in the target subject. Then, the the final segmented bundle is estimated through a refinement step that classifies the streamlines of all the candidate bundles based on majority voting. The workflow of this method, called LAP, is depicted in Figure 3.5. In [Sharmin et al., 2018], they showed that their method outperformed

ROI-based approaches and methods based on the NN algorithm. However, a major limitation of LAP is that it cannot deal with large tractograms since the computation of the assignment is both time- and memory-wise too demanding.

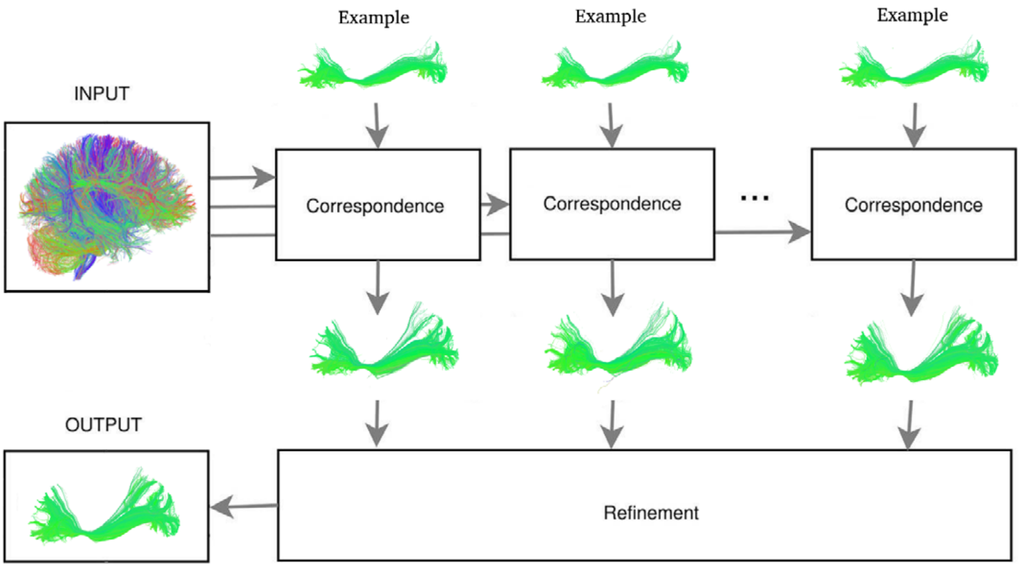


Figure 3.5: Linear Assignment Problem (LAP) workflow. First, the candidate bundles are estimated from the example bundles by finding the correspondence between streamlines of the examples and those of the tractogram individually (top). Then, the final bundle of interest is retrieved through a refinement step (bottom). Image adapted with permission from [Sharmin et al., 2018].

3.2.4 Direct methods

Direct methods are voxel-based methods that segment bundles directly from diffusion MRI images, without the need for streamlines. Some of the earlier methods of this category employed techniques like template matching [Eckstein et al., 2009], surface evolution [Descoteaux and Deriche, 2009], and Markov Random Field optimization [Bazin et al., 2011]. However, these methods have not become very popular mainly due to their limited segmentation accuracy [Wasserthal et al., 2018a].

On the other hand, the recent direct method proposed in [Wasserthal et al., 2018a], called TractSeg, presented evidence of remarkably better segmentation quality than a large selection of other segmentation methods. TractSeg is a voxel-based method based on Convolutional Neural Networks (CNNs) inspired by [Ronneberger et al., 2015], which are known to be the best machine learning models for grid-based domains like voxels. Its input consists of fiber orientation distribution function (fODF) peaks computed in each voxel, while its output is composed of the voxel masks of multiple segmented bundles, see Figure 3.6. Streamlines of the bundles can later be obtained by performing bundle-specific tracking inside the predicted masks using Tract Orientation Maps (TOMs) [Wasserthal et al., 2018b].

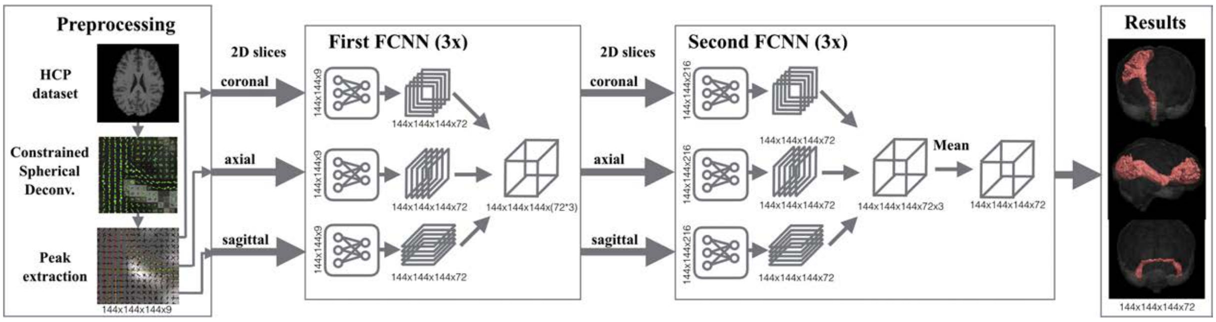


Figure 3.6: TractSeg workflow. First, fiber orientation distribution function (fODF) peaks are extracted from dMRI images. Then, two Convolutional Neural Networks (CNNs) are trained, and finally the bundle masks are predicted. Image taken from [Wasserthal et al., 2018a].

Although having shown to provide unprecedented accuracy of segmentation, therefore setting a new standard in terms of bundle segmentation quality, TractSeg suffers from some limitations. Its major drawback is that, due to its voxel-based nature, it is not very precise in segmenting fine-grained structures of the bundles. As a consequence, it reaches poor performances especially in segmenting the terminal portions of the bundles, and bundles which are not particularly large and rounded, such as small and short bundles.

Chapter 4

Problem Statement

In this Chapter, we describe the open problems we aim to address with this thesis. First, in Section 4.1, we provide some basic notations and definitions in order to formulate the problem of supervised white matter bundle segmentation. Then, in Section 4.2, we present the main research topics of the thesis and we summarize the three main research questions we want to answer with **Q1**, **Q2** and **Q3**.

4.1 Basic notations and definitions

A *streamline* $s = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is a sequence of 3D points, $\mathbf{x}_i = [x_i, y_i, z_i] \in \mathbb{R}^3$, $i = 1 \ldots n$. Usually, n is in the order of $10^1 - 10^2$ and differs across streamlines.

A *tractogram* T is the entire set of streamlines of the white matter of a brain: $T = \{s_1, \ldots, s_M\}$, where in general M is in the order of $10^5 - 10^6$.

A white matter *bundle*, or *tract*, $b \subset T$, $b = \{s_1, \ldots, s_m\}$, is a group of streamlines with a specific anatomical meaning, where $m \ll M$, and m differs across bundles.

A *streamline-streamline* (or *pairwise*) *distance function* between two streamlines s_A and s_B, defined as $d(s_A,s_B)$, is a measurement of their local spatial displacement. Intuitively, the closer the two streamlines, the more $d(s_A,s_B)$ tends to 0.

Given a set of examples of the same anatomical bundle belonging to different subjects, $b_{examples} = \{b_1, \ldots, b_{N_e}\}$, the goal of *supervised white matter bundle segmentation* is to segment the same anatomical bundle in a new target subject, see Figure 4.1.

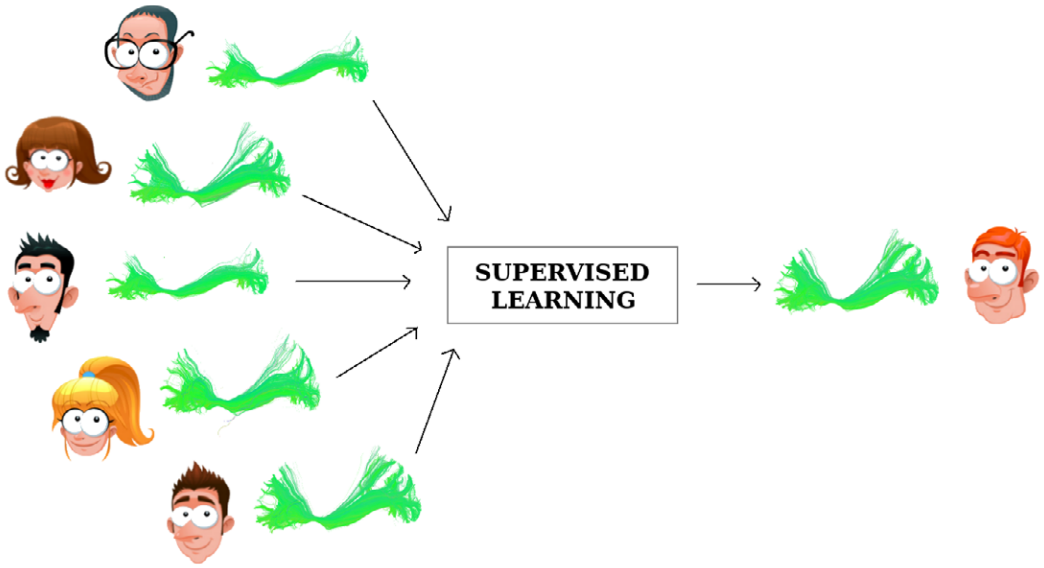


Figure 4.1: Supervised learning approach for white matter bundle segmentation. Some examples of the same kind of bundle, e.g. the inferior fronto-occipital fasciculus, are given in 5 different subjects. With this data, the goal of supervised learning is to find the same kind of bundle, e.g. the inferior fronto-occipital fasciculus, in a new target subject.

In the context of streamline-based supervised learning, given a tractogram T and a set of examples $b_{examples} = \{b_1, \dots, b_{N_e}\}$, an *ideal* bundle segmentation algorithm f is able to determine whether on not a streamline $s \in T$ belongs to the bundle of interest $b \subset T$, i.e. it can establish a mapping $f : T \mapsto \{0, 1\}$ where:

$$f(s) = \begin{cases} 1 & \text{if } s \in b \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

In practice, f is not available, and the goal of a supervised algorithm is to estimate a function g that best approximates f.

4.2 Research topics and questions

As already mentioned, obtaining a good and reliable bundle segmentation is not trivial, mainly because of the intrinsic complexity of the data. In particular, the borders of the bundles are not always well-defined, many bundles overlap with others, and their shape and size have a large intra- and inter-subject variability.

As comprehensively reviewed in Chapter 3, to overcome the limitations of manual approaches [Catani et al., 2002, Wakana et al., 2007], which are very time consuming and affected by human subjectivity, in the last few decades several automatic methods to address the task of bundle segmentation have been developed. Promising automatic methods have proven to be especially those based on supervised learning, which are able to learn how to virtually extract a bundle of interest in a new subject by leveraging multiple examples of the same bundle available in other subjects [Garyfallidis et al., 2018, Sharmin et al., 2018, Wasserthal et al., 2018a]. Nevertheless, these state-of-the-art supervised methods suffer from some important drawbacks, such as dependency on multiple parameters, scalability issues, and inability to adapt to different data settings.

To overcome these limitations, the main goal of this thesis is to develop a robust automatic method for accurate white matter bundle segmentation, by exploiting, combining and extending multiple up-to-date supervised learning techniques, see Section 4.2.3. Important factors involved in the task of streamline classification are (i) the need for an effective *streamline distance function*, Sections 4.2.1, and (ii) the definition of a proper *feature space*, see Section 4.2.2.

4.2.1 Streamline distances assessment

As discussed in Section 3.1.1, streamline distance functions are building blocks for a variety of tasks, among which supervised bundle segmentation. However, since the streamlines are 3D polylines in space that can generally have a different number of points and a different step size, the definition of a pairwise distance function is not unique.

Although each streamline distance has its distinct technical motivation, little has been said to guide the selection of a distance for a specific task. To the best of our knowledge, comparison and evaluation of distances were performed only in the case of *unsupervised* bundle segmentation, i.e. by means of clustering, see Section 3.1.1 for a brief review. Investigations about the most suitable streamline distance function to perform *supervised* bundle segmentation have not been conducted yet. Moreover, computing hundreds of thousands of distances may significantly affect the overall computational time for the segmentation task.

As a consequence, a preliminary useful assessment to address the problem of *supervised* bundle segmentation is to establish which of the distance functions available in the literature is the most effective for the specific task. Additionally, an evaluation of the computational time required to compute each of the compared distance functions could be of interest to practitioners.

4.2.2 Beyond streamline geometry

Correspondence-based methods have shown to outperform connectivity-based techniques because, as reviewed in Section 3.2.2, these last ones are inherently limited by the anatomical variability of the subjects, by the quality of the atlas or parcellation, and by the process of registration [Garyfallidis et al., 2018, Sharmin et al., 2018]. Another disadvantage of connectivity-based techniques is that they do not take into account the shape of the streamlines, but only their closeness to anatomical Regions Of Interest (ROIs). On the other hand, correspondence-based methods are based only on geometrical properties of the streamlines without considering any prior anatomical information of the bundle of interest.

We aim to improve the results of bundle segmentation by considering information about both the streamline geometry, typical of correspondence-based methods, and about the relative anatomical position of the bundle of interest, typical of ROI-based and connectivity-based techniques. The underlying assumption is that streamlines composing the same anatomical bundles not only have a similar shape, but are also characterized by their propensity to interconnect or pass through predefined regions of the brain. In the case of streamline classification, a feature space which simultaneously combines information from atlases, bundle geometries, and connectivity patterns could be effective for the bundle segmentation task.

4.2.3 Towards a robust streamline-based method for bundle segmentation

From the survey of the literature about white matter bundle segmentation techniques in Section 3.2, three state-of-the-art supervised methods have shown to achieve the best segmentation performances: RecoBundles [Garyfallidis et al., 2018], LAP [Sharmin et al., 2018], and TractSeg [Wasserthal et al., 2018a]. Each of them however has its own limitations, which we summarize here:

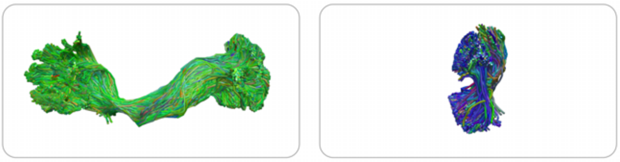
- Limitations of RecoBundles: variation of the Nearest Neighbor (NN) strategy, dependency on multiple parameters, not robust to different data settings.

- Limitations of LAP: one-to-one constraint, inability to address large datasets, not robust to different data settings.

- Limitations of TractSeg: less effective in segmenting fine-grained structures of the bundles, not robust to different data settings.

A common drawback is that they are not robust to different data settings, i.e. the quality of segmentation may be strongly affected by some properties of the bundles, for example by their size, by the tractography technique, e.g. probabilistic or deterministic tracking algorithm, and by the data quality, e.g. research (high-resolution) or clinical quality, see Figure 4.2 for some examples.

A. Bundle size (large versus small)



B. Tracking algorithm (probabilistic versus deterministic)

C. Data quality (research quality versus clinical quality)

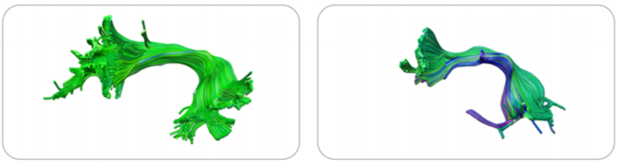


Figure 4.2: Examples of different data settings. A. Two bundles with different size, on the left a large bundle (inferior-fronto-occipital fascicle) and on the right a small bundle (posterior arcuate). B. Two bundles (corticospinal tracts) obtained using different tracking algorithms, on the left with probabilistic and on the right with deterministic tracking. C. Two bundles (arcuate fascicles) segmented from diffusion data of different quality, on the left at research quality and on the right at clinical quality.

It is worth noting that in general, as discussed in Section 2.2.3, the choice of the most appropriate pipeline for tractography is not univocal, but rather is strongly affected by the quality of the available diffusion data, and changes according to the specific application, depending on the desired level of sensitivity/specificity. For example, to avoid unintentional dissection of crucial fibers during neurosurgical sessions, a tractography technique with high sensitivity may be preferred. On the contrary, if the goal is to reduce the possibility to identify spurious connections, a more conservative tractography method may be desirable to reduce false positives [Thomas et al., 2014]. Moreover, even though the interest in large associative bundles is well established in multiple applications [Wandell, 2016, Pestilli, 2018], small and short bundles have recently received increased attention [Guevara et al., 2011, Wu et al., 2016b, Guevara et al., 2017, Bullock et al., 2019]. For example, the relatively smaller bundles connecting the human dorsal and posterior cortices have been recently proven to be of great help in understanding how information flows in the human brain [Wu et al., 2016b, Bullock et al., 2019, Sani et al., 2019].

The main goal of this thesis is indeed to develop a novel streamline-based method which is robust to different data setting, specifically that can deal with either (i) large/small bundles, (ii) probabilistic/deterministic tractography, and (iii) research/clinical data quality. We believe in fact that maintaining a high bundle segmentation quality across different settings is of primary importance for subsequent applications. Additional desired qualities are computational efficiency, speed, and low resource and memory consumption. We aim to achieve the goal by exploiting, combining and extending multiple up-to-date supervised learning techniques, and by leveraging the results of our investigations about streamline distance functions and about the possibility to include anatomical information into a streamline-based bundle segmentation method.

4.2.4 Summary of the main research questions

To summarize, the three main research questions we aim to address within this thesis are:

Q1: Which streamline distance function available in the literature is the most effective for supervised bundle segmentation?

Q2: Why not including prior anatomical information in a streamline-based bundle segmentation method?

Q3: Can we develop a novel robust streamline-based bundle segmentation method which is able to adapt to multiple data settings?

Chapter 5

Methods

In this Chapter, we present the methods used and developed to address the research questions of this thesis. Sections 5.1, 5.2 and 5.3 describe key concepts and algorithms available in the literature that were auxiliary to develop the main methodological contributions of the thesis, which are presented in detail in Sections 5.4 and 5.5. Section 5.6 reports the evaluation metrics used to quantify the bundle segmentation performances, while in Section 5.7 we introduce the notion of fractal dimension (FD) of a bundle, which will be used to discuss some of the experimental results.

5.1 Streamline distance functions and dissimilarity representation

5.1.1 Streamline distance functions

Given two streamlines s_a and s_b, here we provide the mathematical formulation of some of the streamline distance functions frequently used in the literature, which will be compared in Section 7.1. The idea is that streamlines belonging to the same anatomical structure lie at small distances.

1. Mean of closest distances [Corouge et al., 2004]:

$$d_{\mathrm{MC}}(s_a, s_b) = \frac{d_m(s_a, s_b) + d_m(s_b, s_a)}{2} \tag{5.1}$$

   where $d_m(s_a, s_b) = \frac{1}{|s_a|} \sum_{\mathbf{x}_i \in s_a} \min_{\mathbf{x}_j \in s_b} ||\mathbf{x}_i - \mathbf{x}_j||_2$

2. Shorter mean of closest distances [Zhang et al., 2008]:

$$d_{\mathrm{SC}}(s_a, s_b) = \min(d_m(s_a, s_b), d_m(s_b, s_a)) \tag{5.2}$$

3. Longer mean of closest distances [Zhang et al., 2008]:

$$d_{\mathrm{LC}}(s_a, s_b) = \max(d_m(s_a, s_b), d_m(s_b, s_a)) \tag{5.3}$$

4. Minimum average direct flip (MDF) distance [Garyfallidis et al., 2012], after resampling each streamline to a given number of points p, such as $s_a = \{\mathbf{x}_1^a, \ldots, \mathbf{x}_p^a\}$ and $s_b = \{\mathbf{x}_1^b, \ldots, \mathbf{x}_p^b\}$:

$$d_{\mathrm{MDF}}(s_a, s_b) = \min(d_{\mathrm{direct}}(s_a, s_b), d_{\mathrm{flipped}}(s_a, s_b)) \tag{5.4}$$

where $d_{\mathrm{direct}}(s_a, s_b) = \frac{1}{p}\sum_{i=1}^{p} ||\mathbf{x}_i^a - \mathbf{x}_i^b||_2$ and $d_{\mathrm{flipped}}(s_a, s_b) = \frac{1}{p}\sum_{i=1}^{p} ||\mathbf{x}_i^a - \mathbf{x}_{p-i+1}^b||_2$

5. Point Density Model (PDM) [Siless et al., 2013]:

$$d^2_{\mathrm{PDM}}(s_a, s_b) = \langle s_a, s_a \rangle_{pdm} + \langle s_b, s_b \rangle_{pdm} - 2\langle s_a, s_b \rangle_{pdm} \tag{5.5}$$

where

$$\langle s_a, s_b \rangle_{pdm} = \frac{1}{|s_a||s_b|} \sum_{i=1}^{|s_a|} \sum_{j=1}^{|s_b|} K_\sigma(\mathbf{x}_i^a, \mathbf{x}_j^b) \tag{5.6}$$

and $K_\sigma(\mathbf{x}_i^a, \mathbf{x}_j^b) = \exp\left(-\frac{||\mathbf{x}_i^a - \mathbf{x}_j^b||_2^2}{\sigma^2}\right)$ is a Gaussian kernel between the two 3D points.

6. Varifolds distance [Charon and Trouvé, 2013], which is the non-oriented version of the currents distance [Gori et al., 2016], namely it does not need streamlines s_a and s_b to have a consistent orientation:

$$d^2_{\mathrm{varifolds}}(s_a, s_b) = \langle s_a, s_a \rangle_{var} + \langle s_b, s_b \rangle_{var} - 2\langle s_a, s_b \rangle_{var} \tag{5.7}$$

where

$$\langle s_a, s_b \rangle_{var} = \sum_{i=1}^{|s_a|-1} \sum_{j=1}^{|s_b|-1} K_\sigma(\mathbf{p}_i^a, \mathbf{p}_j^b) K_n(\mathbf{n}_i^a, \mathbf{n}_j^b) |\mathbf{n}_i^a|_2 |\mathbf{n}_j^b|_2 \tag{5.8}$$

with $K_n(\mathbf{n}_i^a, \mathbf{n}_j^b) = \left(\frac{(\mathbf{n}_i^a)^T \mathbf{n}_j^b}{|\mathbf{n}_i^a|_2 |\mathbf{n}_j^b|_2}\right)^2$ where $\mathbf{p}_i^a$ (resp. $\mathbf{p}_j^b$) and $\mathbf{n}_i^a$ (resp. $\mathbf{n}_j^b$) are the center and tangent vector of segment i (resp. j) of streamline s_a (resp. s_b). The endpoints of segment i are $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ for $i \in [1, ..., n-1]$.

5.1.2 Dissimilarity representation

As explained in Section 3.1.2, the dissimilarity representation is an embedding technique to embed a streamline into a vectorial space [Olivetti et al., 2012]. The dissimilarity representation describes the streamline with a set of distances from L landmarks, which are usually streamlines taken from a whole brain tractogram. In principle, any streamline distance function d described in 5.1.1 can be used. Given d and the set of landmarks $\Pi = \{l_1, l_2, \ldots, l_L\}$, the dissimilarity representation of a streamline s is defined as:

$$\phi_{\Pi}^{d}(s) = [d(s, l_1), d(s, l_2), \ldots, d(s, l_L)] \tag{5.9}$$

5.2 Nearest Neighbor (NN) algorithm

Given an object, the Nearest Neighbor (NN) algorithm is a greedy strategy that finds, in a given set, its closest (or its most similar) object, i.e. its nearest neighbor. The concept of ”closeness” is frequently expressed in terms of distance measures. More formally, given a query object o^A and a set of objects $S^B = \{o_1^B, \ldots, o_M^B\}$, the NN algorithm computes the closest object $o_i^B \in S^B$ to o^A:

$$o_i^B = NN(o^A, S^B) = \underset{o_i^B \in S^B}{\operatorname{argmin}}\, d(o^A, o_i^B) \tag{5.10}$$

where d is a distance function. In the case of a set of querying objects $S^A = \{o_1^A, \ldots, o_N^A\}$, the NNs of S^A correspond to the union of all the NN $\forall o_i^A \in S^A$. It is worth noting that the result of the queries for two distinct query objects in S^A may correspond to the same object in S^B.

In the context of streamline-based bundle segmentation, the NN strategy selects, for each individual streamline of the example bundle, the closest streamline in the target tractogram by performing a distance minimization. We can consider the two sets of objects as being (i) an example bundle of a subject A, i.e. $b^A = \{s_1^A, \ldots, s_m^A\}$, and (ii) the tractogram of a subject B, i.e. $T^B = \{s_1^B, \ldots, s_M^B\}$, from which we aim to segment the same anatomical bundle. The goal of a bundle segmentation algorithm based on NN is to estimate the target bundle $\hat{b}^B \subset T^B$, which corresponds to the union of the neighboring streamlines of b^A in T^B. Under the assumption that b^A and T^B lie in a common space, the target bundle $\hat{b}^B \subset T^B$ can be estimated as follows:

$$\hat{b}^B = \bigcup_{i=1}^{m} NN(s_i^A, T^B) \qquad \forall s_i^A \in b^A \tag{5.11}$$

The NN algorithm is a particular case of the k-NN algorithm in which k, the number of nearest neighbors to compute at each query, is equal to 1. In the present work, we used the k-NN algorithm in two different situations: with $k = 1$ as just explained for bundle segmentation, and with $k \gg 1$ for bundle supersets computation, e.g. with $k = 2000$. Efficient k-NN queries were possible due to the use of a K-dimensional tree (K-D tree) on a vectorial representation of the streamlines, see [Sharmin et al., 2018]. A K-D tree is a data structure to efficiently store and retrieve vectorial data in a K-dimensional space, a behavior particularly useful when large datasets are involved.

5.3 Linear Assignment Problem (LAP) algorithm

Given two sets of objects, $S^A = \{o_1^A, \ldots, o_M^A\}$ and $S^B = \{o_1^B, \ldots, o_M^B\}$, and the cost matrix $C = \{c_{ij}\}_{ij} \in \mathbb{R}^{M \times M}$, where c_{ij} is the cost of assigning o_i^A to o_j^B, the Linear Assignment Problem (LAP) [Bijsterbosch and Volgenant, 2010] aims to find the one-to-one correspondence between the objects in A and the objects in B by minimizing the total cost:

$$P^* = \underset{P \in \mathcal{P}}{\operatorname{argmin}} \sum_{i,j=1}^{M} c_{ij} p_{ij} \tag{5.12}$$

where $P = \{p_{ij}\}_{ij} \in \mathcal{P}$ is a permutation matrix and P^* is the optimal assignment[1]. One of the most efficient algorithm to solve the LAP is the Jonker-Volgenant algorithm (LAPJV) [Bijsterbosch and Volgenant, 2010]. As opposed to the NN strategy, the Linear Assignment Problem (LAP) algorithm introduces a one-to-one constraint to the correspondence by performing a joint minimization of distances over all the pairs of objects.

In the context of streamline-based bundle segmentation, we can consider the two sets of objects as being (i) an example bundle of a subject A, i.e. $b^A = \{s_1^A, \ldots, s_m^A\}$, and (ii) the tractogram of a subject B, i.e. $T^B = \{s_1^B, \ldots, s_M^B\}$, from which we aim to segment the same anatomical bundle. The goal of a bundle segmentation algorithm based on LAP is to find the optimal correspondence P^* of all the streamlines in b^A with those in T^B. In [Sharmin et al., 2016, Sharmin et al., 2018], the cost matrix is equal to the distance matrix D between the two sets of streamlines, in which each element is given by $c_{ij} = d_{\mathrm{MC}}(s_i^A, s_j^B)$. Intuitively, the closer two streamlines are to each other, the more likely they belong to the same anatomical structure. The segmentation from multiple

---

[1] If the size of the two sets of objects is different, i.e. $|A| \neq |B|$, the problem is called Rectangular Linear Assignment Problem (RLAP) and P becomes a partial permutation matrix [Bijsterbosch and Volgenant, 2010].

examples proposed in [Sharmin et al., 2018] is obtained by merging together the solution of multiple LAP solved individually. This is done through a refinement step that ranks and filters the streamlines based on majority vote.

5.4 Anatomically-Informed Linear Assignment Problem (LAP-anat) algorithm

In this Section, we present our methodological contribution to improve the results of bundle segmentation by taking into account information about both the streamline geometry and the relative anatomical position of the bundle of interest. To achieve this goal, we propose to extend the correspondence-based method proposed in [Sharmin et al., 2018] by including additional anatomical information within the optimization process of the Linear Assignment Problem (LAP) [Bijsterbosch and Volgenant, 2010]. Specifically, the anatomical information is given by the location of the streamline endpoints and the proximity of the streamlines to specific anatomical ROIs predefined in the literature. The novelty consists in linearly combining and optimizing multiple matrices rather that only the distance matrix D as originally proposed in [Sharmin et al., 2018]. We refer to this extension of the LAP algorithm as LAP-anat. In particular, we extend the cost matrix D by adding two weighted anatomically-informed cost matrices: the *endpoint-distance matrix* E and the *ROI-based distance matrix* R. Hence, the new cost matrix becomes $C = \lambda_D D + \lambda_E E + \lambda_R R$, see Figure 5.1 for a graphical intuition about how the individual distances are computed.

5.4.1 Endpoint-based distance matrix

Due to the fact that white matter fibers serve as connections between brain areas at their terminations, from an anatomical and functional point of view and regardless of their geometry, two streamlines with neighboring endpoints are assumed to play a similar role. Based on this concept, we propose to create a new cost matrix E by defining the *endpoint-based* distance, d_{END}, between two streamlines s_a and s_b as the mean Euclidean distance of their corresponding endpoints:

$$d_{\text{END}}(s_a, s_b) = \frac{\min(\|\mathbf{x}_1^a - \mathbf{x}_1^b\|, \|\mathbf{x}_1^a - \mathbf{x}_{n_b}^b\|) + \min(\|\mathbf{x}_{n_a}^a - \mathbf{x}_1^b\|, \|\mathbf{x}_{n_a}^a - \mathbf{x}_{n_b}^b\|)}{2} \tag{5.13}$$

where $\{\mathbf{x}_1^a,\mathbf{x}_{n_a}^a\}$ are the endpoints of s_a and $\{\mathbf{x}_1^b,\mathbf{x}_{n_b}^b\}$ are those of s_b, see Figure 5.1 (middle) for a graphical intuition.

5.4.2 ROI-based distance matrix

In the literature, white matter bundles have been defined with respect to specific ROIs that they cross or touch [Catani et al., 2002, Wakana et al., 2007, Yeatman et al., 2012, Wassermann et al., 2016]. Inspired by these works, in order to create an additional cost matrix R containing information regarding the relative position of a bundle from specific ROIs, we first define the distance between a streamline and a single ROI, as the minimum Euclidean distance between them:

$$d_{min}(s, \mathrm{ROI}) = \min_{\mathbf{x}_i \in s, v_j \in \mathrm{ROI}} \|\mathbf{x}_i - coord(v_j)\| \tag{5.14}$$

where $coord(v)$ are the 3D coordinates of the center of the voxel v belonging to the ROI, see Figure 5.1 (right) for a graphical intuition. In the case of multiple ROIs, we define the distance between a streamline and them as the mean of the distances to each ROI. With this building block, given a set of N ROIs defining a bundle, the ROI-based distance between two streamlines s_a and s_b can be defined as:

$$d_{\mathrm{ROIs}} = |\frac{1}{N}\sum_{i=1}^{N} d_{\mathrm{min}}(s_a, \mathrm{ROI}_i) - \frac{1}{N}\sum_{i=1}^{N} d_{\mathrm{min}}(s_b, \mathrm{ROI}_i)| \tag{5.15}$$

meaning that two streamlines with similar distances from the ROIs are likely to belong to the same white matter bundle.

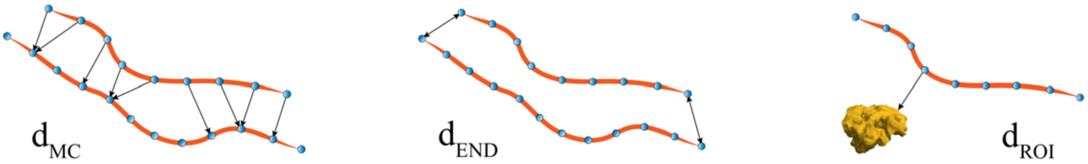


Figure 5.1: Graphical intuition about how the individual distances are computed to form the distance matrix D (left), the endpoint-distance matrix E (middle), and the ROI-based distance matrix R (right).

5.5 Classifyber, a binary linear classification algorithm for streamlines

In this Section, we present the main methodological contribution of this thesis, i.e. Classifyber, a binary linear classification algorithm for streamlines. Classifyber is a novel method that performs automatic bundle segmentation as a supervised learning problem, meaning that the algorithm learns how to segment from expert-based examples. The name *Classifyber* is the linguistic blend of *Classify* and *fiber*, which explains the basic principle of its functioning: to classify whether or not a given streamline/fiber[2] belongs to the bundle of interest.

In general, as briefly introduced in Section 2.4.2, a classifier aims to classify given objects into predefined classes. Given some examples representing pairs of objects x with the corresponding class label y, i.e. $(x_1,\ y_1),\ \ldots,\ (x_N,\ y_N)$, the *ideal* classifier f is a function that assigns to each element x its class label $y = f(x)$. If the label y can only take two values, the classification is known as binary classification. In practice, since f is only ideal, the goal of a classification algorithm is to estimate a function g that best approximates f. The function g is estimated from labeled examples by training a model to minimize a loss function, which quantifies the goodness of the prediction.

In the case of binary linear classification, a linear classifier l takes as input a vector $\mathbf{v} \in \mathbb{R}^d$ and returns its predicted category, i.e. 0 or 1:

$$l(\mathbf{v}) = \begin{cases} 1 & \text{if } a_1 v_1 + \cdots + a_d v_d + a_0 > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.16}$$

where the coefficients of the linear classifier $a_0, \ldots, a_d$ are estimated by minimizing the errors in classification on a training set (plus regularization terms that may differ between different algorithms).

In the context of streamline-based bundle segmentation, given a streamline s and a tractogram T, the binary linear classifier l aims to determine whether or not a streamline $s \in T$ belongs to the bundle of interest $b \subset T$, i.e. $l(\mathbf{v}) = 1$ if $s \in b$, 0 otherwise, where $\mathbf{v}$ is a vectorial representation of the streamline s.

5.5.1 Feature Space

In order to use a linear classifier on a streamline s, it is necessary to transform the streamline into a vector $\mathbf{v}$ that contains the necessary information for the task of bundle segmentation. In other words, we need to define an effective *feature space* to represent streamlines as vectors. This is a key step in the proposed method, where we extract geometrical and anatomical information from the streamline and create this multidimensional vector. The peculiarity of this novel specifically-designed feature space is that takes into

---

[2] In some literature, the name *fiber* refers to *axon* and in other literature to *streamline*. Here we refer to the latter for linguistic convenience.

account both geometrical properties of the streamlines and their vicinity to anatomical ROIs, a solution that provided to be successful within the optimization process of the LAP-anat algorithm presented in Section 5.4. Moreover, whenever a streamline distance function is involved, as a consequence of the preliminary assessment about streamline distances that will be reported in Section 7.1, we employ the minimum average direct flip distance d_{MDF}. Additional streamline features that can be considered are Fractional Anisotropy (FA), and rigid body invariants like curvature, and torsion. Specifically, we propose to describe each streamline using a combination of the following features:

(i) distances from Global Landmarks (d_{GL})

(ii) distances from Local Landmarks (d_{LL})

(iii) Endpoint-based distances (d_{END})

(iv) ROI-based distances (d_{ROI})

(v) FA profile

(vi) curvature and torsion profiles

Distances from Global Landmarks (d_{GL})

One possible way to describe a streamline as a vector is to compute its dissimilarity representation as originally proposed in [Olivetti et al., 2012] and presented in Section 5.1.2, i.e. each streamline is described with an L-dimensional vector of distances from L global landmarks. We use the the minimum average direct flip distance d_{MDF} on streamlines resampled at 20 points, while the global landmarks correspond to $L = 100$ streamlines chosen in a tractogram of one random subject using the policy subset farthest first (SFF), which provides a uniform coverage of the brain. It is worth noting that, since the set of landmarks is a reference system, it has to be the same for every subject.

Given d_{MDF} the MDF distance and $\Pi = \{\tilde{l_1}, \tilde{l_2}, \ldots, \tilde{l_L}\}$ the set of global landmarks, the dissimilarity representation of a streamline s is defined as:

$$d_{\mathrm{GL}}(s, \Pi) = \phi_{\Pi}^{d_{\mathrm{MDF}}}(s) = [d_{\mathrm{MDF}}(s, \tilde{l_1}), d_{\mathrm{MDF}}(s, \tilde{l_2}), \ldots, d_{\mathrm{MDF}}(s, \tilde{l_L})] \tag{5.17}$$

Distances from Local Landmarks (d_{LL})

An alternative way of choosing the landmarks is to select them locally, i.e. in the region surrounding the bundle of interest. A denser local sampling may help in discriminating nearby streamlines in the area in which the bundle of interest is located. In this case, the $L = 100$ local landmarks are chosen in one random subject using the SFF policy in a superset of the bundle of interest. This superset is computed by considering the neighboring streamlines belonging to the corresponding tractogram retrieved by a k-NN procedure (k=2000) applied to each streamline of the bundle. As for the global landmarks, fixed a bundle, the local landmarks need to be the same for every subject.

Given d_{MDF} the MDF distance and $\Gamma = \{\hat{l_1}, \hat{l_2}, \ldots, \hat{l_L}\}$ the set of local landmarks, the dissimilarity representation of a streamline s is defined as:

$$d_{\text{LL}}(s, \Gamma) = \phi_{\Gamma}^{d_{\text{MDF}}}(s) = [d_{\text{MDF}}(s, \hat{l_1}), d_{\text{MDF}}(s, \hat{l_2}), \ldots, d_{\text{MDF}}(s, \hat{l_L})] \tag{5.18}$$

Endpoint-based distances (d_{END})

Since a specific anatomical bundle b has the property to interconnect two areas of the brain, we expect that streamlines belonging to that bundle have terminal points spatially close to each other. For this reason, we assume the spatial locations of the terminal points of a streamline being informative to solve the classification task. We indeed propose to describe a streamline s with L features corresponding to the mean Euclidean distances between its endpoints and those of the $L = 100$ global landmarks.

Given a streamline s and a global landmark $\tilde{l}$, the endpoint-based distance $d_{\text{END}}(s, \tilde{l})$ is defined as in Equation 5.13. Consequently, given the set of global landmarks Π, a streamline s can be described as follows:

$$d_{\text{END}}(s, \Pi) = \phi_{\Pi}^{d_{\text{END}}}(s) = [d_{\text{END}}(s, \tilde{l_1}), d_{\text{END}}(s, \tilde{l_2}), \ldots, d_{\text{END}}(s, \tilde{l_L})] \tag{5.19}$$

ROI-based distances (d_{ROI})

Besides the geometrical information of a streamline and location of its terminal points, it is possible to consider anatomical information quantifying its proximity to ROIs pertaining to the bundle of interest. This approach is meant to partially emulate the connectivity-based segmentation techniques in [Wakana et al., 2007, Yeatman et al., 2012], which filter each bundle of interest by a combination of two planar waypoint ROIs that delineate its path before its divergence towards the cortex.

Given a streamline s and a single ROI, their distance $d_{min}(s, \text{ROI})$ is defined as the minimum Euclidean distance between them, see Equation 5.14. Consequently, given a set of two ROIs $\Lambda = \{\text{ROI}_1, \text{ROI}_2\}$, a streamline s can be described as follows:

$$d_{\text{ROI}}(s, \Lambda) = \phi_{\Lambda}^{d_{min}}(s) = [d_{min}(s, \text{ROI}_1), d_{min}(s, \text{ROI}_2)] \tag{5.20}$$

Fractional Anisotropy (FA) profile

Inspired by the works in [Colby et al., 2012, Yeatman et al., 2012], we treat the Fractional Anisotropy (FA) profile as a streamline descriptor. After a resampling of the streamline at 100 equally-spaced points, the FA profile is computed by taking at each point the FA value of its corresponding voxel, see Figure 2.10 for an example. In this case, streamlines in a bundle must have the same orientation.

Curvature and torsion profiles

Inspired by the works in [Corouge et al., 2004, Batchelor et al., 2006], we consider local shape descriptors to embed a streamline into a vector. Local shape descriptors are derived from the Frenet-Serret framework, which describes the geometric properties of a differentiable curve in $\mathbb{R}^3$:

$$\begin{cases} \frac{d\mathbf{T}}{ds} = \kappa\mathbf{N} \\ \frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T} + \tau\mathbf{B} \\ \frac{d\mathbf{B}}{ds} = -\tau\mathbf{N} \end{cases} \tag{5.21}$$

where $\mathbf{T}$ is the unit vector tangent to the curve, $\mathbf{N}$ is the normal unit vector, $\mathbf{B}$ is the binormal unit vector (cross product of $\mathbf{T}$ and $\mathbf{N}$), d/ds is the derivative with respect to arclength, κ is the curvature, and τ is the torsion. After a resampling of the streamline at 100 equally-spaced points, we compute at each point curvature and torsion. The resulting vector is composed of 200 values, i.e. 100 curvature values and 100 torsion values. Also in this case, streamlines in a bundle must have the same orientation.

It is worth noting that with the above presented streamline representations, different combinations are possible to build an effective feature space[3]. For example, in Figure 5.2 is illustrated a vectorial representation of a streamline with four sets of features: distances from Global Landmarks (d_{GL}), distances from Local Landmarks (d_{LL}), Endpoint-based distances (d_{END}), and ROI-based distances (d_{ROI}).

Therefore, given the tractograms and the expert-made segmented bundles of multiple subjects, we first transform all streamlines into vectors and label them with 1 or 0, to indicate whether or not they belong to the bundle of interest, and then train a classifier to segment a specific bundle, e.g. the corticospinal tract (CST). Notice that, in order to segment different kinds of bundles, it is necessary to train different instances of Classifyber, each with a set of examples of the desired kind of bundle. Afterwards, given a tractogram

---

[3] The effectiveness of different combinations of features is investigated in Section A.2.2

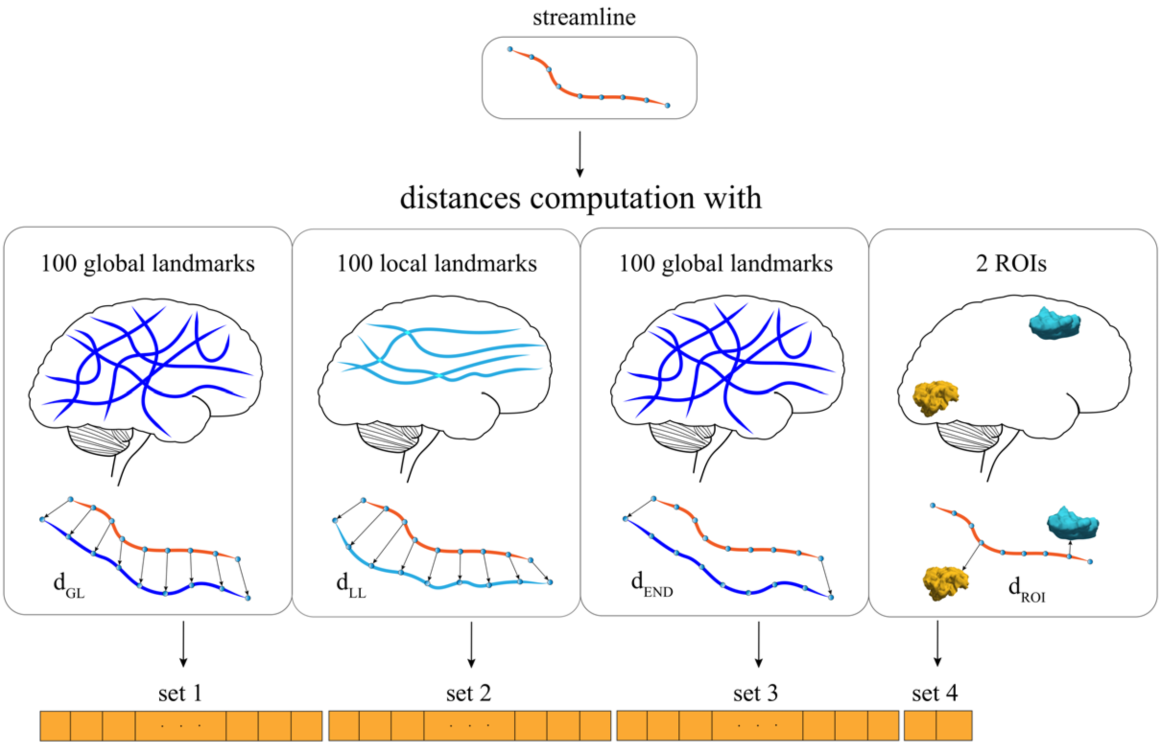


Figure 5.2: Vectorial representation of a streamline. Set 1 and set 2 contain the distances of the streamline with 100 global (d_{GL}) and 100 local (d_{LL}) landmarks, respectively. Set 3 contains the distances (d_{END}) between the endpoints of the streamline and the 100 global landmarks. Set 4 contains the distances (d_{ROI}) between the streamline and the two ROIs pertaining to the bundle of interest.

of a new subject, we predict whether or not the streamlines belong to that bundle. This procedure, which we divided into a *training phase* and a *test phase*, is summarized below and illustrated in Figure 5.4.

5.5.2 Training phase

The training phase is composed of three steps:

Step (a1): Bundle superset

The entire set of streamlines in each tractogram is reduced to a subset of those in the region of the bundle of interest. The main purpose of this reduction is to avoid extremely *imbalanced data*, which decreases the accuracy of classification.

Typically, the ratio between the number of streamlines of a bundle (class 1) and all the other streamlines in the tractogram (class 0) is around 1 : 500, so extremely imbalanced. A typical simple technique to promote effective training is to remove examples far away from the boundary between the two classes and to get a more even class ratio, which is what we obtain by retaining only the streamlines in the region of the bundle of interest. Such operation is computationally intensive, but we adopt the very fast solution described in [Sharmin et al., 2018]. Specifically, the bundle superset of an example bundle is computed by considering the neighboring streamlines belonging to the corresponding tractogram retrieved by a k nearest neighbors (k-NN) procedure applied to each streamline of the bundle. We found $k = 2000$ to be a good compromise between computational cost reduction and size of the resulting superset with respect to the bundle and the tractogram. Usually, with $k = 2000$, the bundle superset, which is a subset of the entire tractogram, is approximately 30 times bigger that the bundle and 20 times smaller than the whole brain tractogram. Efficient k-NN queries are possible due to the use of a K-D tree as explained in Section 5.2. Moreover, this extra cost in time is massively outweighed by the 20x gain in time when computing the next steps, (a2) and (a3).

Then, to each streamline is assigned a class label 1 if it belongs to the bundle, 0 otherwise, see Figure 5.4 (A), where they are represented in green and red respectively.

Step (a2): Feature extraction

Each streamline of the superset is then transformed into a vector, as explained in Section 5.5.1. The choice of the combination of features to build the vectorial representation of the streamlines was the result of a preliminary assessment described in Section A.2.2. Specifically, the combination that gave the best results in terms of classification performances is the one shown in Figure 5.2, i.e. composed of distances from Global Landmarks (d_{GL}), distances from Local Landmarks (d_{LL}), Endpoint-based distances (d_{END}), and ROI-based distances (d_{ROI}).

The entire set of vectors, i.e. the *training set*, is then z-scored independently for each feature.

Step (a3): Training

A binary Logistic Regression (LR) classifier is trained, using the stochastic average gradient (SAG) solver [Schmidt et al., 2017] available in the python package scikit-learn [Pedregosa et al., 2011]. The choice of the classification algorithm and of the solver was the result of a preliminary assessment described in Section A.2.3.

Logistic Regression is a classification algorithm that estimates discrete (categorical) values (binary values like 0/1, yes/no, true/false) on the basis of a given set of independent variables $\mathbf{v} = [v_1, \dots, v_d] \in \mathbb{R}^d$. The function $f(\mathbf{v})$ that maps the relationship between dependent and independent variables is known as the *logistic function*, or *sigmoid function*, see Equation 5.22 and Figure 5.3:

$$f(\mathbf{v}) = \frac{1}{1 + e^{-(a_1 v_1 + \dots + a_d v_d + a_0)}} \tag{5.22}$$

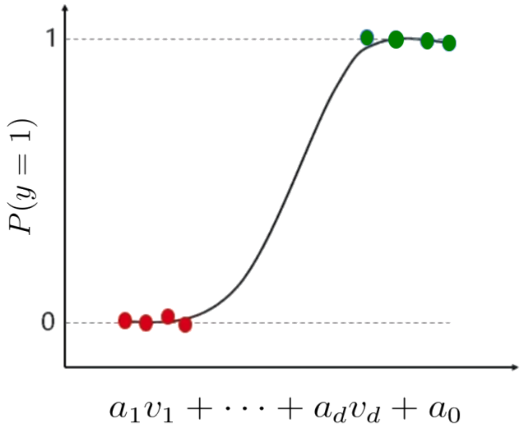


Figure 5.3: Standard sigmoid function that maps the relationship between dependent and independent variables. Red dots are predicted to belong to class 0, while green dots to class 1.

Typically, the coefficients of the classifier $a_0, \dots, a_d$ (weights) are estimated by minimizing the errors in classification on a training set using maximum-likelihood estimation. A powerful optimization technique is the stochastic average gradient (SAG) method [Schmidt et al., 2017], which is particularly useful when dealing with large datasets. The predicted output corresponds to the probability of a certain class and thus ranges from 0 to 1. However, it can be easily transformed into a discrete output by setting a decision threshold.

One crucial aspect when dealing with estimators is the tuning of hyperparameters, i.e. the parameters that are not directly learned from the learning process. Notice that it is common that some of these parameters can largely affect the model performances, while others can be left to their default values. An efficient parameter search strategy is the so called *grid search strategy*, which exhaustively generates candidates from a grid of parameter values. The parameters are then optimized by a cross-validated (CV) grid search over the parameter grid on left out data. Different classification algorithms require different hyperparameters. For example, some of the hyperparameters of the LR classification algorithm are the type of solver, the maximum number of iterations of the solver, and the class weight, i.e. the parameter to lessen the negative effects of the residual class imbalance. In our case, default parameter values are used, except for the maximum number of iterations of the solver, which we increase to 1000 to ensure convergence, as well as the class weight, which we set in all cases to 1:3. These choices are the result of a preliminary investigation using a grid search strategy on left out data and were kept for all the experiments.

5.5.3 Test phase

The test phase is also composed of three steps:

Step (b1): Bundle superset

Similarly to step (a1) of the training phase, we reduce the whole target tractogram to a superset of the target bundle. The main reason for this step is to reduce the computational cost of segmenting the target bundle. Obviously, in this case we do not know the target bundle in advance, so the superset is only *expected* to contain the target bundle, with very high probability. In this case, first, a candidate bundle superset is computed as in step (a1) but considering, in the target tractogram, the neighboring streamlines of one of the *example* bundles. This procedure is then repeated using 5 of the examples bundles. Second, the final bundle superset is obtained by merging all the candidate bundle supersets. Retrospectively, in all experiments, the superset obtained in this way was approximately 40 times larger than the target bundle and always containing all the streamlines of the target bundle. In Figure 5.4 (B), the streamlines of the target superset are depicted in orange because the labels are unknown.

Step (b2): Feature extraction

Each streamline of the bundle superset is embedded into a vector, as described in Section 5.5.1 and as was done in step (a2). All the vectors are z-scored feature-by-feature using means and standard deviations obtained in step (a2) of the training phase.

Step (b3): Test

By exploiting the linear classifier obtained from the training phase in step (a3), each streamline of the superset is predicted to be either part of the bundle (class 1) or not (class 0). Streamline with a probability higher than the decision threshold, which is equal to 0.5 by default, are those composing the predicted bundle in the target tractogram, see Figure 5.4 (B).

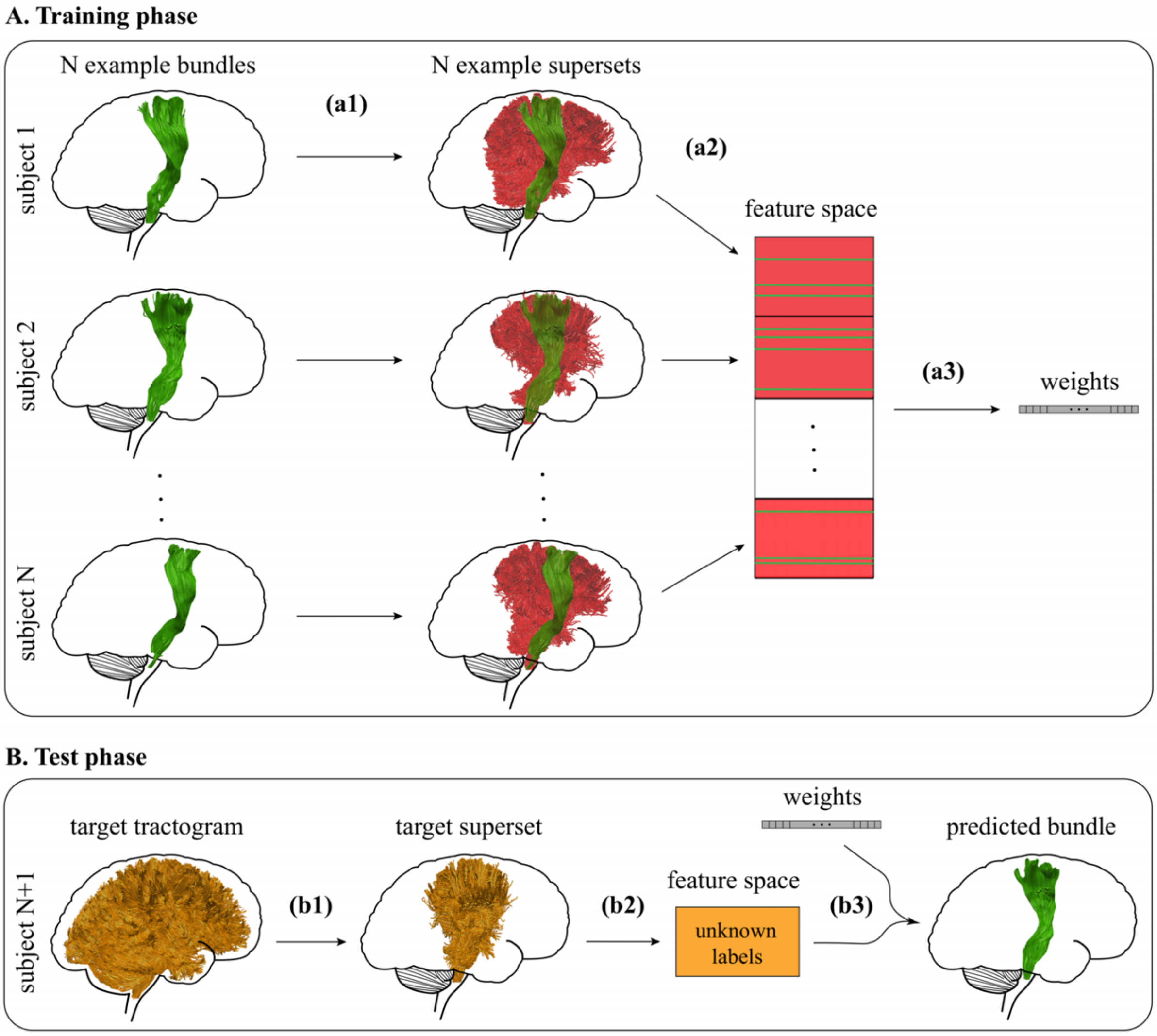


Figure 5.4: Classifyber workflow. A. Schematic illustration of the training phase of Classifyber for a given bundle (CST) over N different subjects. Step (a1): bundle superset. Streamlines belonging to the bundle are depicted in green (class 1), while those not belonging to the bundle are depicted in red (class 0). Step (a2): feature extraction. Step (a3): training of a linear Logistic Regression (LR) classifier. The outcome of this phase is a vector of weights. B. Schematic illustration of the test phase of Classifyber on a single target subject. Step (b1): bundle superset. All the streamlines are depicted in orange because the labels are unknown. Step (b2): feature extraction. Step (b3): test using the resulting weights of the training phase. The outcome of this phase is the predicted bundle (CST) in the target subject.

5.6 Evaluation metrics

5.6.1 Dice Similarity Coefficient (DSC)

To quantitatively evaluate the performance of different segmentation methods we use a procedure commonly adopted in this literature, see for example [Garyfallidis et al., 2018, Sharmin et al., 2018, Wasserthal et al., 2018a]. We compute the degree of voxel overlap between the automatically segmented bundle $\hat{b}$ and the expert-based segmented bundle b, through the Dice Similarity Coefficient (DSC) [Dice, 1945] at the voxel level:

$$DSC = 2\frac{|v(\hat{b}) \cap v(b)|}{|v(\hat{b})| + |v(b)|} \tag{5.23}$$

where $|v()|$ is the number of voxels in a mask. The DSC ranges from 0 to 1 and the more the score is close to 1, the more the two bundles $\hat{b}$ and b are similar.

5.6.2 Receiving Operating Characteristic (ROC) curve

The Receiving Operating Characteristic (ROC) curve analysis is a standard tool for test validation in the field of medical imaging segmentation that plots the sensitivity/specificity curve of the method under evaluation for different cut-off points [Southall et al., 2000].

In a binary classification problem, the outcomes can be labeled either as positive (p) or negative (n). The ROC curve plots the sensitivity, i.e. the true positive rate (TPR), against the specificity, i.e. false positive rate (FPR), by combining four values: the true positives (TP), i.e. the number of times both the predicted outcome and the actual value are p, the false positives (FP), i.e. the number of times the predicted value is p but the actual value is n, the true negatives (TN), i.e. the number of times both the predicted outcome and the actual value are n, and the false negatives (FN), i.e. the number of times the predicted value is n but the actual value is p. TPR and FPR can be formulated as follows:

$$TPR = \frac{TP}{TP + FN} \tag{5.24}$$

$$FPR = \frac{FP}{FP + TN} \tag{5.25}$$

Performances of multiple methods can be compared through the scalar value represented by the Area Under the Curve (AUC). Higher AUC values mean better performances.

5.7 Fractal Dimension (FD)

The concept of *fractal dimension* (FD) [Mandelbrot, 1982] can be used to quantify the degree of irregularity of an object. Specifically, it is an index that quantifies the complexity of an object as a ratio of the change in detail to the change in scale. Fractal analysis has already been applied in multiple fields, such as meteorology, marketing and medicine. With respect to the human brain, it has been used to characterize the brain white matter morphology [Zhang et al., 2006] and multiple sclerosis [Esteban et al., 2007].

Unlike the common notion of dimension, which is called *topological dimension*, the *fractal dimension* can take non-integer values, meaning that a general object fills its space in a qualitatively and quantitatively different manner than a standard geometrical object [Falconer, 2014]. For example, for standard geometrical objects like a straight line, a 2D flat square or a 3D cube, the FD is 1, 2 and 3, respectively, which corresponds with the notion of topological dimensions. However, irregular lines or curves can have FD greater than 1 and asymptotically 2, if their resulting shape is close to a 2D surface. In the same way, a convoluted 2D surface that resembles a 3D object, or a 3D object with several holes, have FD between 2 and 3. For instance, [Zhang et al., 2006] estimated the FD of the 3D voxel mask of the white matter of human brains and obtained values between 2.1 and 2.5.

A standard way to compute the FD of a 3D object is the *box-counting method* [Mandelbrot, 1982, Falconer, 2014], which we used in this work to determine the FD of the voxel mask of white matter bundles. The method is based on the idea of repeatedly covering a given object with boxes of size σ to quantify how the number of boxes changes when σ changes. The FD of the object is then calculated with the following equation, in double-log scale:

$$\mathrm{FD}_{\mathrm{box}} = -\lim_{\sigma \to 0} \frac{\log \mathrm{count}(\sigma)}{\log \sigma} \tag{5.26}$$

where $\mathrm{count}(\sigma)$ is the number of the necessary boxes. As an example, consider the two bundles illustrated in Figure 5.5. To compute their FD, we repeatedly cover their volume with boxes of different sizes (i.e. different side lengths), ranging from 7 to 1 voxel, the minimum possible value corresponding to a single voxel. The relation between the number of boxes required to cover the volume of the two bundles when varying σ is shown in Figure 5.6. The orange and the green lines were obtained by fitting with a regression line the data points corresponding to the two bundles on the left and right side of Figure 5.5, respectively. Their slope corresponds to the FD of their respective voxel masks.

We provide the code to estimate the box-counting dimension (with examples) as open source software at `https://github.com/FBK-NILab/fractal_dimension`.

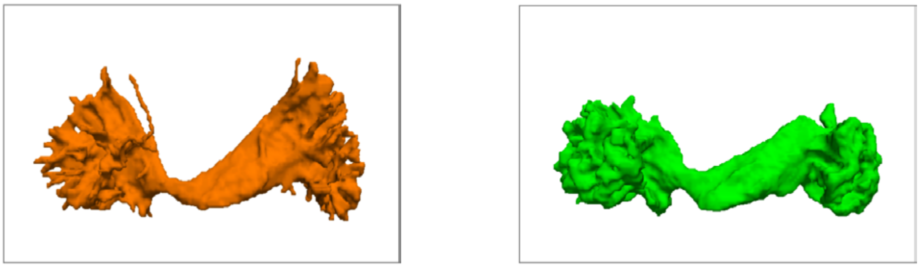


Figure 5.5: Two segmentations of an inferior fronto-occipital fascicles (IFOF) in the same subject: manual segmentation by an expert neuroanatomist with deterministic tractography (left) and semi-automatic segmentation with probabilistic tractography (right).

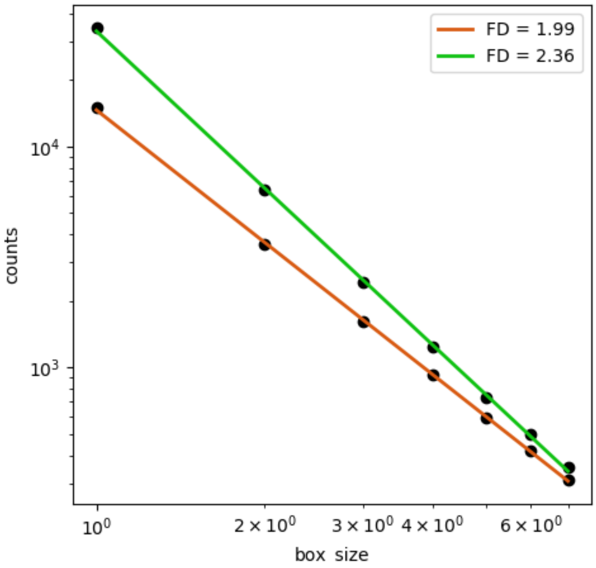


Figure 5.6: Linear regression analysis to compute the fractal dimension (FD) of the voxel masks of the two white matter bundles of Figure 5.5 with the box-counting method. The orange line refers to the bundle on the left side of Figure 5.5, while the green line refers to the bundle on the right side. In both cases, the number of boxes (count) is computed varying the box size from 7 to 1 voxel.

The bundle on the left side of Figure 5.5 has FD $= 1.99$, while the bundle on the right side has FD $= 2.36$. The more a bundle is rounded and smooth, the higher is the FD of its voxel mask. As a matter of fact, when decreasing the box size, the number of boxes required to cover a smooth bundle increases more rapidly rather than to cover a wrinkled bundle.

Chapter 6

Materials

This Chapter is devoted to all the data sources and materials used for the experiments of this thesis. In Section 6.1, we report the Regions of Interest (ROIs) atlas and the parcellation atlas exploited to derive the waypoint ROIs and the terminal ROIs related to the bundles of interest. In Section 6.2, we describe the tractography datasets derived from the Human Connectome Project (HCP), while Section 6.3 describes an in-house tractography clinical dataset acquired at Santa Chiara's Hospital of Trento, Italy.

6.1 ROIs and parcellation atlases

6.1.1 MNI_JHU_tracts_ROIs_atlas

The *MNI_JHU_tracts_ROIs_atlas* is an ROIs atlas composed of the two planar waypoint ROIs that delineate the path of a bundle before it diverges towards the cortex, as defined in [Wakana et al., 2007]. Such ROIs were drawn on a group-average dataset in MNI space and define the path of the following 20 major bundles:

- left and right thalamic radiation (Left_TR and Right_TR)
- left and right corticospinal tract (Left_CST and Right_CST)
- left and right cingulum cingulate (Left_CGC and Right_CGC)
- left and right cingulum hippocampus (Left_CGH and Right_CGH)
- callosum forceps major (FP) and callosum forceps minor (FA)
- left and right inferior fronto-occipital fasciculus (Left_IFOF and Right_IFOF)
- left and right inferior longitudinal fasciculus (Left_ILF and Right_ILF)

- left and right superior longitudinal fasciculus (Left_SLF and Right_SLF)
- left and right uncinate fasciculus (Left_UF and Right_UF)
- left and right arcuate fasciculus (Left_AF and Right_AF)

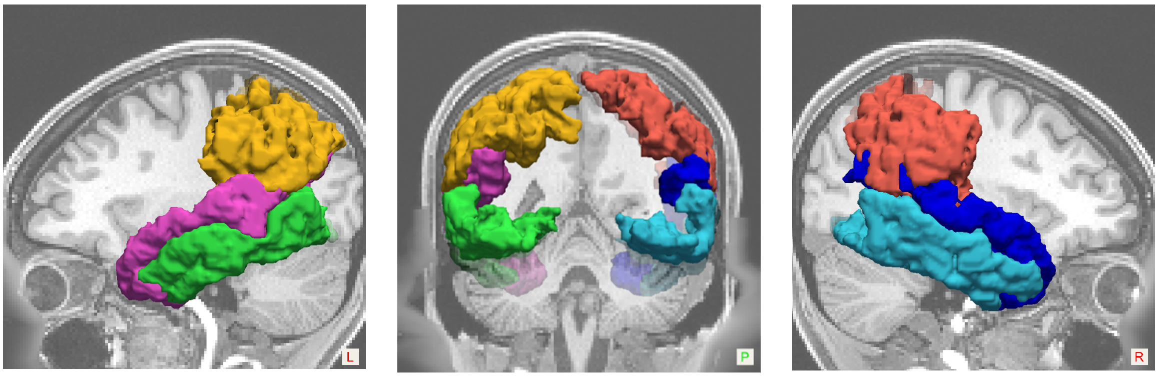
This atlas is used in this thesis to retrieve two ROIs pertaining to each of the major bundles considered in Sections 6.2 and 6.3. Figure 6.3 and Figure 6.4 show an example of the two planar waypoint ROIs associated to the Right_TR, the Left_CST, the Left_AF, the Left_CGC, the Left_CGH and the Right_UF.

The atlas is freely available at `https://github.com/vistalab/vistasoft/tree/master/mrDiffusion/templates/MNI_JHU_tracts_ROIs`.

6.1.2 MNI152_ICBM2009c_reconstructed_atlas

The *MNI152_ICBM2009c_reconstructed_atlas* is a parcellation atlas that is a curated FreeSurfer parcellation of the MNI152 ICBM2009c nonlinear asymmetric template [Gorgolewski, 2016, Fonov et al., 2011]. This atlas is used in this thesis to build the terminal ROIs of the minor bundles considered in Section 6.2.3, as was done in [Bullock et al., 2019]. Specifically, the derived terminal ROIs are six: parietal ROIs left and right, latero-temporal ROIs left and right, and temporal ROIs left and right. These ROIs are illustrated with different colors in an example in Figure 6.1.

The atlas is freely available at `https://figshare.com/articles/FreeSurfer_reconstruction_of_the_MNI152_ICBM2009c_asymmetrical_non-linear_atlas/4223811`.



sagittal view left　coronal view　sagittal view right

Figure 6.1: Terminal ROIs derived from the MNI152_ICBM2009c_reconstructed_atlas in one random HCP subject. Parietal ROIs left (gold) and right (red), latero-temporal ROIs left (magenta) and right (blue), temporal ROIs left (light green) and right (light blue).

6.2 Human Connectome Project (HCP) tractography datasets

This Section presents the description of the three tractography datasets derived from the Human Connectome Project (HCP) dataset that were used in the experimental parts of this thesis. The datasets are presented in chronological order:

1. HCP-EuDX (Section 6.2.1), generated in 2016, and used in the experiments to address the research question **Q1**
2. HCP-IFOF (Section 6.2.2), generated in 2017, and used in the experiments to address the research question **Q3**
3. HCP-ensemble (Section 6.2.3), generated in 2018, and further subdivided into:
    (i) HCP-canonical, used in the experiments to address the research question **Q2** and for the reproducibility study
    (ii) HCP-minor, used in the experiments to address the research question **Q3**

The Human Connectome Project (HCP) is one of the main data sharing initiatives of the 21st century devoted to map the human brain [Van Essen et al., 2013]. The main goal was to create an extensive and accurate database, and to make it freely available to the scientific community. With more than 20 active research studies involving multiple institutions worldwide, the HCP consortium provides the largest font of data to advance the study of anatomical and functional connectivity in healthy subjects, as well as to facilitate understanding of brain disorders such as autism, Alzheimer's disease, schizophrenia and dyslexia. Multimodal acquisitions are provided, including structural MRI, functional MRI (both resting-state and task related), dMRI, Magnetoencephalography (MEG), Electroencephalography (EEG) and a substantial amount of behavioral data. High resolution cleaned data with a high signal-to-noise ratio have been possible thanks to customized acquisition sequences and procedures [Van Essen et al., 2012].

One of the largest releases of data comprises scans of 1200 young adult healthy subjects (age 22-35), acquired in multiple locations. Other collections include data acquired across the lifespan (prenatal, 0-5, 6-21, and 36-100+) and data of a cohort of subjects at risk for, or suffering from, diseases or disorders affecting the brain. In the present thesis, all structural and diffusion MRI data used from HCP belongs to the release of healthy subjects. In particular, we used structural and dMRI HCP data [Sotiropoulos et al., 2013], acquired on a Siemens 3T scanner. Raw diffusion data consists of 3 x 90 gradient directions with b-values = 1000, 2000, and 3000 s/mm^2 and 18 non-weighted volumes, with voxel size

$= 1.25\ mm^3$ isotropic. Anterior Commissure - Posterior Commissure (ACPC) alignment, motion and eddy current correction have been already performed by the HCP consortium [Glasser et al., 2013].

6.2.1 HCP-EuDX dataset

Tractography 30 random HCP subjects. Single shell extraction, 90 gradients with b = 1000 s/mm^2. The reconstruction step was performed using the constrained spherical deconvolution (CSD) algorithm [Tournier et al., 2007] and the tracking step using the Euler Delta Crossing (EuDX) algorithm [Garyfallidis et al., 2012] with 10^6 seeds, approximately 100-140K streamlines.

Bundles Left and right cingulum bundle (Left_CB and Right_CB), left and right inferior fronto-occipital fasciculus (Left_IFOF and Right_IFOF), left and right uncinate fasciculus (Left_UF and Right_UF), part 2 and part 7 of the corpus callosum (CC_2 and CC_7), left arcuate fasciculus (Left_AF).

Expert-based segmentation First, automatic segmentation was performed using White Matter Query Language (WMQL) [Wassermann et al., 2016]. Then, we selected the same bundles chosen in [Olivetti et al., 2016], in which they aimed to avoid extreme variability of the same bundle across subjects, due to the limitations of WMQL. The criterion adopted for the selection was the streamline count, i.e. only bundles with a homogeneous number of streamlines across subjects were retained.

6.2.2 HCP-IFOF dataset

Tractography 30 random HCP subjects. 90 directions, single shell with b=2000 s/mm^2, constraint spherical deconvolution (CSD), deterministic local tracking [Garyfallidis et al., 2014, Berman et al., 2008], step size=0.625 mm, white matter seeding, approximately 500K streamlines.

Bundles Left and right inferior fronto-occipital fasciculus (Left_IFOF and Right_IFOF).

Expert-based segmentation One expert neurosurgeon (A.D.B.) manually segmented the bundles following the guidelines in [Sarubbo et al., 2013] and [Hau et al., 2016], which proposed a classification of the IFOF in different subcomponents. Specifically, the bundle is composed of two layers: the first layer is superficial and antero-superiorly directed, with terminations in the inferior frontal gyrus, while the second layer is deeper and consists of three components (anterior, middle and posterior).

6.2.3 HCP-ensemble dataset

Tractography 192 random HCP subjects. 90 directions, single shell with b=2000 s/mm^2, constraint spherical deconvolution (CSD), ensemble probabilistic tracking [Takemura et al., 2016] with curvature parameters=0.25, 0.5, 1, 2 and 4 mm, step size=0.625 mm, 750K streamlines.

From this tractography dataset, we derived two different bundle datasets: HCP-canonical and HCP-minor.

HCP-canonical

Bundles 20 major bundles listed in Section 6.1 in 130 random HCP subjects.

Expert-based segmentation We built a visually inspected ground truth dataset using a semi-automatic technique as follows. First, from each of the 130 tractograms, we segmented the 20 major associative bundles using the Automated Fiber Quantification (AFQ) algorithm [Yeatman et al., 2012]. A first cleaning procedure is applied bundle-wise to remove streamlines that deviate more than 5 standard deviations from the bundle centroid or that are more than 4 standard deviations above the mean streamline length, as was done in [Yeatman et al., 2012]. Then, in order to have consistent bundles across subjects, given a bundle, we identified those that do not deviate more than the 20% from the median number of streamlines, obtaining on average 50 segmentations per bundle. Finally, we visually inspected each segmentation and filtered out the outliers in order to have 30 segmentations per bundle. We then selected 12 bundles (6 left and 6 right) per subject, which we subdivided into two groups based on their streamline count[1], see Figure 6.2:

- *HCP-canonical-large*: left and right thalamic radiation (Left_TR and Right_TR), left and right corticospinal tract (Left_CST and Right_CST), left and right arcuate fasciculus (Left_AF and Right_AF), see some examples in Figure 6.3.
- *HCP-canonical-small*: left and right cingulum cingulate (Left_CGC and Right_CGC), left and right cingulum hippocampus (Left_CGH and Right_CGH), left and right uncinate fasciculus (Left_UF and Right_UF), see some examples in Figure 6.4.

---

[1] This is not an absolute definition of small and large bundles, but only a relative definition within the group of bundles considered in this work.

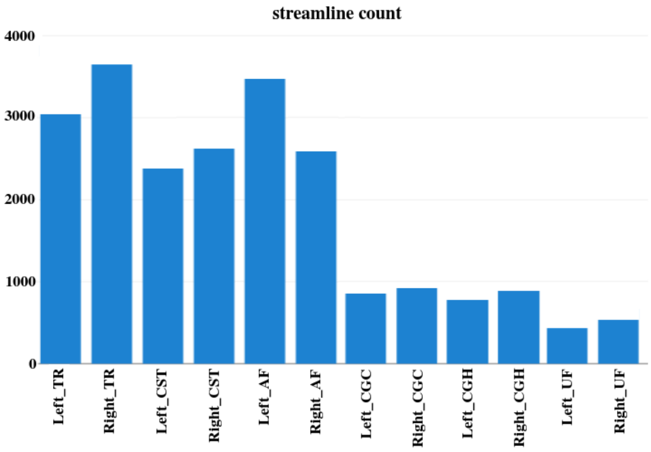
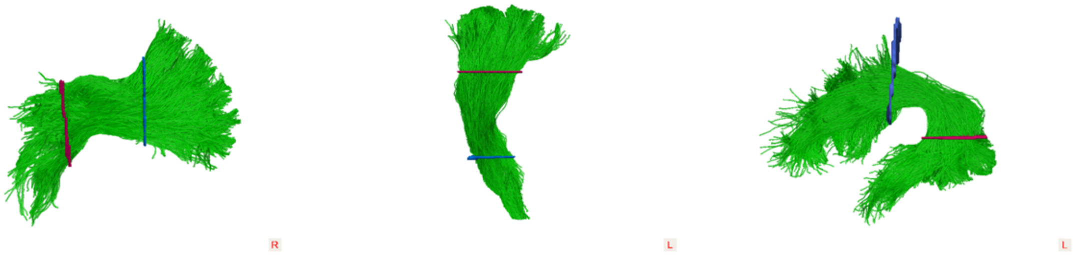


Figure 6.2: Streamline count of the 12 bundles selected in the HCP-canonical dataset in one random HCP subject. The 6 bundles with a higher number of streamlines are considered ”large” (HCP-canonical-large, depicted on the left side), while the other 6 bundles are considered ”small” (HCP-canonical-small, depicted on the right side).



Right_TR Left_CST Left_AF

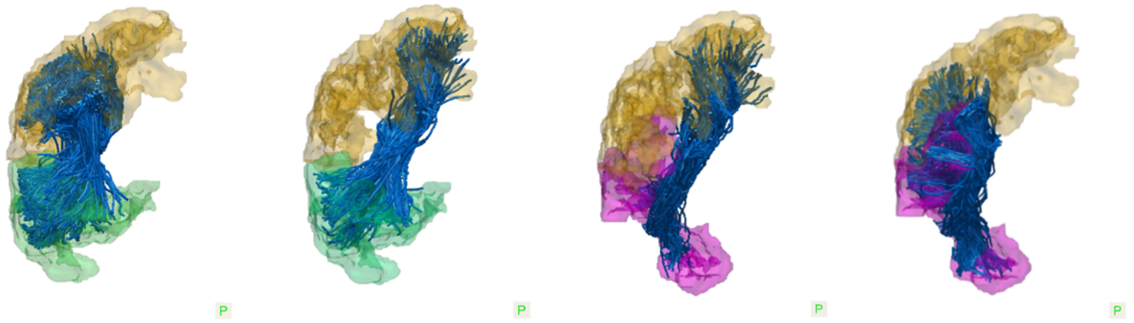
Figure 6.3: Three bundles of the dataset HCP-canonical-large with their associated ROIs that define their path before they diverge towards the cortex.

Left_CGC Left_CGH Right_UF

Figure 6.4: Three bundles of the dataset HCP-canonical-small with their associated ROIs that define their path before they diverge towards the cortex.

HCP-minor

Bundles Left and right posterior arcuate fasciculus (Left_pArc and Right_pArc), left and right temporo-parietal connection to the superior parietal lobule (Left_TP-SPL and Right_TP-SPL), left and right middle longitudinal fasciculus–superior parietal lobule component (Left_MdLF-SPL and Right_MdLF-SPL), left and right middle longitudinal fasciculus–superior angular gyrus component (Left_ MdLF-Ang and Right_MdLF-Ang), see some examples in Figure 6.5, with their streamline count in Figure 6.6.



Left_pArc Left_TP-SPL Left_MdLF-SPL Left_MdLF-Ang

Figure 6.5: Examples of four bundles of the HCP-minor dataset with their associated terminal ROIs. The Left_pArc and the Left_TP-SPL connect the left parietal ROI with the left temporal ROI, while the Left_MdLF-SPL and the Left_MdLF-Ang connect the left parietal ROI with the left latero-temporal ROI.

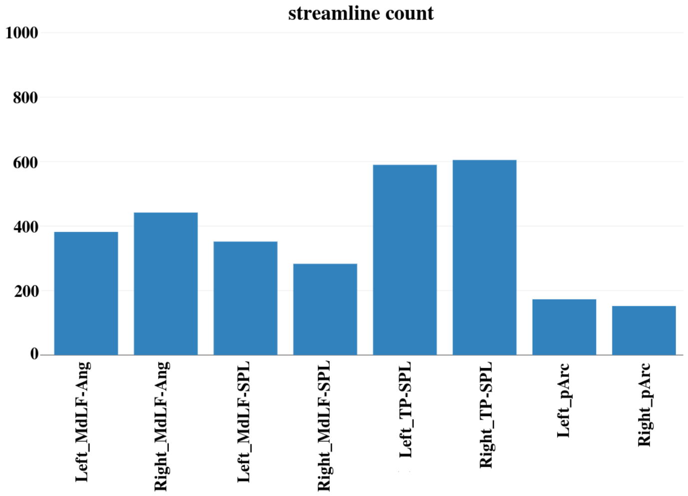


Figure 6.6: Streamline count of the 8 minor bundles of the HCP-minor dataset in one random HCP subject.

Expert-based segmentation We obtained the segmentations of 192 randomly selected HCP subjects using the procedure proposed in [Bullock et al., 2019]. We then filtered out segmented bundles that for any reason were not considered plausible from the neuroanatomical point of view with a semi-automatic technique, as follows. First, we automatically discarded those subjects which had at least one bundle that deviated more than ±2 standard deviations from the mean of the bundle distribution of the number of voxels and number of streamlines of the population of the 192 subjects. After this step, the number of subjects retained was 121. Then, an expert (D.B.) performed visual inspection of each individual bundle to detect anomalies in the segmentations, with the following procedure. For each subject and for each bundle, an anatomical plot of the streamlines corresponding to the relevant bundle was obtained. These streamline bundle models were overlain on an anatomical T1 background for reference, and were viewed from an axial, sagittal, and coronal perspective (slice number 48, 74 and 85 respectively). Bundles were assigned an omnibus score corresponding to an equal weighting of the characteristics of anatomical typicality and overall fullness. These scores ranged from 1 to 5 such that 1 indicated a rating of *bad*, 2 indicated a rating of *poor*, 3 indicated a rating of *OK*, 4 indicated a rating of *good*, 5 indicated a rating of *great*. *Anatomical typicality* was stipulated to refer to the degree to which the bundle manifested the typical connectivity motif associated with prior characterizations of the bundle (see [Bullock et al., 2019] for an overview of the relevant literature for each bundle). *Fullness* was stipulated to refer to the degree to which the bundle's constituent streamlines fully and accurately occupied the volume of white matter described or depicted in prior characterizations of the bundle. Finally, we kept those subjects whose *all* bundles obtained a score of 2 or higher, remaining with a total of 105 subjects.

6.3 Clinical tractography dataset

In this Section, we describe the in-house clinical dataset built from patients with brain tumors that we used in the experiments to address the research question **Q3**. Acquisitions were done between 2015 and 2017 at Santa Chiara's Hospital of Trento, Italy, 1.5T scanner, image resolution 0.9 x 0.9 x 2.4 mm, 60 gradient directions with b-value=1000 s/mm^2 and 1 volume with b=0. Data were corrected for eddy-current and motion, and an additional step of rescaling was applied to obtain an isotropic voxel resolution of 2 mm. In Table 6.1 we report the clinical data of the 10 patients considered (mean age 44.4 ± 15.5 years old).

case	age	type of lesion	segmented bundle(s)
A	36	diffuse astrocytoma, grade II	Left_IFOF
B	56	gliosarcoma, grade IV	Left_AF
C	29	cavernous hemangioma	Left_AF
D	23	ganglioglioma, grade I	Left_IFOF, Left_AF
E	37	cavernous hemangioma	Left_IFOF
F	32	glioblastoma, grade IV	Left_IFOF, Left_AF
G	39	cavernous hemangioma	Left_IFOF
H	59	glioblastoma, grade IV	Left_IFOF, Left_AF
I	60	glioblastoma, grade IV	Left_IFOF, Left_AF
J	73	adenocarcinoma metastasis	Left_AF

Table 6.1: Clinical data of the 10 patients of the Clinical dataset.

Tractography 10 patients with brain tumor. 60 directions, single shell b=1000 s/mm^2, diffusion tensor imaging (DTI) reconstruction, Euler Delta Crossing (EuDX) tracking method [Garyfallidis et al., 2014], 10^6 seeds, approximately 100K streamlines.

Bundles Left inferior fronto-occipital fasciculus (Left_IFOF) and left arcuate fasciculus (Left_AF).

Expert-based segmentations One expert neurosurgeon (S.S.) manually segmented the bundles in the lesioned hemisphere of the patients, who were affected by brain tumor, see Table 6.1 for further details. The lesion however did not affect the shape of the bundles consistently. Bundles where successively refined to remove outliers using the interactive segmentation tool Tractome [Porro-Muñoz et al., 2015] and visually inspected, remaining with 7 instances for each bundle.

Figure 6.7: Two examples of expert-based segmented Left_AF (left) and Left_IFOF (right) in two random patients of the Clinical dataset.

Chapter 7

Experiments and Results

In this Chapter, we report the main[1] experiments conducted and their associated results in order to address each of the three main research questions of this thesis, which were summarized in Section 4.2.4. Specifically:

- Section 7.1 addresses the research question **Q1**, and its content was published in "*Comparison of Distances for Supervised Segmentation of White Matter Tractography*" [Olivetti et al., 2017].
- Section 7.2 addresses the research question **Q2**, and its content was published in "*Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation*" [Bertò et al., 2019].
- Section 7.3 addresses the research question **Q3**, and its content "*Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation*" is under review in a scientific journal.

7.1 Comparison of streamline distances for bundle segmentation

This Section aims to answer to the research question **Q1**: Which streamline distance function available in the literature is the most effective for supervised bundle segmentation?

As discussed in Section 4.2.1, several streamline distance functions have been proposed in the literature. However, to the best of our knowledge, no comparison has be made in

---

[1] For further details and results, refer to Part II, which comprises the main scientific papers published, under review and in preparation resulting from our work.

order to determine which one works better for supervised bundle segmentation.

Here, we aim to address the gap in the literature by comparing 8 different distance functions available in the literature and by quantifying their performances within a supervised segmentation task. To this end, following the ideas in [Yoo et al., 2015, Sharmin et al., 2016], we adopted the supervised bundle segmentation framework, where the bundle of interest is automatically segmented in a target subject by leveraging an example of that bundle segmented by an expert in a different subject.

7.1.1 Experimental design

We ran multiple experiments on 10 subjects[2] randomly selected from the HCP-EuDX dataset that we described in Section 6.2.1, which is composed of tractograms and expert-based segmented bundles. The 9 kinds of bundles considered for each subject were: left and right cingulum bundle (Left_CB and Right_CB), left and right inferior fronto-occipital fasciculus (Left_IFOF and Right_IFOF), left and right uncinate fasciculus (Left_UF and Right_UF), part 2 and part 7 of the corpus callosum (CC_2 and CC_7), and left arcuate fasciculus (Left_AF). The 8 streamline distance functions compared were: d_{MC}, d_{SC}, d_{LC}, $d_{\mathrm{MDF},12}$, $d_{\mathrm{MDF},20}$, $d_{\mathrm{MDF},32}$, d_{PDM} and $d_{\mathrm{varifolds}}$, whose mathematical formulation was given in Section 5.1.1. We took into account three variants of the d_{MDF}, which differ from the number of points (12, 20 or 32) used to resample the streamlines. For d_{PDM} and $d_{\mathrm{varifolds}}$ we set $\sigma = 42mm$, according to [Siless et al., 2013].

Specifically, the experiments were conducted as follows. First, each pair of tractograms was co-registered using the streamline linear registration (SLR) algorithm [Garyfallidis et al., 2015]. Then, we performed bundle segmentation using the Nearest Neighbor (NN) procedure as described in Section 5.2, across the 90 pairs of subjects, for each of the 9 different bundles, and each of the 8 different distance functions. Fast NN queries were possible due to the use of a K-D tree on the vectorial representation of the streamlines computed using the dissimilarity representation as explained in Section 5.2. To build such representation, according to [Olivetti et al., 2012], we selected in each target tractogram 40 landmarks with the subset farthest first (SFF) policy. Finally, following a common practice to evaluate the bundle segmentation performances, we measured the degree of overlap between the estimated bundle and the ground truth bundle through the Dice Similarity Coefficient (DSC) at the voxel level, see Section 5.6.

The experiments were developed in Python code, on top of DIPY[3]. The code of

---

[2] Each subject was considered both as example and target.

[3] `http://nipy.org/dipy` [Garyfallidis et al., 2014].

all experiments is available under a Free/OpenSource license at `http://github.com/emanuele/prni2017_comparison_of_distances`.

7.1.2 Results

In Table 7.1, we report the mean DSC obtained with the NN bundle segmentation method across the 90 pairs of subjects, for each of the 9 different bundles, and each of the 8 different distance functions considered. For each bundle and distance function, we observe a standard deviation of DSC of approximately 0.10, which corresponds to a standard deviation of the mean of 0.01. This value includes the variances due to: the anatomical variability across subjects, the limitations of the White Matter Query Language (WMQL) segmentation technique used as ground truth and, in minor part, the approximation introduced by the dissimilarity representation.

	d_{MC}	d_{SC}	d_{LC}	$d_{\mathrm{MDF},12}$	$d_{\mathrm{MDF},20}$	$d_{\mathrm{MDF},32}$	d_{PDM}	$d_{\mathrm{varifolds}}$
Left_CB	0.61	0.60	0.59	0.59	0.59	0.59	0.59	0.56
Right_CB	0.60	0.59	0.58	0.58	0.57	0.58	0.57	0.55
Left_IFOF	0.49	0.48	0.47	0.48	0.48	0.47	0.48	0.49
Right_IFOF	0.47	0.46	0.45	0.45	0.45	0.45	0.46	0.44
Left_UF	0.52	0.54	0.55	0.52	0.52	0.53	0.57	0.60
Right_UF	0.49	0.52	0.51	0.49	0.49	0.49	0.52	0.56
CC_7	0.58	0.56	0.61	0.64	0.63	0.63	0.59	0.67
CC_2	0.49	0.50	0.52	0.53	0.53	0.54	0.57	0.59
Right_AF	0.51	0.49	0.51	0.51	0.50	0.50	0.52	0.50
means	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.53**	**0.54**	**0.55**

Table 7.1: Mean DSC voxel table across the 90 pairs of subjects, for each of the 9 different bundles (rows), and each of the 8 different distance functions (columns) considered.

In Table 7.2, we report the time (in seconds) required by a modern desktop computer to compute 90000 pairwise streamline distances using the 8 distance functions considered. The differences in time are due to both the different computational cost of the formulas in Section 5.1.1 and their implementation. d_{MC}, d_{SC}, d_{LC} and d_{MDF}, available from DIPY, were implemented in Cython. d_{PDM} and $d_{\mathrm{varifolds}}$ were implemented by us in Python and NumPy[4].

	d_{MC}	d_{SC}	d_{LC}	$d_{\mathrm{MDF},12}$	$d_{\mathrm{MDF},20}$	$d_{\mathrm{MDF},32}$	d_{PDM}	$d_{\mathrm{varifolds}}$
time (seconds)	0.5	0.5	0.5	0.03	0.04	0.05	16	28

Table 7.2: Computational time to compute 90000 pairwise streamline distances.

---

[4] `http://www.numpy.org`.

These results show that there are no major differences in the segmentation accuracy, measured as DSC, when using different distance functions. However, the computational times of the distance functions can be very different.

7.2 Anatomically-Informed Multiple Linear Assignment Problems (LAP-anat) for bundle segmentation

This Section aims to answer to the research question **Q2**: Why not including prior anatomical information in a streamline-based bundle segmentation method?

As discussed in Section 4.2.2, current methods for bundle segmentation consider either only geometrical properties of the bundle of interest, or only prior knowledge about its relative anatomical position.

We aim to improve the results of bundle segmentation by considering information about both the shape of the streamlines and about the relative anatomical position of the bundle of interest. To this end, we propose an extension of the bundle segmentation method based on the Linear Assignment Problem (LAP) proposed in [Sharmin et al., 2018], which has proved to outperform the methods based on the NN procedure. Specifically, our goals are (i) to include additional anatomical information within the optimization process, hence showing that it is possible to combine the best of connectivity-based and streamline-based methods; (ii) to show that large bundles can be accurately segmented more easily than small bundles, and that anatomical information helps in reducing such difference.

7.2.1 Experimental design

We ran multiple experiments on the 130 subjects[5] of the HCP-canonical dataset that we described in Section 6.2.3, which is composed of tractograms and expert-based segmented bundles. Notice that the 12 segmented bundles were divided into two groups based on their streamline count: the *HCP-canonical-large* group comprised left and right thalamic radiation (Left_TR and Right_TR), left and right corticospinal tract (Left_CST and Right_CST), left and right arcuate fasciculus (Left_AF and Right_AF), while the *HCP-canonical-small* group comprised left and right cingulum cingulate (Left_CGC and Right_CGC), left and right cingulum hippocampus (Left_CGH and Right_CGH), left and right uncinate fasciculus (Left_UF and Right_UF).

---

[5] Each subject could potentially be used both as example and target.

We conducted multiple experiments using the LAP method of [Sharmin et al., 2018] (LAP), described in Section 5.3, and the proposed method (LAP-anat), described in Section 5.4, on a total of 360 segmented bundles. In both cases, each pair of tractograms were aligned with an affine registration using the streamline linear registration (SLR) algorithm [Garyfallidis et al., 2015]. We used an example set composed of 5 bundles, since it was proved that a larger example set has no substantial impact on the final result of the segmentation [Sharmin et al., 2018]. In the LAP-anat method, we added the two anatomically-informed distance matrices, i.e. the endpoint-based distance matrix E and the ROI-based distance matrix R, to the original cost matrix D, as explained in Section 5.4. The parameters λ_D, λ_E and λ_R were set in order to let all the values of the matrices span in the same range (which would approximately correspond to $\lambda_D = 1$, $\lambda_E = 0.4$ and $\lambda_R = 1.6$). To build the ROI-based distance matrix, for each bundle, we considered the two waypoint ROIs defined in the MNI_JHU_tracts_ROIs_atlas, see [Wakana et al., 2007] and Section 6.1, which we transferred into the individual subject space through a non-linear registration using the Advanced Normalization Tool (ANTs) [Avants et al., 2008]. We then compared the performances of the two methods through the Dice Similarity Coefficient (DSC) at the voxel level, see Section 5.6.

All the experiments were developed in Python code and ran using cloud computing resources provided by brainlife.io. To support scientific reproducibility and openness, code and dataset are freely available for reproducibility at `https://doi.org/10.25663/brainlife.app.122` and `https://doi.org/10.25663/brainlife.pub.3` respectively.

Additional experiments were performed in order to assess the individual contribution of the three cost matrices and to quantify the different performances when using tractograms aligned linearly with SLR or non-linearly with ANTs, see Section A.1.

7.2.2 Results

In Figure 7.1 we report the mean DSC results for the two methods that we compared across 30 subjects, both for large bundles (HCP-canonical-large, depicted on the left side of the graph) and for small bundles (HCP-canonical-small, depicted on the right side of the graph). For the large bundles, both methods registered a standard deviation of the mean ≤ 0.005. For the small bundles, with the LAP method we observed a standard deviation of the mean between 0.009 and 0.015, and with the proposed LAP-anat method between 0.007 and 0.011. Figure 7.2 shows a paradigmatic example in which the LAP method fails to retrieve most of the streamlines terminating in the cortical areas, which instead are correctly identified by the proposed LAP-anat method.

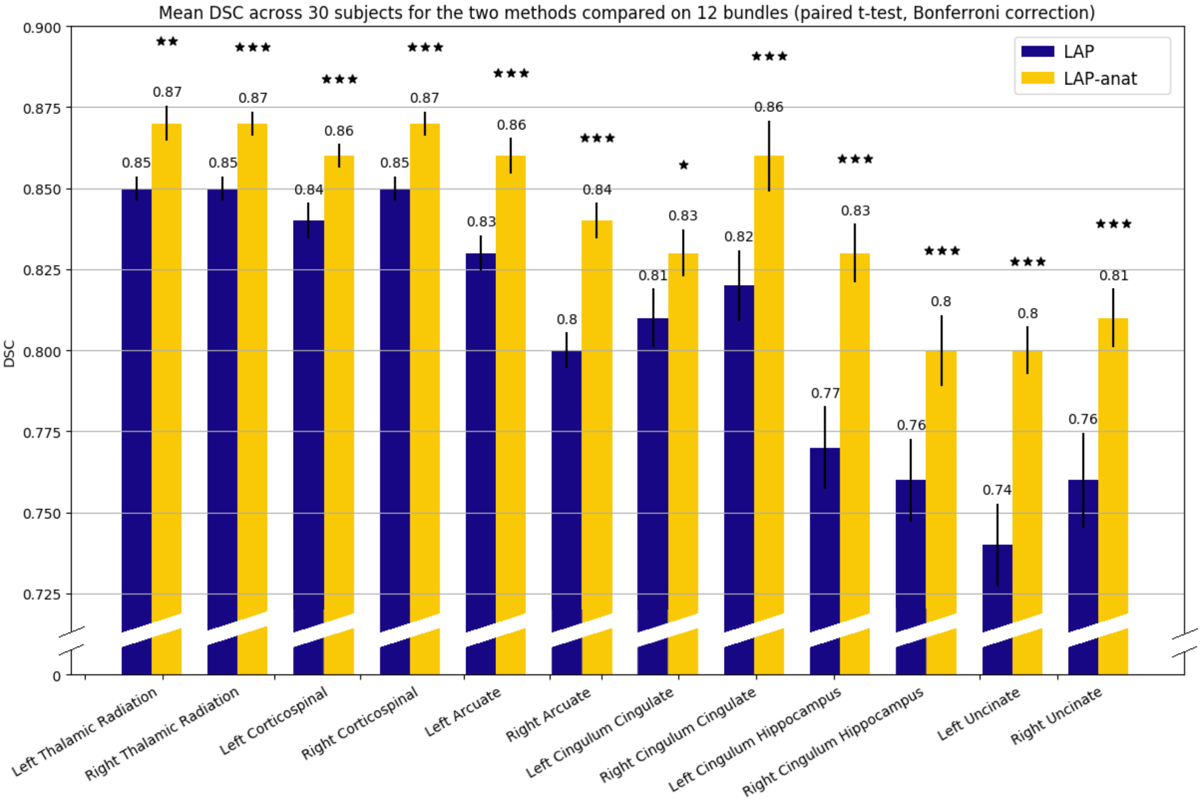


Figure 7.1: Mean DSC and standard deviation of the mean obtained across 30 subjects with the original method (LAP, depicted in blue) an the proposed method (LAP-anat, depicted in yellow). *** means $p < 0.0001$, ** means $p < 0.001$ and * means $p < 0.05$. Bundles belonging to the group HCP-canonical-large are depicted on the left side, while bundles belonging to the group HCP-canonical-small are depicted on the right side.

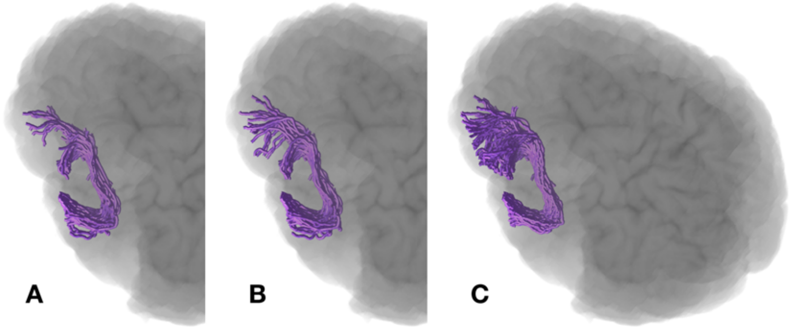


Figure 7.2: Comparative paradigmatic example of a segmented uncinate fasciculus (UF) obtained with A) the original method (LAP), B) the proposed method (LAP-anat), and C) the ground truth.

These results illustrate that, on average, the proposed LAP-anat method outperforms the LAP method of [Sharmin et al., 2018], for all the bundles considered. For the large bundles we observe a mean improvement in the DSC score of +0.025, while for the small bundles of +0.045. With LAP-anat, in all the cases we obtained a mean DSC between 0.80 and 0.87, which means that the overlap with the ground truth is at least 80%.

The results of the additional experiments are presented in Section A.1.

7.3 Classifyber, a robust streamline-based linear classifier for bundle segmentation

This Section aims to answer to the research question **Q3**: Can we develop a novel robust streamline-based bundle segmentation method which is able to adapt to multiple data settings?

As discussed in Section 4.2.3, most of the state-of-the-art supervised methods suffer from some important drawbacks, such as dependency from multiple thresholds, scalability issues, and inability to adapt to different data settings.

To overcome these limitations, we propose *Classifyber*, a new streamline-based method for white matter bundle segmentation. The core algorithm of Classifyber is a single streamline linear classifier, and thus is has multiple advantages: linear methods in fact are generally well known and easy to understand, they are very fast and require minimal resources, their software implementations are commonly available and, as opposed to non-linear methods, they can be interpreted. Our main goals with the work presented here are to segment multiple kinds of white matter bundles using different automatic segmentation methods, and to prove that the proposed method, Classifyber, substantially increases the quality of segmentation with respect to other methods, consistently across different datasets. Further investigations about the fractal dimension of bundles, the size of the training set, and the computational time are also presented.

7.3.1 Experimental design

The experiments were conducted on four different tractography datasets composed of tractograms and expert-based segmented bundles, three of which are novel: HCP-minor, HCP-IFOF, and Clinical datasets, described in Sections 6.2.3, 6.2.2, and 6.3 respectively. The fourth dataset[6], which we called HCP-major dataset, is (part of) the dataset used in

---

[6] Available at `https://doi.org/10.5281/zenodo.1477956`.

the experiments of [Wasserthal et al., 2018a]. This dataset is composed of probabilistic tractograms with 10 million streamlines of 105 HCP subjects, with a collection of 72 bundles segmented through a semi-automatic procedure, see [Wasserthal et al., 2018a].

For the three HCP datasets, we computed the non-linear warp to register the structural T1-weighted images of every subject of each dataset to the MNI152 T1 template using the Advanced Normalization Tool (ANTs) [Avants et al., 2008]. The choice of the registration method was the result of a preliminary assessment described in Section A.2.1. For the clinical dataset, we computed a streamline linear registration (SLR) to the whole brain template of [Yeh et al., 2018][7], because non-linear registration of clinical data is debated, see [Garyfallidis et al., 2015]. In both cases, we applied the registrations to tractograms and bundles. All ROIs derived from the atlases described in Section 6.1 were registered as well to the MNI152 T1 template with a linear registration using the FMRIB's Linear Image Registration Tool (FLIRT) [Jenkinson et al., 2002].

Below we describe the experimental setup of Classifyber and of the state-of-the-art automatic segmentation methods that we considered in our comparison[8], together with their necessary variants. For each dataset, the entire pool of subjects was randomly divided into two groups: the *training set* and the *test set*. Bundles of the training set are used as examples to learn from, while bundles of the test set are used to assess the performance of the different methods. Notice that the exact same test sets were kept for all the methods compared.

Classifyber In order to build the feature space of Classifyber, we retrieved the ROIs pertaining to each bundles from the available atlases described in Section 6.1. For the dataset HCP-minor, the two ROIs considered for each bundle are the two terminal ROIs, i.e. the cortical regions that the bundle of interest connects, derived from [Bullock et al., 2019]. Specifically, the MdLF-Ang and MdLF-SPL connect the parietal region to the lateral-temporal region, while the TP-SPL and pArc connect the parietal region to the temporal region. Each region was built by merging specific cortical parcellations of the MNI152_ICBM2009c_reconstructed_atlas. For the other three datasets, the ROIs considered are the two planar waypoint ROIs defined in the MNI_JHU_tracts_ROIs_atlas. Bundle segmentation was performed as described in detail in Section 5.5.

---

[7] Available at `http://brain.labsolver.org`.

[8] The bundle segmentation methods compared here are those that in the literature have shown to achieve the best segmentation performances, as discussed in Section 4.2.3.

TractSeg We adopted the openly available pretrained network of [Wasserthal et al., 2018a], which was trained on 84 subjects, and tested it on the dMRI data of the target subjects. We used the default parameters and the postprocessing option, which removes holes and isolated voxels in the predicted voxel mask of the bundles.

TractSeg-retrained When the bundle to be segmented was not available among those covered by TractSeg, we re-trained the FCNN on new examples with a procedure discussed in a private communication with the authors of [Wasserthal et al., 2018a]. We refer to this variant as *TractSeg-retrained*. First, we trained a single FCNNs per dataset with default parameters, 250 epochs, fraction of validation subjects = 0.2 and data augmentation. Then, we tested the method enabling the postprocessing option.

RecoBundles-atlas We contacted the authors of [Garyfallidis et al., 2018] and received the indication to use the bundle models provided by the atlas of [Yeh et al., 2018][9] as the example bundles, and specifically 30 (out of 80) of them. We denote as *RecoBundles-atlas* this use of the RecoBundles algorithm. We used the best configuration of parameter values found from an extensive preliminary assessment. This configuration uses default parameter values with the exception of disabling the local SLR option (most probably because all the datasets were already coregistered in MNI space) and using the minimum average mean distance (d_{MAM}) instead of the minimum average direct flip distance (d_{MDF}).

RecoBundles When the bundle to be segmented is not available among the 30 selected bundles from the atlas of [Yeh et al., 2018], we fell back to the original indication in [Garyfallidis et al., 2018] and used the same example bundles adopted as input for the other methods. We denote this use of the algorithm plainly as *RecoBundles*. Due to the fact that RecoBundles accepts only one bundle as example, to quantify the quality of segmentation when multiple bundles are available in the training set, we adopted a procedure similar to the one used in the experiments of [Wasserthal et al., 2018a]. Specifically, we treated the N example bundles as models for N separate runs of the algorithm over the target subject, thus obtaining N different predictions of the same bundle. We then evaluated the segmentation accuracy by computing the mean DSC across the N bundles. As for RecoBundles-atlas, we used the best configuration of parameter values found from an extensive preliminary assessment.

---

[9] Available at `http://brain.labsolver.org`.

LAP We ran the algorithm following the original procedure in [Sharmin et al., 2018] and we set the parameter k, the only parameter of the method, corresponding to the number of nearest neighbors streamlines to compute the superset, equal to 2000 (default $k = 500$), since the the total number of streamlines of the tractograms considered in our experiments are approximately 4 times higher than those in the original study.

LAP-anat We ran the algorithm following the original procedure in [Bertò et al., 2019] and, as we did for LAP, we set $k = 2000$. The parameters of λ_D, λ_E and λ_R were set in order to let all the values of the matrices span in the same range, as was done in [Bertò et al., 2019]. The pairs of ROIs considered for each bundle were the same used for Classifyber.

7.3.2 Experiments and results on HCP-minor dataset

From the HCP-minor dataset, we considered only subjects for which all bundles received an expert-made score of 3 or higher, according to the procedure explained in Section 6.2.3, resulting in a set of 40 subjects. We randomly split this pool of subjects in 15 for training and 25 for test. We conducted multiple experiments as described in Section 7.3.1, and then evaluated the segmentation performances through the DSC score, see Section 5.6.

In Table 7.3 and Figure 7.3, we quantify the mean quality of segmentation in terms of DSC across the minor bundles considered in this set of experiments for RecoBundles, TractSeg-retrained, LAP, LAP-anat, and Classifyber across 25 subjects. TractSeg and RecoBundles-atlas were excluded because they do not address minor bundles. The quality of segmentation obtained by Classifyber is very high and outperforms all the other methods.

	RecoBundles	TractSeg-ret.	LAP	LAP-anat	Classifyber
Right_pArc	0.76 ± 0.04	0.77 ± 0.03	0.80 ± 0.03	0.83 ± 0.03	$\mathbf{0.88} \pm 0.03$
Left_MdLF-Ang	0.71 ± 0.04	0.72 ± 0.06	0.79 ± 0.05	0.80 ± 0.07	$\mathbf{0.87} \pm 0.03$
Left_pArc	0.73 ± 0.05	0.75 ± 0.03	0.79 ± 0.04	$\mathbf{0.81} \pm 0.05$	$\mathbf{0.85} \pm 0.05$
Right_MdLF-Ang	0.68 ± 0.04	0.70 ± 0.03	0.76 ± 0.03	0.79 ± 0.03	$\mathbf{0.84} \pm 0.03$
Left_MdLF-SPL	0.63 ± 0.06	0.67 ± 0.05	0.73 ± 0.04	0.76 ± 0.04	$\mathbf{0.82} \pm 0.04$
Right_TP-SPL	0.62 ± 0.08	0.68 ± 0.06	0.72 ± 0.05	0.76 ± 0.06	$\mathbf{0.82} \pm 0.05$
Left_TP-SPL	0.63 ± 0.06	0.67 ± 0.04	0.70 ± 0.05	0.75 ± 0.04	$\mathbf{0.81} \pm 0.04$
Right_MdLF-SPL	0.60 ± 0.05	0.64 ± 0.04	0.70 ± 0.03	0.73 ± 0.04	$\mathbf{0.80} \pm 0.04$

Table 7.3: Quantitative comparison over HCP-minor dataset: DSC (mean ± sd) across 25 target subjects for RecoBundles, TractSeg-retrained, LAP, LAP-anat and Classifyber. Highest quality of segmentation in bold face.

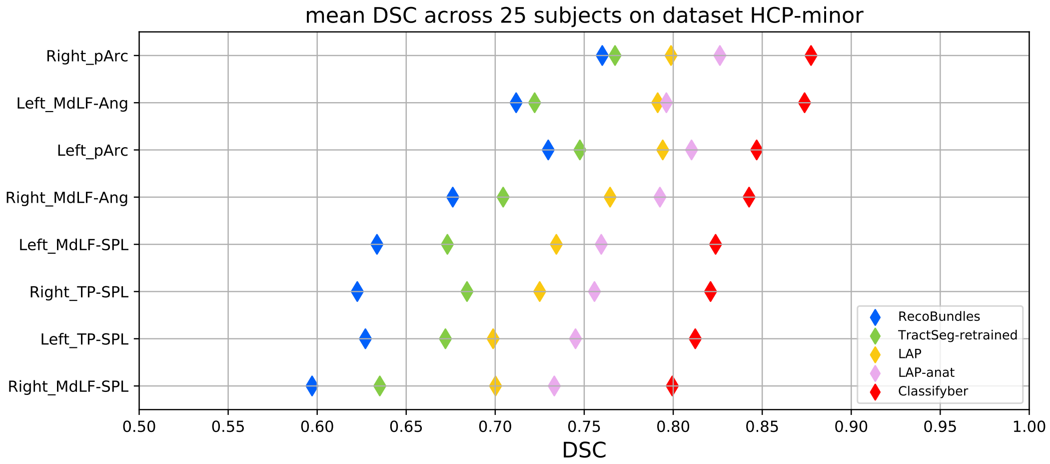


Figure 7.3: Quantitative comparison over HCP-minor dataset: mean DSC across 25 target subjects for RecoBundles (blue), TractSeg-retrained (light green), LAP (yellow), LAP-anat (pink) and Classifyber (red).

When using other methods, we observe a consistent bias in the predictions: frequently RecoBundles tends to overestimate the bundle producing several false positives streamlines, while TractSeg-retrained correctly identifies the core part of the bundles, but fails to retrieve part of the cortical terminations. An illustrative example of this behavior is shown in Figure 7.4, in which the Right_MdLF-SPL is overestimated by RecoBundles and it is missing most of the terminations in the latero-temporal ROI by TractSeg-retrained.

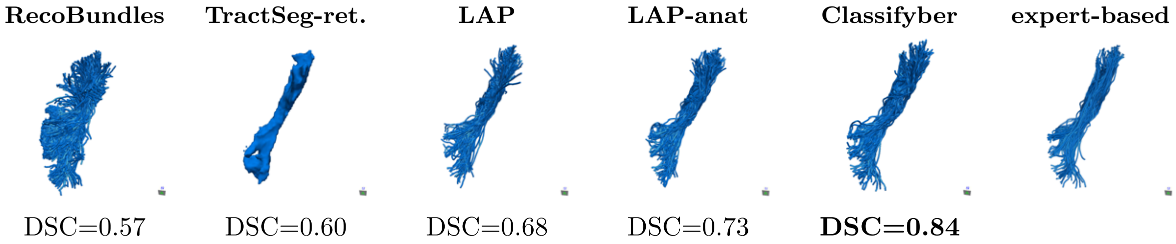


Figure 7.4: Qualitative comparison of segmented bundles in one target subject of the HCP-minor dataset. One instance of Right_MdLF-SPL predicted by RecoBundles, TractSeg-retrained, LAP, LAP-anat, and Classifyber with the expert-based segmentation. Highest quality of segmentation in bold face.

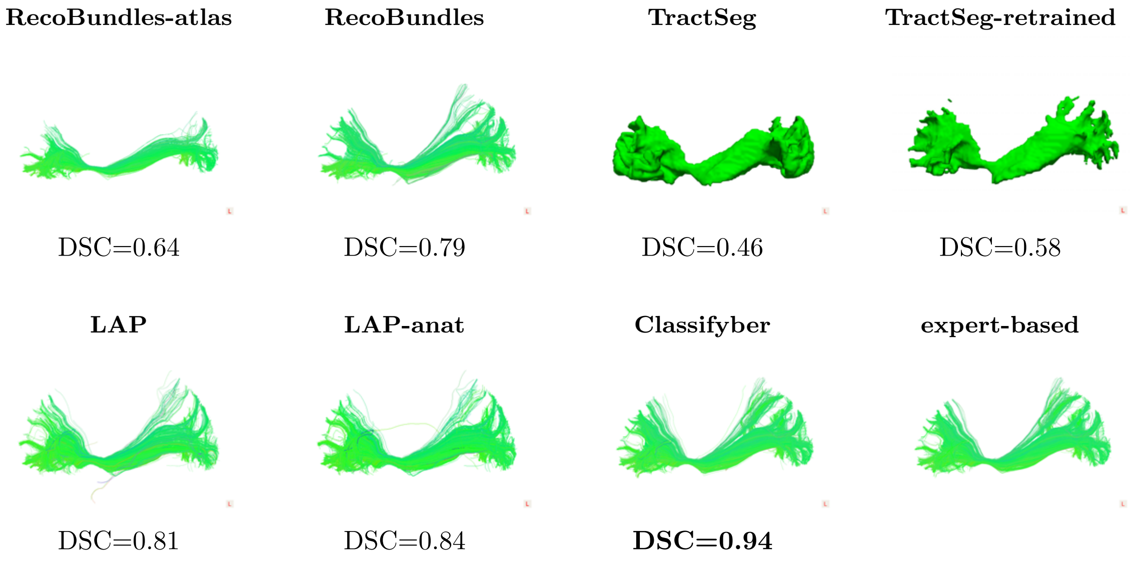
7.3.3 Experiments and results on HCP-IFOF dataset

For the HCP-IFOF dataset, we randomly split the pool of subjects in 15 for training and 15 for test. We conducted multiple experiments as described in Section 7.3.1, and then evaluated the segmentation performances through the DSC score, see Section 5.6.

In Table 7.4 we report the result of comparing Classifyber with all other methods and variants: RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, and LAP-anat. The average DSC across 15 subjects shows the superiority of Classifyber. Additionally, a qualitative visual comparison is reported in Figure 7.5, which illustrates that the Left_IFOF estimated with RecoBunldes-atlas (first panel), is clearly missing the middle and posterior subcomponents with respect to the expert-based segmented bundle (last panel). A very similar behavior is observed in the bundle predicted by TractSeg (third panel).

	RecoB.-atlas	RecoBundles	TractSeg	TractSeg-ret.	LAP	LAP-anat	Classifyber
Left_IFOF	0.45 ± 0.14	0.80 ± 0.04	0.48 ± 0.04	0.61 ± 0.03	0.81 ± 0.04	0.82 ± 0.07	**0.91** ± 0.03
Right_IFOF	0.62 ± 0.18	0.72 ± 0.06	0.41 ± 0.06	0.57 ± 0.04	0.73 ± 0.05	0.77 ± 0.06	**0.89** ± 0.03

Table 7.4: Quantitative comparison over the HCP-IFOF dataset: DSC (mean ± sd) across 15 target subjects for RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber. Highest quality of segmentation in bold face.



RecoBundles-atlas **RecoBundles** **TractSeg** **TractSeg-retrained**

DSC=0.64 DSC=0.79 DSC=0.46 DSC=0.58

LAP **LAP-anat** **Classifyber** **expert-based**

DSC=0.81 DSC=0.84 **DSC=0.94**

Figure 7.5: Qualitative comparison of segmented bundles in one target subject of the HCP-IFOF dataset. One instance of Left_IFOF predicted by RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber with the expert-based segmented bundle. Highest quality of segmentation in bold face.

7.3.4 Experiments and results on HCP-major dataset

From the HCP-major dataset, we selected the same 21 test subjects used in the experiments presented in [Wasserthal et al., 2018a]. In this way, we could directly compare our new results on the major bundles with theirs and, at the same time, we could test the reproducibility of their results. Of the 84 remaining subjects, 15 were randomly selected and used as training set for Classifyber. The kinds of bundles considered are those for which the two waypoint ROIs are available in the MNI_JHU_tracts_ROIs_atlas. We could not execute LAP and LAP-anat on this dataset because such methods are unfeasible on tractograms of 10 million streamlines. We conducted multiple experiments as described in Section 7.3.1, and then evaluated the segmentation performances through the DSC score, see Section 5.6.

In Table 7.5 and Figure 7.6 we report the mean quality of segmentation as DSC for RecoBundles-atlas, TractSeg and Classifyber over the major bundles considered, across 21 subjects[10]. In this case, the segmentation quality obtained by Classifyber outperforms the one obtained by RecoBundles-atlas, and it is comparable to the one obtained by TractSeg, for all the bundles considered. Moreover, Figure 7.7 shows a qualitative comparison of one instance of Right_CST segmented with the three different methods. It is visible that Classifyber reaches a comparable quality of segmentation to TractSeg, even though it uses only 15 subjects as examples.

	RecoB.-atlas	TractSeg	Classifyber
Right_CST	0.62 ± 0.07	**0.85** ± 0.02	**0.87** ± 0.02
Left_CST	0.62 ± 0.11	**0.85** ± 0.03	**0.86** ± 0.10
Right_UF	0.57 ± 0.24	0.79 ± 0.03	**0.86** ± 0.03
Right_AF	0.53 ± 0.11	**0.83** ± 0.02	**0.86** ± 0.03
Left_UF	0.55 ± 0.27	0.77 ± 0.03	**0.84** ± 0.04
Left_IFOF	0.67 ± 0.06	0.80 ± 0.02	**0.84** ± 0.03
Left_ILF	0.57 ± 0.07	0.77 ± 0.02	**0.84** ± 0.04
Right_IFOF	0.76 ± 0.04	0.80 ± 0.02	**0.84** ± 0.03
Left_AF	0.71 ± 0.05	**0.84** ± 0.03	**0.83** ± 0.04
Right_ILF	0.42 ± 0.13	0.75 ± 0.03	**0.82** ± 0.04

Table 7.5: Quantitative comparison over HCP-major dataset: DSC (mean ± sd) across 21 target subjects for RecoBundles-atlas, TractSeg and Classifyber. Highest quality of segmentation in bold face.

---

[10] For what concerns TractSeg, these results accurately reproduce those in [Wasserthal et al., 2018a].

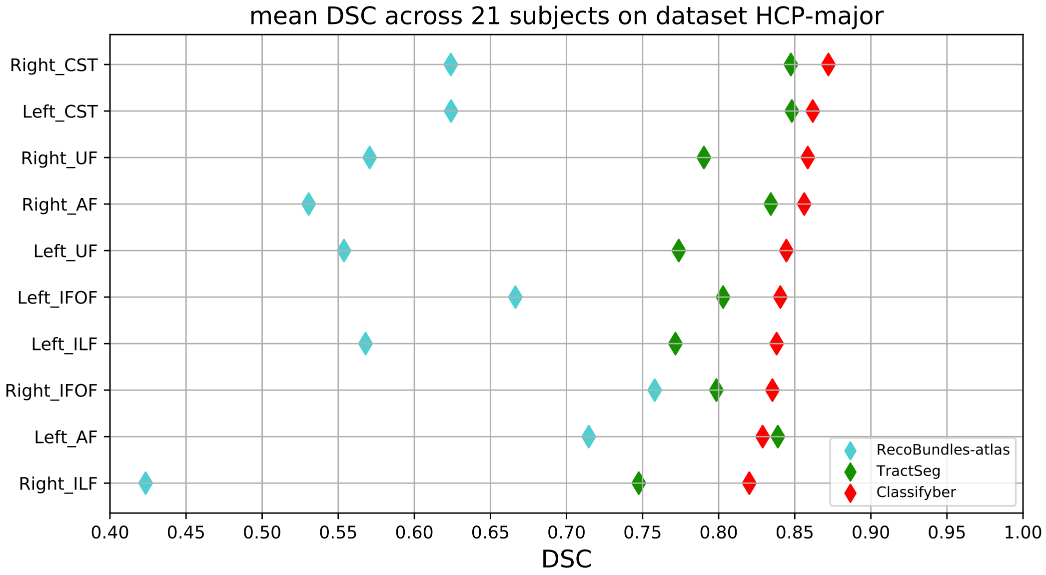


Figure 7.6: Quantitative comparison over HCP-major dataset: mean DSC across 21 target subjects for RecoBundles-atlas (light blue), TractSeg (green) and Classifyber (red).

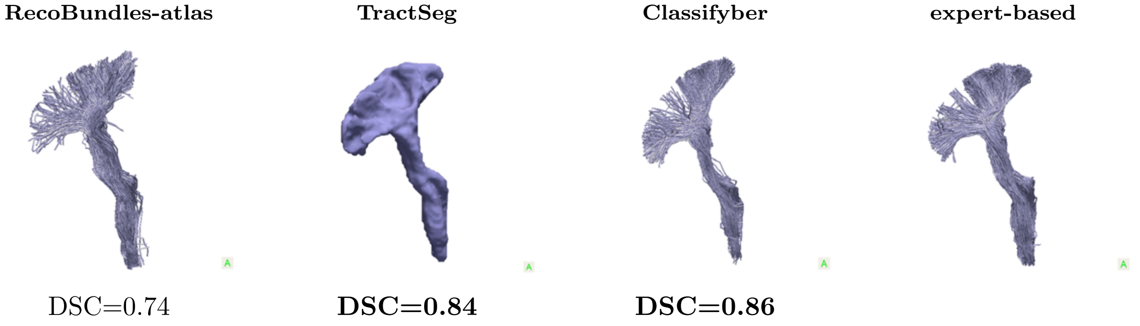


Figure 7.7: Qualitative comparison of segmented bundles in one target subject of the HCP-major dataset. One instance of Right_CST predicted by RecoBundles-atlas, TractSeg, and Classifyber with the expert-based segmented bundle. Highest quality of segmentation in bold face.

7.3.5 Experiments and results on Clinical dataset

Due to the small number of subjects of the Clinical dataset, instead of splitting the pool of 7 subjects into training and test sets, we ran Classifyber in two different ways: (i) we trained Classifyber on the IFOFs and AFs of the HCP-major dataset[11] and then segmented the 7 patients in the Clinical dataset. We refer to this case as *Classifyber-major*. (ii) We performed a cross-validation study with the leave-one-subject-out (LOSO)

---

[11] We chose this dataset because it is part of the exact same dataset used for training TractSeg, to have fair comparison between the two methods.

strategy, using *only* 6 subjects from the Clinical dataset as training set and the remaining subject as test set, repeatedly. We refer to this case as *Classifyber-LOSO*. In this latter case we also aimed to show the ability of Classifyber to accurately segment bundles even when trained on a very small number of segmentations, in this case only 6. To conclude, for the IFOF, we ran one additional experiment where Classifyber was trained on the IFOFs of the HCP-IFOF dataset. We refer to this case as *Classifyber-IFOF*. We conducted multiple experiments as described in Section 7.3.1, and then evaluated the segmentation performances through the DSC score, see Section 5.6.

In Table 7.6 we report the quantitative comparison in terms of mean DSC for Classifyber and TractSeg. The comparison is focused on TractSeg because in [Wasserthal et al., 2018a] it is stated that the method works well also on clinical quality data, without the need for retraining the network. Individually, over the 14 segmented bundles, Classifyber always obtained a higher DSC than TractSeg, for all the different training sets, i.e. for all three different variants: Classifyber-major, Classifyber-IFOF and Classifyber-LOSO.

	TractSeg	Classifyber-major	Classifyber-IFOF	Classifyber-LOSO
Left_IFOF	0.42 ± 0.05	0.71 ± 0.09	0.82 ± 0.07	$\mathbf{0.89} \pm 0.03$
Left_AF	0.23 ± 0.02	0.74 ± 0.13	-	$\mathbf{0.92} \pm 0.03$

Table 7.6: Quantitative comparison over the Clinical dataset: DSC (mean ± sd) across 7 target subjects for TractSeg, Classifyber-major, Classifyber-IFOF and Classifyber-LOSO. Highest quality of segmentation in bold face.

7.3.6 Experiments and results on fractal dimension

With this experiments, we wanted to investigate the reason of consistent success or consistent failure of some of the segmentation methods compared on specific datasets. To this end, we studied how the performance of the different segmentation methods is affected by the fractal dimension (FD) of the target bundles. We computed the FD of the voxel mask of each target bundle as segmented by experts and compared it with the quality of segmentation (DSC) obtained for that bundle by each automatic segmentation method, across all experiments (approximately 500 bundles). For TractSeg and RecoBundles, that number was larger because we investigated also the variants TractSeg-retrained and RecoBundles-atlas, while for LAP and LAP-anat it was smaller because it was not possible to execute the methods on the HCP-major dataset, where supersets substantially exceeded 100 thousands streamlines.

In Figure 7.8, we present the relationship between the FDs and the DSC scores of each method when segmenting all bundles in the experiments over the four datasets described above, i.e. on approximately 500 bundles. In the same figure, we also show the linear interpolation of such values as a summary of all experiments presented in this work, reporting the Pearson correlation coefficient (R) between FD and DSC.

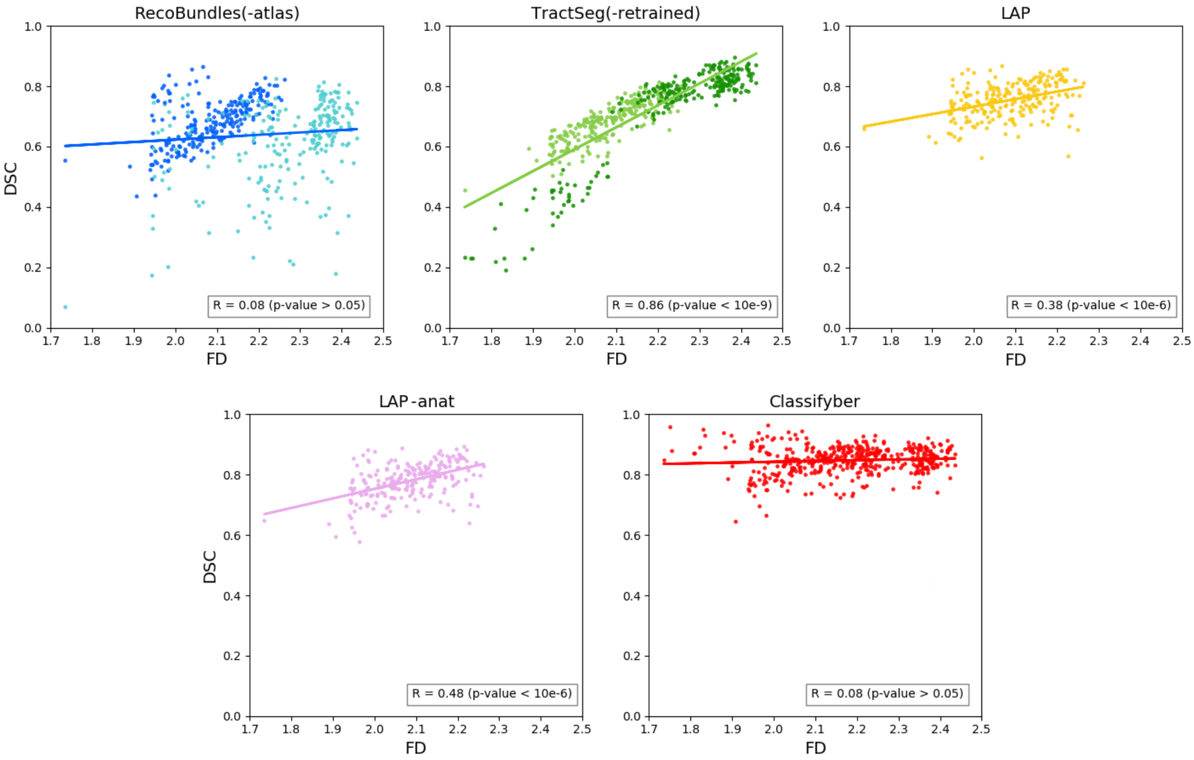


Figure 7.8: DSC vs FD across all methods for all the predicted bundles of the experiments in the work of Classifyber. Top: RecoBundles and RecoBundles-atlas (blue and light blue), TractSeg and TractSeg-retrained (green and light green), and LAP (yellow). Bottom: LAP-anat (pink) and Classifyber (red). R is the Pearson correlation coefficient and related p-value between FD and DSC over all the predicted bundles, i.e. approximately 500 segmentations.

The results show that the quality of segmentation of TractSeg, the only voxel-based segmentation method, is strongly dependent on the FD of the bundle to be segmented. LAP and LAP-anat also show some degree of dependency, while RecoBundles and Classifyber are not affected by the FD of bundles.

7.3.7 Size of the training set

Within the HCP-minor dataset, we also studied how much the quality of segmentation of Classifyber is affected when changing the number of subjects of the training set from 1 to 60. In this case, we considered also subjects for which all bundles received an expert-made score at least of 2.

In Figure 7.9 we show how the mean DSC of Classifyber over multiple bundles changes with the number of training subjects. We observe that the quality of segmentation has no substantial increase beyond approximately 15 subjects and plateaus at 30 subjects.

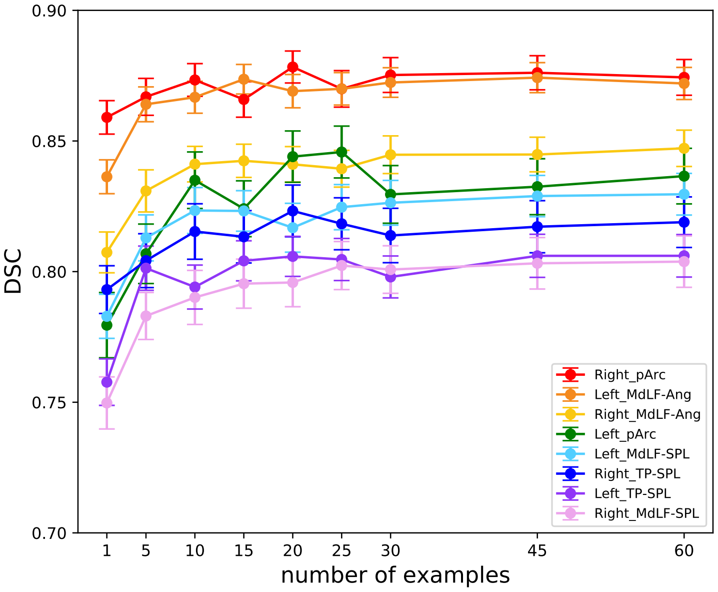


Figure 7.9: DSC (mean ± sd of the mean) across 25 test subjects of the HCP-minor dataset when varying the number of examples, from 1 to 60, for Classifyber. Each of the bundle is depicted with a different color.

7.3.8 Analysis of the computing time

In Table 7.7 we report the time (in minutes) required by each segmentation method for the training phase and for segmenting one IFOF of the HCP-IFOF dataset. We chose this dataset because it is the only dataset on which we compared all segmentation methods and variants.

We observed that for Classifyber the training time was linearly correlated with the number of training streamlines. The main cost of the computation in both the training

	training phase	segmentation	total
RecoBundles-atlas	0	0.5	0.5
RecoBundles	0	3	3
Classifyber	34	3	37
LAP	0	130	130
LAP-anat	0	142	142
TractSeg-retrained(*)	175(**)	5	180
TractSeg(*)	720(***)	5	725

Table 7.7: Time in minutes required to train each method and to segment one IFOF for: RecoBundles-atlas, RecoBundles, TractSeg, TractSeg-retrained, LAP, LAP-anat, and Classifyber, when having 15 training examples. (*) GPU accelerated. (**) segmenting 2 kinds of bundles at the same time. (***) Training on 84 subjects to segment 72 kinds of bundles at the same time.

and test phases is the preparation of the feature space. For the test phase, almost all of the 3 minutes were spent preparing the target tractogram for the linear classifier (steps (b1) and (b2)), while the actual prediction (step (b3)) only required less than 1 second.

In contrast to Classifyber, RecoBundles, LAP and LAP-anat do not require training time, because their underlying learning algorithms, i.e. nearest neighbor and linear assignment respectively, are *lazy learning algorithms* that postpone the computation to when the testing/segmentation step is required. In the case of RecoBundles, the segmentation step required between 0.5 and 3 minutes, on the example discussed above. LAP and LAP-anat required approximately 140 minutes and are thus the slowest of the methods compared. TractSeg adopts a different approach because it segments 72 bundles in parallel. The training time of TractSeg is vastly larger than all other methods, requiring 7 hours on a GPU. In the case of TractSeg-retrained, on the examples of HCP-IFOF dataset, the training phase required approximately 3 hours on GPU, see Table 8. Both TractSeg and TractSeg-retrained required approximately 5 minutes to segment a new bundle.

All computations of all experiments described in this work were executed on the high-performance computing (HPC) cluster provided by Indiana University, allocating 16 cores of Intel Xeon CPU E5-2680 2.50GHz and 32Gb of RAM, a setup equivalent to a powerful personal workstation typically available in research labs and clinics. For TractSeg, we also allocated one NVIDIA GPU RTX 2080Ti.

Chapter 8

Towards Reproducibility in Neuroscience

This Chapter is devoted to reproducibility for Neuroscience. Section 8.1 briefly illustrates the general reproducibility challenge within the scientific community. In Section 8.2, we describe our contributions to promote openness and facilitate reproducibility in neuroscience. Lastly, in Section 8.3, we report a reproducibility study conducted by us about comparison of lazy learning methods for white matter bundle segmentation.

8.1 Reproducibility challenge

One of the greatest challenges of data-intensive science is to enhance knowledge discovery by assisting humans and machines to access to scientific data and their associated algorithms and workflows [Wilkinson et al., 2016]. Since current scientific research is based on previous published findings, being able to reproduce an experimental workflow and replicate its results is essential to advance discovery [Pernet and Poline, 2015]. The agreement of measurements and observations performed by different researchers is on the basis of a rigorous scientific method. In the last few years, several authors have questioned the reliability of scientific studies [Ioannidis, 2005, Button et al., 2013], even claiming that most of them contain false findings [Ioannidis, 2005]. The probability that a scientific finding is indeed true is related to the statistical power and significance of the study, and on its prior probability of being true [Ioannidis, 2005]. Reproduction can be considered as a possible means to contrast false findings, since the probability of a scientific finding being true increases with the number of its reproduction [Ioannidis, 2005, Pernet and Poline, 2015].

Replicability of results requires however a considerable effort towards code and data sharing initiatives through specific websites and repositories. This practice was not so common until a few years ago, but to-date is becoming standard within the neuroscientific community [Tournier et al., 2012, Garyfallidis et al., 2014, Avesani et al., 2019]. Free and open source software like *DIPY* (`dipy.org`) [Garyfallidis et al., 2014], *MRtrix* (`mrtrix.org`) [Tournier et al., 2012] and *scikit-learn* (`scikit-learn.org`) [Pedregosa et al., 2011] were essential for all the work presented in this thesis, facilitating our scientific findings. In general, to let data Findable, Accessible, Interoperable and Reusable (FAIR), a set of guiding principles have been recently designed, which are known as FAIR Data Principles [Wilkinson et al., 2016]. The primary intent was to enhance knowledge discovery by assisting humans and machines in the detection of, access to, and integration and analysis of scientific data[1]. Being concise and domain-independent, these principles can benefit a wide range of diverse stakeholders in the current digital ecosystem.

8.2 Promoting open sharing of scientific data and code

Following the FAIR Data Principles, a significant effort has been made by us to carry out the research activities of this thesis according to the guidelines of openness and reproducibility.

Specifically, we provide the source code we developed to address the research questions **Q1**, **Q2**, and **Q3** (presented in 4.2.4) as open source software, see Table 8.1.

research question - contribution	web link
Q1 - Comparison of distances	`https://github.com/emanuele/prni2017_comparison_of_distances`
Q2 - LAP-anat	`https://github.com/giulia-berto/app-lap-anat`
Q3 - Classifyber	`https://github.com/FBK-NILab/app-classifyber`

Table 8.1: Web links to the software repositories with the code developed to address the three research questions of the present thesis.

Additionally, we have integrated our pipelines of analysis, composed of: (i) code, (ii) data, and (iii) environment of execution, into the online open platform *BrainLife* (`brainlife.io`).

BrainLife is a cloud computing platform which tackles the reproducibility challenges in

---

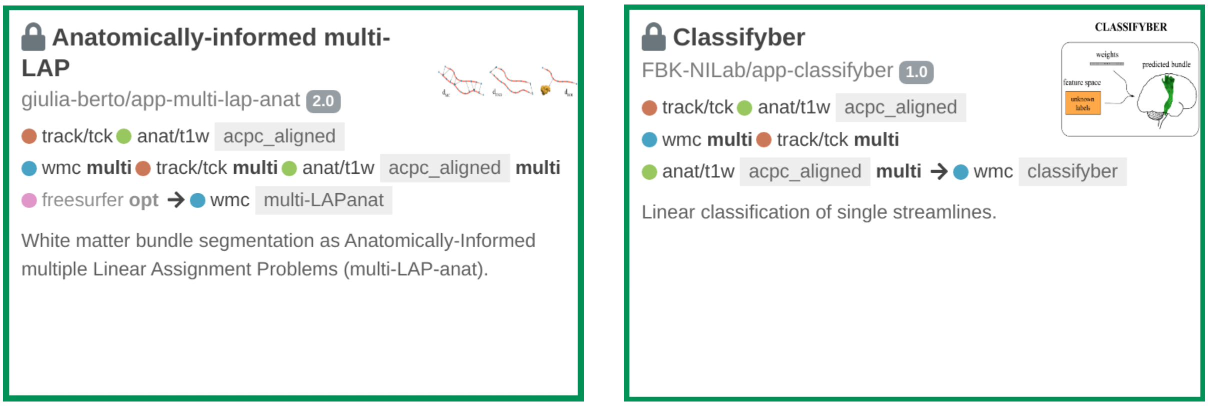
[1] The term *data* does not only refer to data in its conventional sense, but also to algorithms, tools, and workflows that led to that data [Wilkinson et al., 2016].

neuroscience by jointly providing software for data and analyses management, sharing, and publication. Pipelines and workflows integrated in the platform allow researchers efficient execution of consecutive steps of analysis on multiple datasets, thus representing a unique mechanism for tracking the reproducibility of scientific experiments. To analyze data, BrainLife makes use of web applications, which are open services with self-contained pieces of software that can be reused in a cloud computing environment [Avesani et al., 2019].

Table 8.2 reports the main BrainLife web applications we developed. These applications implement the two main methodological contributions of the present thesis, i.e. (i) the LAP-anat bundle segmentation method, and (ii) Classifyber, the linear classifier for bundle segmentation. Classifyber is available both as the full algorithm that implements the training and test phases (called *Classifyber* in Table 8.2), and as a pre-trained method ready to segment bundles (called *Classifyber - segmentation* in Table 8.2). In Figure 8.1 there is a graphical example of the appearance of web applications on the `brainlife.io` website.

contribution	BrainLife application name	BrainLife web link
LAP-anat	Anatomically-informed multi-LAP	`https://doi.org/10.25663/brainlife.app.227`
Classifyber	Classifyber	`https://doi.org/10.25663/brainlife.app.228`
	Classifyber - segmentation	`https://doi.org/10.25663/brainlife.app.265`

Table 8.2: Link to the main BrainLife web applications we developed, implementing the two main methodological contributions of the present thesis.



Anatomically-informed multi-LAP
giulia-berto/app-multi-lap-anat 2.0
● track/tck ● anat/t1w acpc_aligned
● wmc **multi** ● track/tck **multi** ● anat/t1w acpc_aligned **multi**
● freesurfer **opt** → ● wmc multi-LAPanat
White matter bundle segmentation as Anatomically-Informed multiple Linear Assignment Problems (multi-LAP-anat).

Classifyber
FBK-NILab/app-classifyber 1.0
● track/tck ● anat/t1w acpc_aligned
● wmc **multi** ● track/tck **multi**
● anat/t1w acpc_aligned **multi** → ● wmc classifyber
Linear classification of single streamlines.

Figure 8.1: Graphical example of how the web applications Anatomically-informed multi-LAP (left) and Classifyber (right) appear on the BrainLife website.

To conduct the experiments described in Chapter 7 and in Section 8.3, we also developed other minor auxiliary BrainLife web applications, see Table 8.3.

task accomplished	BrainLife application name	BrainLife web link
LAP and NN	multi-LAP and multi-NN	`https://doi.org/10.25663/brainlife.app.174`
ANTs registration	ANTs transformation and registration	`https://doi.org/10.25663/bl.app.26`
DSC evaluation	DSC evaluation	`https://doi.org/10.25663/bl.app.57`
ROC curve	ROC curve	`https://doi.org/10.25663/brainlife.app.180`

Table 8.3: Link to minor auxiliary BrainLife web applications we developed.

Moreover, we share as well few of the tractography datasets that we generated and presented in Section 6.2 through the online platform `brainlife.io`. In particular, we share tractograms and bundles of the HCP-canonical dataset and of the HCP-minor dataset, see Table 8.4. We firmly believe that data sharing is an essential step towards scientific reproducibility and data upcycling within multiple scientific communities.

dataset	BrainLife publication link
HCP-canonical	`https://doi.org/10.25663/brainlife.pub.3`
HCP-minor	`https://doi.org/10.25663/brainlife.pub.11`

Table 8.4: Link to the tractography datasets we share through the online platform BrainLife.

8.3 A reproducibility study: comparison of lazy learning methods for white matter bundle segmentation

Aim

The BrainLife data processing mechanism facilitated our attempt to replicate some of the findings contained in [Sharmin et al., 2018], with the aim of verifying the claim that “LAP [Linear Assignment Problem]-based segmentation [...] is substantially more accurate than the NN [Nearest Neighbor] strategy”. The primary objective of this reproducibility study was indeed to independently assess, by replicating the same empirical study, whether this claim held even in the case of diverse experimental conditions, such as for example a different preprocessing pipeline and tracking algorithm.

Experimental design

Original and replication study adopted the same experimental design. The original study used tractograms and bundles of 30 subjects of the HCP-EuDX dataset described in Section 6.2.1, while the replication study used tractograms and bundles of 30 subjects of the HCP-canonical dataset described in Section 6.2.3. In both cases, the set of 30

HCP subjects was equally split into a set of examples and a set of target subjects. The segmented bundles of the examples subjects were used to train the automatic segmentation methods to extract the bundles in the tractograms of the target subjects. The segmented bundles of the target subjects were instead used as ground truth for evaluation of the performances of the automatic segmentation methods. Automatic segmentation of 10 different bundles was performed using two automatic lazy learning methods, the LAP-based method [Sharmin et al., 2018] and the NN-based method [Yoo et al., 2015], and their performances were compared through the Receiver Operating Characteristic and Area Under the Curve (ROC-AUC) analysis, as explained in Section 5.6.2. This was done for each bundle and for each of the 15 target subjects independently. The ROC curves were then macro-averaged over the 15 target subjects to obtain a summary of the performances for each bundle and each method.

Results

Figure 8.2 shows an example of the ROC curve of the two automatic segmentation methods under comparison when segmenting the right Cingulum bundle (original plot on the left panel and replicated plot on the right panel). In each plot, the yellow curve corresponds to the LAP-based method and the green curve corresponds to the NN-based method. In the legend, the averaged value of the AUC score over the 15 target subjects for the two segmentation methods is reported (auc_LAP for LAP and auc_NN_DR_MAM for NN).

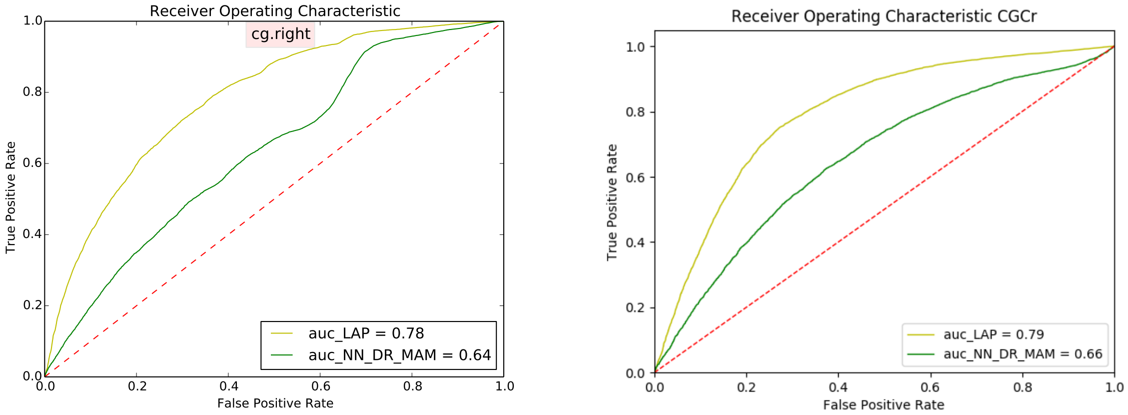


Figure 8.2: Examples of averaged Receiver Operating Characteristic (ROC) curve related to the LAP-based segmentation method (yellow) and NN-based segmentation method (green) when segmenting the right Cingulum bundle. Original plot on the left panel and replicated plot on the right panel.

Overall, in the original study, the averaged values of the AUC score over the 15 target subjects for the LAP-based method span from 0.69 to 0.83 while those for the NN-based method span from 0.54 to 0.66. In the replication study, they span from 0.70 to 0.85 and from 0.62 to 0.70 respectively. These results show that, for each of the bundles considered, the AUC score related to the LAP-based method is consistently higher than the AUC score related to the NN-based methods, confirming that the segmentation based on LAP outperforms the segmentation based on NN.

Chapter 9

Discussion and Conclusion

This Chapter concludes Part I of the present thesis. In Section 9.1, we discuss the results of the research activities that have been presented in the previous Chapters. Section 9.2 provides the conclusion of this thesis, and finally, some future activities that can arise from the work presented in this thesis are collected in Section 9.3.

9.1 Discussion

In this Section we discuss the results of the experiments presented in Chapter 7 conducted to address the three main research questions of the thesis, and the results of the reproducibility study presented in Section 8.3.

9.1.1 Discussion about the comparison of streamline distance functions for supervised bundle segmentation

We compared several streamline distance functions available in the literature in order to determine which one is the most effective for the task of supervised bundle segmentation.

The results reported in Table 7.1 clearly show that there are no major differences in the segmentation accuracy, measured as DSC, when using different streamline distance functions. The highest mean DSC value, corresponding to 0.55 for $d_{varifolds}$, is not significantly higher than the values obtained by the other distances. At the *local* level, different distances have a geometrically different concept of proximity possibly leading to different nearest neighbors, which however we observed do not lie far apart from each other. As a consequence, at a *higher/aggregated* level of bundle, it should not be a surprise that they lead to a comparable quality of segmentation, measured as DSC. In this circumstance, a voxel-based measure of bundle overlap like the DSC revealed indeed to be insensitive to the choice of the streamline distance function.

From a computational point of view, Table 7.2 shows that the computational times of the streamline distance functions can be very different. For example, there are more than two orders of magnitude between the computational time of d_{MDF} and the one of $d_{varifolds}$. To conclude, for the supervised bundle segmentation task based on a voxel-based measure like the DSC, we suggest that practitioners select fast distance functions, such as d_{MDF}, d_{MC}, d_{SC} or d_{LC}, over slower ones, such as d_{PMD} and $d_{varifolds}$.

9.1.2 Discussion about the Anatomically-Informed Multiple Linear Assignment Problems (LAP-anat) for bundle segmentation

We aimed to improve the results of bundle segmentation by considering information about both the shape of the streamlines and about the relative anatomical position of the bundle of interest. To achieve this goal, we proposed to extend the correspondence-based method proposed in [Sharmin et al., 2018] by including additional anatomical information within the optimization process of the Linear Assignment Problem (LAP). We referred to this extension of the LAP algorithm as LAP-anat.

Figure 7.1 illustrates that, on average, the proposed LAP-anat method outperforms the original LAP method of [Sharmin et al., 2018], for all the bundles considered. Streamlines composing the same anatomical bundles not only have a similar shape, but have also a propensity to interconnect or pass through predefined ROIs of the brain. Including this information into the optimization process is helpful in particular to identify those streamlines that may have a less similar shape from the example, but that are near to ROIs that are known from the literature defining the bundle of interest. Additionally, taking into account the endpoint-based distance is useful to select all the streamlines that terminate in the same cortical region.

Moreover, Figure 7.1 provides also evidence that the proposed LAP-anat method is more effective than the original LAP method especially in segmenting small bundles, which are usually more sensitive to registration errors and are generally harder to segment than large bundles. Overall, these results confirm the assumption that, for all the bundles considered, including additional information about the relative anatomical position of bundles helps to improve streamline-based bundle segmentation, in particular for small bundles.

9.1.3 Discussion about Classifyber, a robust streamline-based linear classifier for bundle segmentation

To overcome the limitations of state-of-the-art automatic method for bundle segmentation, we proposed Classifyber, a new streamline-based method for white matter bundle segmentation. We conducted multiple experiments using different automatic segmentation methods and demonstrated that the proposed one, Classifyber, substantially increases the quality of segmentation with respect to other methods, and that is robust to different data settings.

Here we discuss the results obtained both by Classifyber and by all the other segmentation methods compared in Section 7.3.

General comments At the global level, all the results on the comparison among automatic segmentation methods presented in Section 7.3 indicate one main message: Classifyber clearly outperforms other methods in all cases, by a substantial margin, and segments bundles very accurately. This is observed to occur across different kinds of bundles, tractography techniques, expert-made segmentations, and quality of dMRI data. The summary results in Figure 7.8, which report on the y-axes the DSC score for each of the hundreds of individual bundles segmented across all the experiments of Section 7.3, show that Classifyber obtained scores ranging from 0.65 to 0.96, with a mean and standard deviation of 0.85 ± 0.05. This is the highest quality of segmentation among the different automatic segmentation methods by a large or substantial margin, in almost all cases, see the results at the level of individual target bundles reported in Sections 7.3.2, 7.3.3, 7.3.4 and 7.3.5. On the datasets HCP-minor and HCP-IFOF, we believe that the gap in terms of DSC score between Classifyber and the second best method, which is LAP-anat, the other method we proposed, can be explained by the lack of the one-to-one constraint and the more advanced optimization models used by Classifyber.

Figure 7.8 also reports that the results obtained by LAP and LAP-anat are consistently superior to those obtained by RecoBundles and TractSeg, at least on the datasets HCP-minor and HCP-IFOF. For both these datasets, LAP-anat outperforms LAP with an average improvement in the DSC score of $+0.04$[1]. Moreover, the figure shows that RecoBundles and TractSeg have a large amount of variability in the quality of segmentation across the different experiments: their DSC scores range from 0.07 to 0.90, with means of 0.64 ± 0.14 and 0.71 ± 0.13 respectively. Surprisingly, TractSeg reaches a low (or very low)

---

[1] These results accurately reproduce those of the comparison between LAP and LAP-anat on the HCP-canonical dataset reported in Section 7.2.2.

quality of segmentation on small and wrinkled bundles. We discuss this point in detail below, where we discuss the FD of bundles.

Dataset shift It is well known that training classification algorithms on examples that systematically differ from the examples in the test set substantially reduces the quality of classification [Quionero-Candela et al., 2009, Stacke et al., 2019]. This issue, called *dataset shift*, was previously mentioned for bundle segmentation in [Wasserthal et al., 2018a] and has no simple solution. In our case, this happened in some cases when we used the methods RecoBundles-atlas, TractSeg, Classifyber-major and Classifyber-IFOF.

As an example, consider the results of RecoBundles-atlas and TractSeg on the HCP-IFOF dataset reported in Table 7.4. A possible explanation of the poor performances of these methods is that the anatomical shape of the bundles used as examples differs from the shape of the manually segmented bundles of the test dataset, i.e. the HCP-IFOF dataset. Specifically, the example used by RecoBundles-atlas, i.e. the IFOF of the atlas of [Yeh et al., 2018], comes from clustering followed by expert labeling. The examples used by TractSeg come from a semi-automatic refinement of the segmentation provided by WMQL [Wassermann et al., 2016], while the examples in HCP-IFOF are manually segmented by an expert neurosurgeon and follow the definition in [Sarubbo et al., 2013] and [Hau et al., 2016][2].

A similar behavior was observed for the variants Classifyber-major and Classifyber-IFOF when segmenting bundles of the Clinical dataset. When examples are partly different from the ones in the Clinical dataset, the DSC (which for Classifyber was around 0.9) dropped accordingly to around 0.8 for Classifyber-IFOF and to 0.7 for Classifyber-major, see Table 7.6. In this case, the three different datasets, i.e. HCP-IFOF, HCP-major and Clinical, do not only differ with respect to the expert-based segmentation technique, but also with respect to the tractography algorithm and to the quality of the dMRI data (refer to Sections 6.2 and 6.3 for all the details).

The use of RecoBundles-atlas and TractSeg is justified by the fact that (i) we received the indication from the authors of RecoBundles to use the models provided by the atlas of [Yeh et al., 2018], and that (ii) in [Wasserthal et al., 2018a] they claim that the pre-trained network of TractSeg works properly on multiple datasets, including clinical ones. However, due to their poor performances obtained across multiple datasets discussed above, we believe that bundle atlases and pre-trained networks have to be adopted with caution for supervised bundle segmentation.

---

[2] These anatomical differences are justified by the fact that the anatomical definition of some white matter bundles, among which the IFOF, is in evolution [Sarubbo et al., 2013, Forkel et al., 2014, Wu et al., 2016a].

Beyond dataset shift The methods RecoBundles and TractSeg-retrained where adopted mainly to conduct a fair comparison with Classifyber, which only uses example subjects belonging to the same dataset of the test subjects (properly split into training and test sets). As a proof of the fact that the dataset shift issue substantially reduces the quality of segmentation, consider the results reported in Table 7.4 on the HCP-IFOF dataset. The segmentation quality reached by RecoBundles and TractSeg-retrained is remarkably higher than that obtained by RecoBundles-atlas and TractSeg, with a mean improvement in the DSC score of +0.15/0.20. However, these results are still considerably inferior to those obtained by Classifyber, which has a mean DSC margin of at least +0.11 over these segmentation methods, see Table 7.4.

Fractal dimension of bundles While conducting hundreds of automatic segmentations with different methods, we noticed that TractSeg had consistent success or consistent failure on specific datasets. This method, which segments bundles by levaraging their volumetric representation as voxel masks using convolutional neural networks (CNNs), very accurately segmented the bundles in HCP-major, but obtained only medium or poor results in other datasets, see Tables 7.3, 7.4 and 7.6. Figure 7.8 shows that the segmentation quality reached by this method is deeply affected by a specific geometric property of the voxel mask of the target bundle: its fractal dimension (FD, see Section 5.7). TractSeg accurately segmented bundles which are smooth and rounded, i.e., with high FD, while it produced poor segmentations when they are wrinkled and irregular, i.e., with low FD, see Figure 5.5. As a matter of fact, in the qualitative comparison in Figure 7.5, we observe that TractSeg is not very precise in segmenting fine-grained structures of the bundles, in particular their terminal portions.

We believe that this tendency is related to the operations of convolution and max-pooling of the fully convolutional neural networks (FCNNs) applied to voxels within TractSeg. In the domain of computer vision, it has been observed multiple times that FCNNs are biased towards rounded segmentations of objects, which can loose details and fine-grained structure, in particular because of the max-pooling operation, see for example [Sabour et al., 2017, Kim et al., 2018, Wei et al., 2019]. This problem is inherent in U-net [Ronneberger et al., 2015], which is at the core of TractSeg.

Figure 7.8 shows that both LAP and LAP-anat are also slightly affected by the FD of bundles, though much less than TractSeg. However, such result might not be entirely reliable because a large portion of the segmentations are missing due to the limitation of LAP and LAP-anat to address large tractograms.

Voxel-based vs streamline-based bundle segmentation On the other hand, as clearly shown in Figure 7.8, both RecoBundles and Classifyber are insensitive to the FD of the voxel masks of bundles. We speculate that the reason for this is related to the streamline-based nature of such methods and, more specifically, to the fact that they operate via *single* streamline classification. By predicting whether or not each streamline of the tractogram belongs to the target bundle, there is not a specific constraint to produce round/smooth voxel-structures as observed with TractSeg or to jointly consider all target streamlines during the prediction as in LAP and LAP-anat. Streamlines represent a spatial statistic of the dMRI signal that approximates the underlying anatomical connectivity patterns, thus enriching the mere volumetric information provided by voxel masks.

Towards a better brain connectivity characterization Our results provided evidence that Classifyber outperforms other automatic bundle segmentation methods, especially on small bundles of the HCP-minor dataset. Due to their relatively smaller size and shorter length with respect to long associative bundles, these bundles have been less studied in the past [Bullock et al., 2019]. However, they have recently proven to be of great help in understanding how information flows in the human brain [Wu et al., 2016b, Bullock et al., 2019, Sani et al., 2019]. For example, over the past few years, the functions attributed to them has expanded from pure language perception to more complex processes [Wu et al., 2016b], such as attention [Sani et al., 2019], and transferring information from superior to inferior horizontal bundles [Bullock et al., 2019]. As a consequence, accurate segmentation of small bundles is nowadays of paramount importance to better characterize the human brain connectivity.

Size of the training set As an additional result of this work, we observed that Classifyber requires only a small number of example bundles to obtain high quality of segmentation. In fact, Figure 7.9 shows that, on the HCP-minor dataset, there is no substantial gain in the quality of segmentation beyond 15 training examples. In the experiments on the Clinical dataset, Classifyber reached an extremely high segmentation quality using only 6 example subjects, with a mean DSC around 0.9, see Table 7.6.

RecoBundles, LAP and LAP-anat require a very small number of training subjects: 1 bundle/model for RecoBundles and around 5-10 for LAP and LAP-anat, according to [Sharmin et al., 2018]. On the contrary, TractSeg was trained on 84 subjects. Although in [Wasserthal et al., 2018a] there are no clear guidelines on the number of subjects to be used for training, it is well known that deep learning models need a very large training set, which is often not available in clinical settings and whose expert-labeling is usually very time consuming.

Analysis of the computing time Among the methods compared in this work, deciding which one is faster is not straightforward: on the one hand, streamline-based methods like Classifyber, RecoBundles, LAP and LAP-anat require the tractogram as input. If only raw dMRI data is provided, the time to build the tractogram should be accounted for the total time of the computation. On the other hand, TractSeg uses the GPU and requires a specific pre-processing of dMRI data as input, which needs approximately 30 minutes of computation per subject. Moreover, to obtain the predicted bundle as streamlines, bundle-specific tracking must be computed afterwards [Wasserthal et al., 2018b].

Overall, if the target tractogram is available, RecoBundles is the fastest segmentation method in our comparison, see Table 7.7. Alternatively, if pre-trained methods are available, TractSeg and Classifyber are also similarly as fast as RecoBundles. LAP-anat is the slowest segmentation method but, if training has to be done, TractSeg ranks last.

9.1.4 Discussion about the comparison of lazy learning methods for bundle segmentation

The aim of the reproducibility study presented in Section 8.3 was to replicate some of the findings of [Sharmin et al., 2018]. In particular, the goal was to independently assess, by replicating the same empirical study, whether the claim "LAP [Linear Assignment Problem]-based segmentation [...] is substantially more accurate than the NN [Nearest Neighbor] strategy" held even in the case of diverse experimental conditions.

The results reported in Section 8.3 show that, for each of the bundles considered, the AUC score related to the LAP-based method is consistently higher than the AUC score related to the NN-based methods, confirming that the segmentation based on LAP outperforms the segmentation based on NN. If a local systematic displacement between homologous bundles still persists after tractogram alignment, the NN strategy may underestimate the size of the bundle of interest. On the other hand, the LAP method is able to correctly follow such displacement thanks to the introduction of the one-to-one constraint in the optimization process [Sharmin et al., 2018].

The small differences of the averaged values of the AUC score over the 15 target subjects observed between original and replication study can be explained by the different characteristics and variability of the dataset used in the two studies. Nevertheless, our results demonstrate that the original claim can be generalized, bringing thus added value to the original scientific findings.

9.2 Conclusion

The main goal of this thesis was to develop a robust automatic method for accurate white matter bundle segmentation, by exploiting, combining and extending multiple up-to-date supervised learning techniques.

We compared several streamline distance functions available in the literature, and we showed that there are no major differences in the segmentation accuracy, measured as DSC at the voxel level, when using different streamline distance functions for the task of supervised white matter bundle segmentation. However, since their computational time can be very different, we suggested that practitioners select fast distance functions, such as d_{MDF}, d_{MC}, d_{SC} or d_{LC}, over slower ones, such as d_{PMD} and $d_{varifolds}$.

We combined, in a supervised streamline-based bundle segmentation method, the similarity between streamlines with the anatomical information from ROIs. Two additional distance functions were defined: the *endpoint-based* distance, d_{END}, which measures the average distance between the endpoints of two streamlines, and the *ROI-based* distance, d_{ROI}, which measures the distance between a streamline and an ROI. We demonstrated that including additional information about the relative anatomical position of bundles helps to improve streamline-based bundle segmentation, in particular for small bundles, which are notoriously harder to segment. Due to their location, number and functions, small and short bundles are not well characterized in the literature [Oishi et al., 2008, Guevara et al., 2011, Wu et al., 2016b]. However, small bundles have recently received increased attention and have proven to be of great help to better characterize the human brain connectivity [Guevara et al., 2017, Bullock et al., 2019].

We proposed Classifyber, a novel streamline-based linear classifier that segments white matter bundles from dMRI data and expert-made examples. Classifyber is the first automatic classification segmentation method that exploits both the shape of streamlines, obtained with tractography techniques from dMRI data, and the anatomical information of the bundles, in the form of connectivity patterns and specific ROIs. Classifyber substantially raises the quality of segmentation as compared to the state-of-the-art bundle segmentation methods available in the literature, by a large margin, and more importantly, across very diverse data settings, such as when there are differences in bundle size, tracking algorithm, and/or quality of dMRI data.

We provided evidence that some voxel-based bundle segmentation methods are deeply affected by a geometrical property of the shape of the bundles, i.e. by their fractal dimension (FD). For those methods, smooth and rounded bundles, i.e., with high FD, are easier to segment, while wrinkled and irregular bundles, i.e., with low FD, are harder to segment. We believe that this is due to an inherent bias of fully convolutional neural networks (FCNNs) applied on voxels. In contrast with voxel-based methods such as that presented in [Wasserthal et al., 2018a], we believe that accurate segmentation of bundles from dMRI data must leverage tractography techniques and also include information about streamlines. Streamlines represent a spatial statistic of the dMRI signal that approximates the underlying anatomical connectivity patterns, thus enriching the mere volumetric information provided by voxel masks.

Particular attention was also paid to curating the bundle datasets used to conduct the experiments. Although it has been shown that the lack of reliable anatomical ground truth information is a major issue for tractography applications [Maier-Hein et al., 2017], having bundle datasets refined manually or semi-automatically by expert neuroanatomists can partially tackle this problem and lead to more accurate and reproducible results.

Moreover, we performed a reproducibility study with the aim of reproducing some of the findings contained in the recent work of [Sharmin et al., 2018]. Our results demonstrated that the original claim can be generalized, thus bringing added value to the original scientific findings. Open sharing of scientific data and code are of acknowledged importance to facilitate reproducibility in neuroscience and to increase the quality of scientific research beyond the standard expectations. Our entire work, including the extensive comparison with other state-of-the-art bundle segmentation methods, would have not been possible without the availability of data and code shared by the community. Free and open source software like DIPY (`dipy.org`) [Garyfallidis et al., 2014], MRtrix (`mrtrix.org`) [Tournier et al., 2012] and scikit-learn (`scikit-learn.org`) [Pedregosa et al., 2011] were essential for all the research activities presented in this thesis project, facilitating our scientific findings. In addition, the use of the platform BrainLife (`brainlife.io`) enabled us to easily build consistent pipelines of analysis and to take advantage in the best possible way of the high-performance computing clusters provided [Avesani et al., 2019].

9.3 Future work

In this last Section, we present potential extensions and future activities that can arise from the work presented in this thesis.

A common limitation of the proposed bundle segmentation methods LAP-anat and Classifyber is that they can address only those kinds of bundles whose pertaining ROIs are defined from the literature. A possible way to overcome this issue could be to consider, for each streamline, its distance from a set of ROIs that are *independent* from the bundle of interest. In other words, similarly to the approach used to compute the first two sets of values composing the vectorial representation for Classifyber, we may account for a set of *prototypical* ROIs and compute their distance from each of the streamlines we want to classify. These ROIs should ideally form a parcellation of the whole brain.

In future, we plan to test nonlinear classification algorithms in order to investigate potential improvements in the segmentation quality of Classifyber. The current linear model used within Classifyber is indeed a limitation of the proposed method. Nevertheless, linear models are fast and light and, according to the results presented here, sufficient to substantially advance the state-of-the-art in automatic white matter bundle segmentation.

We considered a Clinical dataset composed of a limited number of bundles in a few patients affected by brain tumors. The lesions however did not affect the shape of the bundles consistently. We plan to extend our work on a larger pool of clinical patients, even by considering the cases of severely injured brains. In addition, an in depth investigation about the issue of *dataset shift*, i.e. whether we can use examples from healthy subjects to segment bundles in pathological situations, could open new interesting perspectives in clinical settings.

Moreover, we plan to implement our automatic methods into web applications specifically targeted for clinical applications. Having accurate automatically segmented bundles could be of invaluable help for clinicians, who, most of the time, still only rely on manual bundle segmentation.

Part II

Papers

Chapter 10

Comparison of Distances for Supervised Segmentation of White Matter Tractography

Olivetti, E., Bertò, G., Gori, P., Sharmin, N., Avesani, P.

Published in: 2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI). IEEE, pp. 1–4.

Chapter 11

Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation

Bertò, G., Avesani, P., Pestilli, F., Bullock, D., Caron, B., Olivetti, E.

Published in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 135–138.

Chapter 12

Classifyber, a robust streamline-based linear classifier for white matter bundle segmentation

Bertò, G., Bullock, D., Astolfi, P., Hayashi, S., Zigiotto, L., Annicchiarico, L., Corsini, F., De Benedictis, A., Sarubbo, S., Pestilli, F., Avesani, P., Olivetti, E.

Under review in NeuroImage.

Chapter 13

A reproducibility study: Comparison of lazy learning methods for white matter bundle segmentation

Bertò, G., Avesani, P.

In preparation, part of a larger contribution titled "*BrainLife.io, a free online platform for reproducible neuroscience*" (Pestilli F. et al.), to be submitted to a scientific journal.

Appendix A

Supplementary materials

This Chapter concludes this thesis and comprises some supplementary materials. Specifically, Section A.1 describes some additional results related to the work presented in Chapter 11, while Section A.2 describes some preliminary results related to the contribution presented in Chapter 12.

A.1 Additional results related to LAP-anat

Here we present some additional (unpublished) results related to the work presented in Chapter 11, in which we proposed the bundle segmentation method LAP-anat. The experimental design is the same of that described in Section 7.2.1.

A.1.1 Contribution of the individual cost matrices

With this experiment we wanted to quantify the contribution of the individual cost matrices involved in the optimization problem for bundle segmentation based on the anatomically-informed Linear Assignment Problem (LAP-anat), i.e. the distance matrix D, the endpoint-based distance matrix E, and the ROI-based distance matrix R. To this end, we performed automatic bundle segmentation with the proposed LAP-anat method, as was done in 7.2.1, and with its two other variants that we called LAP-end and LAP-roi. The difference between LAP-anat and its two variants lies only in the formulation of the total cost matrix:

- the total cost matrix of LAP-anat is $C = \lambda_D D + \lambda_E E + \lambda_R R$
- the total cost matrix of LAP-end is $C = \lambda_D D + \lambda_E E$
- the total cost matrix of LAP-roi is $C = \lambda_D D + \lambda_R R$

As was done in 7.2.1, the parameters λ_D, λ_E and λ_R were set in order to let all the values of the matrices span in the same range.

In Tables A.1 and A.2 we report the results of the comparison in terms of DSC over the HCP-canonical-large dataset and the HCP-canonical-small dataset, respectively[1].

	Left_TR	Right_TR	Left_CST	Right_CST	Left_AF	Right_AF
LAP-anat	0.87 ± 0.03	0.87 ± 0.02	0.86 ± 0.02	0.87 ± 0.02	0.86 ± 0.03	0.84 ± 0.03
LAP-end	0.87 ± 0.02	0.86 ± 0.02	0.85 ± 0.02	0.86 ± 0.02	0.84 ± 0.04	0.81 ± 0.03
LAP-roi	0.85 ± 0.02	0.85 ± 0.02	0.83 ± 0.02	0.84 ± 0.02	0.85 ± 0.04	0.83 ± 0.03

Table A.1: Quantitative comparison over the HCP-canonical-large dataset: DSC(mean ± sd) across 30 subjects with the proposed method LAP-anat and the two variants LAP-end and LAP-roi.

	Left_CGC	Right_CGC	Left_CGH	Right_CGH	Left_UF	Right_UF
LAP-anat	0.83 ± 0.04	0.86 ± 0.06	0.83 ± 0.05	0.80 ± 0.06	0.80 ± 0.04	0.81 ± 0.05
LAP-end	0.82 ± 0.06	0.85 ± 0.06	0.78 ± 0.07	0.78 ± 0.07	0.78 ± 0.05	0.78 ± 0.06
LAP-roi	0.83 ± 0.05	0.86 ± 0.05	0.81 ± 0.05	0.79 ± 0.07	0.79 ± 0.05	0.81 ± 0.05

Table A.2: Quantitative comparison over the HCP-canonical-small dataset: DSC(mean ± sd) across 30 subjects with the proposed method LAP-anat and the two variants LAP-end and LAP-roi.

These results show that the segmentation quality of the two variants LAP-end and LAP-roi is comparable of that of LAP-anat, for all the bundles considered. However, we observe consistently slightly better results with LAP-anat, meaning that both the additional anatomical distance matrices, E and R, help to improve the results of bundle segmentation.

A.1.2 Comparing tractogram registration with SLR vs ANTs

With this experiment we wanted to quantify the differences in the quality of bundle segmentation when using a different method for tractogram registration. To this end, we consider two bundles segmentation methods, LAP and LAP-anat, and two initial tractogram registration algorithms, the Streamline Linear Registration (SLR) [Garyfallidis et al., 2015] and the non-linear Advanced Normalization Tool (ANTs) [Avants et al., 2008]. We performed automatic bundle segmentation both with LAP and LAP-anat, in combination with either (i) an initial SLR registration, as was done in 7.2.1, or (ii) an initial non-linear registration with ANTs (based on T1 images). We refer to the two cases of (ii) with ANTs-LAP and ANTs-LAP-anat.

---

[1] In both tables, the results of LAP-anat are the same as those reported in Figure 7.1.

In Tables A.3 and A.4 we report the results of the comparison in terms of DSC over the HCP-canonical-large dataset and the HCP-canonical-small dataset, respectively[2].

	Left_TR	Right_TR	Left_CST	Right_CST	Left_AF	Right_AF
LAP	0.85 ± 0.02	0.85 ± 0.02	0.84 ± 0.03	0.85 ± 0.02	0.83 ± 0.03	0.80 ± 0.03
ANTs-LAP	0.86 ± 0.02	0.86 ± 0.02	0.85 ± 0.02	0.86 ± 0.02	0.84 ± 0.03	0.82 ± 0.03
LAP-anat	0.87 ± 0.03	0.87 ± 0.02	0.86 ± 0.02	0.87 ± 0.02	0.86 ± 0.03	0.84 ± 0.03
ANTs-LAP-anat	0.88 ± 0.02	0.88 ± 0.02	0.87 ± 0.02	0.87 ± 0.02	0.86 ± 0.03	0.84 ± 0.03

Table A.3: Quantitative comparison over the HCP-canonical-large dataset: DSC(mean ± sd) across 30 subjects with LAP, ANTs-LAP, LAP-anat, and ANTs-LAP-anat.

	Left_CGC	Right_CGC	Left_CGH	Right_CGH	Left_UF	Right_UF
LAP	0.81 ± 0.05	0.82 ± 0.06	0.77 ± 0.07	0.76 ± 0.07	0.74 ± 0.07	0.76 ± 0.08
ANTs-LAP	0.82 ± 0.06	0.82 ± 0.07	0.80 ± 0.04	0.79 ± 0.05	0.78 ± 0.06	0.79 ± 0.06
LAP-anat	0.83 ± 0.04	0.86 ± 0.06	0.83 ± 0.05	0.80 ± 0.06	0.80 ± 0.04	0.81 ± 0.05
ANTs-LAP-anat	0.84 ± 0.05	0.86 ± 0.06	0.84 ± 0.04	0.82 ± 0.05	0.81 ± 0.05	0.82 ± 0.05

Table A.4: Quantitative comparison over the HCP-canonical-small dataset: DSC(mean ± sd) across 30 subjects with LAP, ANTs-LAP, LAP-anat, and ANTs-LAP-anat.

Over the datasets considered, when using the same bundle segmentation method, we observe that performing a non-linear tractogram registration with ANTs rather than using SLR leads to comparable results in terms of DSC, for all the bundles taken into account. The numerical gap between the first and the last row of both the Tables A.3 and A.4 is mainly due to the use of a different bundle segmentation method.

A.2 Preliminary results related to Classifyber

Here we present some preliminary (unpublished) results related to the contribution presented in Chapter 12, in which we proposed the bundle segmentation method Classifyber. As in Chapter 12, the bundle segmentation framework adopted is the linear classification of single streamlines as performed by Classifyber, see Section 5.5, but with some modifications. As opposed to Chapter 12, the following results were obtained only on a limited number of subjects of the HCP-minor dataset: 5 subjects were used for training and 4 subjects were used for test[3].

---

[2] In both tables, the results of LAP and LAP-anat are the same as those reported in Figure 7.1.

[3] The 4 test subjects selected here are not included in the pool of the 25 test subjects considered in the experiments reported in Section 7.3

A.2.1 Comparing tractogram registration with ACPC vs SLR vs ANTs

Similarly to the experiment presented in Section A.1.2, we wanted to quantify the differences in the quality of bundle segmentation when using a different method for tractogram registration. The three tractogram registration methods compared are (i) the Anterior Commissure - Posterior Commissure (ACPC) alignment, (ii) SLR, and (iii) ANTs (based on T1). In this preliminary assessment, the streamlines were transformed into vectors only considering their distances from 100 global landmarks (GL).

In Figure A.1, we report the results of the comparison in terms of mean DSC across 4 subjects of the HCP-minor dataset.

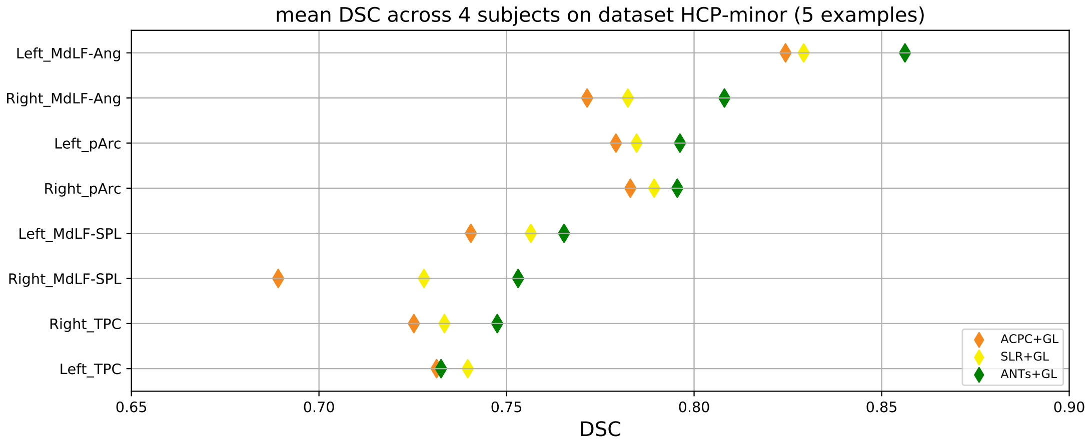


Figure A.1: Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification on subjects registered with ACPC alignment, SLR registration, and ANTs registration.

Figure A.1 shows that, for almost all the bundles of the HCP-minor dataset, the best bundle segmentation performances, in terms of DSC, are obtained when registering the tractograms with ANTs. In this case, the results are consistently higher than performing the ACPC alignment, and generally higher than performing the SLR registration. For this reason, we adopted ANTs as the main tractogram registration method for the work[4] of Classifyber described in Chapter 12.

---

[4] In that work, only for the Clinical dataset we adopted the SLR registration, because non-linear registration of clinical data is debated [Garyfallidis et al., 2015].

A.2.2 Effectiveness of different combinations of features

As a preliminary experiment, we investigated several combinations of features among those described in Section 5.5.1, in order to determine which of them are the most effective for the classification task. In all the cases, due to the results shown in Figure A.1, the registration method used was ANTs. We performed bundle segmentation adopting the linear classification framework of Classifyber to compare the 7 following combinations of features (for the description of the sets of features refer to Section 5.5.1):

- ANTs+GL uses the set of features (i)
- ANTs+GL+ROI+FS combines the sets of features (i)+(iv)+(vi)
- ANTs+GL+END+ROI+FA combines the sets of features (i)+(iii)+(iv)+(v)
- ANTs+GL+ROI combines the sets of features (i)+(iv)
- ANTs+GL+ROI+END combines the sets of features (i)+(iii)+(iv)
- ANTs+GL+LP+ROI combines the sets of features (i)+(ii)+(iv)
- ANTs+GL+LP+END+ROI combines the sets of features (i)+(ii)+(iii)+(iv)

The results of bundle segmentation in terms of DSC for the 7 different combinations of features are reported in Figure A.2.

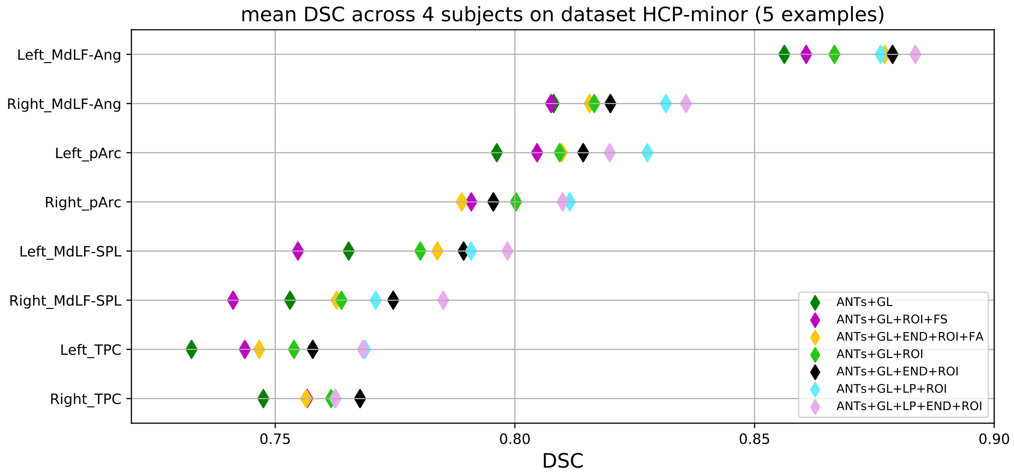


Figure A.2: Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification considering 7 different combinations of features. Notice that the results of the segmentation method depicted in green (ANTs+GL) are the same as those reported in Figure A.1, depicted in the same color.

From Figure A.2, for the task of linear streamline classification for bundle segmentation, we can infer the following facts:

- including the sets of features related to the Fractional Anisotropy (FA), and to the curvature and torsion derived from the Frenet-Serret framework (FS) does not help to improve the performances
- as shown in Section A.1, including the anatomical information in the form of both endpoint-based distances (END) and ROI-based distances (ROI) helps to improve the performances
- including also the distances from Local Landmarks (LL) is useful to improve the performances
- the combination of features ANTs+GL+LP+END+ROI is on average the most effective strategy for the specified task

Given all these facts, the combination of features ANTs+GL+LP+END+ROI was selected as the default feature space for all the experiments of the work of Classifyber described in Chapter 12.

A.2.3 Comparison between classification algorithms and solvers

As a further preliminary assessment, we compared two different classification algorithms together with two different solvers to address the bundle segmentation task. The algorithms we considered are *Logistic Regression* (LR) and *Stochastic Gradient Descent* (SGD). The solvers we considered with LR are 'liblinear', which uses a coordinate descent (CD) algorithm, and 'sag', which uses Stochastic Average Gradient (SAG) descent. When using the SGD algorithm, the 'early-stopping' option, which terminates training when validation score is not improving, was also tested. To summarize, we performed bundle segmentation adopting the linear classification framework of Classifyber to compare the 4 following combinations of classification algorithms and solvers:

- LR liblinear, which uses the LR algorithm with 'liblinear' solver
- LR SAG, which uses the LR algorithm with 'sag' solver
- SGD, which uses the SGD algorithm
- SGD early-stopping, which uses the SGD algorithm with the 'early-stopping' option

In all cases, we adopted the most effective combinations of features obtained from the comparison in Section A.2.2, i.e. ANTs+GL+LL+END+ROI. When using the SGD algorithm, we also performed a cross-validated grid search to optimize the *class weight*[5], i.e. the weights associated with the two classes, and *alpha*[6], i.e. the constant that multiplies the regularization term.

The results of bundle segmentation in terms of DSC for the 4 different combinations of algorithms and solvers are reported in Figure A.3.

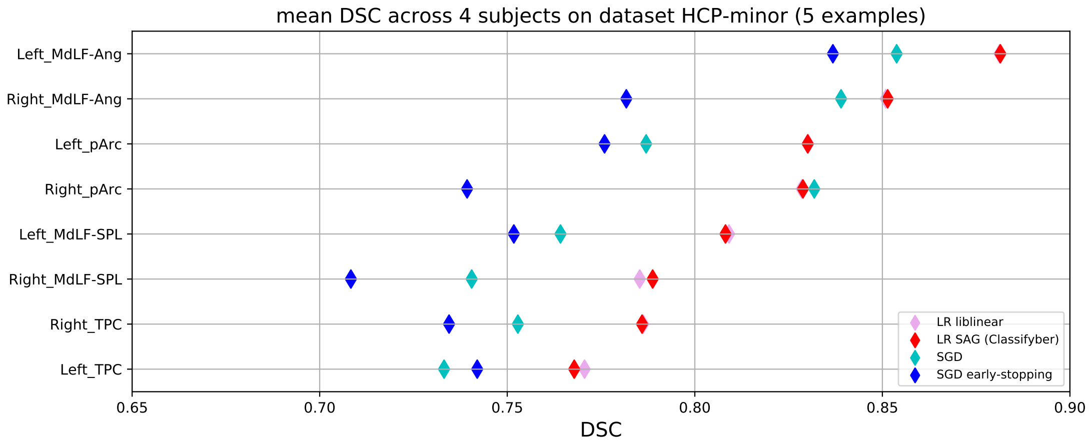


Figure A.3: Quantitative comparison over the HCP-minor dataset: mean DSC across 4 subjects obtained for bundles predicted with linear classification considering 4 different combinations of classification algorithms and solvers. Notice that the results of the segmentation method depicted in pink (LR liblinear) are the same as those reported in Figure A.2, depicted in the same color but on a different pool of subjects. The strategy depicted in red (LR SAG) is the one adopted for Classifyber.

Figure A.3 shows that, when using the SGD classification algorithm, a lower quality of segmentation is reached, in terms of DSC, with respect to the use of the LR algorithm, for all the bundles considered. Moreover, the 'early-stopping' option did not help to improve the performances. On the other hand, with the LR algorithm we obtained high DSC scores, for all the bundles considered. Although for the two solvers 'liblinear' and 'sag' we observe comparable segmentation performances, the latter is more memory-efficient and thus more suitable when dealing with large datasets, such as those employed for the work of Classifyber described in Chapter 12.

---

[5] Class_weight_grid = [balanced, 0.5, 1, 3, 5, 10, 50, 500].

[6] Alpha_grid = [1e-6, 1e-7, 1e-8, 1e-9].

To conclude, the preliminary experiments described in Sections A.2.1, A.2.2 and A.2.3 were essential to build the core algorithm of Classifyber, which uses (i) data previously registered with ANTs, (ii) the combinations of features ANTs+GL+LL+END+ROI, and (iii) the LR classification algorithm with 'sag' solver.

Bibliography

[Aha, 2013] Aha, D. W. (2013). *Lazy Learning.* Springer Science & Business Media. Google-Books-ID: b1CqCAAAQBAJ.

[Alexander et al., 2001] Alexander, D. C., Pierpaoli, C., Basser, P. J., and Gee, J. C. (2001). Spatial transformations of diffusion tensor magnetic resonance images. *IEEE transactions on medical imaging*, 20(11):1131–1139.

[Alexandroni et al., 2016] Alexandroni, G., Zimmerman Moreno, G., Sochen, N., and Greenspan, H. (2016). Coresets vs clustering: comparison of methods for redundancy reduction in very large white matter fiber sets. volume 9784, pages 97840A–97840A–9.

[Alt and Godau, 1995] Alt, H. and Godau, M. (1995). Computing the fréchet distance between two polygonal curves. *International Journal of Computational Geometry & Applications*, 05(01n02):75–91.

[Andersson et al., 2007] Andersson, J. L. R., Jenkinson, M., and Smith, S. (2007). Non-linear registration, aka Spatial normalisation. Technical Report TR07JA2, University of Oxford.

[Assemlal et al., 2011] Assemlal, H.-E., Tschumperlé, D., Brun, L., and Siddiqi, K. (2011). Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *Medical Image Analysis*, 15:369–396.

[Avants et al., 2008] Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

[Avesani et al., 2019] Avesani, P., McPherson, B., Hayashi, S., Caiafa, C. F., Henschel, R., Garyfallidis, E., Kitchell, L., Bullock, D., Patterson, A., Olivetti, E., Sporns, O., Saykin, A. J., Wang, L., Dinov, I., Hancock, D., Caron, B., Qian, Y., and Pestilli,

F. (2019). The open diffusion data derivatives, brain data upcycling via integrated publishing of derivatives and reproducible open cloud services. *Scientific Data*, 6(1):1–13.

[Bach et al., 2014] Bach, M., Fritzsche, K. H., Stieltjes, B., and Laun, F. B. (2014). Investigation of resolution effects using a specialized diffusion tensor phantom. *Magnetic Resonance in Medicine*, 71(3):1108–1116.

[Basser, 1995] Basser, P. J. (1995). Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. *NMR in biomedicine*, 8(7-8):333–344.

[Basser et al., 1994] Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267.

[Basser et al., 2000] Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic resonance in medicine*, 44(4):625–632.

[Batchelor et al., 2006] Batchelor, P. G., Calamante, F., Tournier, J. D., Atkinson, D., Hill, D. L. G., and Connelly, A. (2006). Quantification of the shape of fiber tracts. *Magnetic Resonance in Medicine*, 55(4):894–903.

[Bazin et al., 2011] Bazin, P.-L., Ye, C., Bogovic, J. A., Shiee, N., Reich, D. S., Prince, J. L., and Pham, D. L. (2011). Direct segmentation of the major white matter tracts in diffusion tensor images. *NeuroImage*, 58(2):458–468.

[Behrens et al., 2007] Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage*, 34(1):144–155.

[Behrens et al., 2003] Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50(5):1077–1088.

[Bells et al., 2011] Bells, S., Cercignani, Deoni, Assaf, Pasternak, Evans, Leemans, and And (2011). Tractometry – Comprehensive Multi-modal Quantitative Assessment of White Matter Along Specific Tracts.

[Berman et al., 2008] Berman, J. I., Chung, S., Mukherjee, P., Hess, C. P., Han, E. T., and Henry, R. G. (2008). Probabilistic streamline q-ball tractography using the residual bootstrap. *NeuroImage*, 39(1):215–222.

[Bertò et al., 2019] Bertò, G., Avesani, P., Pestilli, F., Bullock, D., Caron, B., and Olivetti, E. (2019). Anatomically-Informed Multiple Linear Assignment Problems for White Matter Bundle Segmentation. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 135–138.

[Bijsterbosch and Volgenant, 2010] Bijsterbosch, J. and Volgenant, A. (2010). Solving the Rectangular assignment problem and applications. *Annals of Operations Research*, 181:443–462.

[Bishop, 2013] Bishop, C. M. (2013). *Pattern recognition and machine learning.* Springer, 1st ed. 2006. corr. 2nd printing 2011 edition. Published: Hardcover.

[Brown, 1828] Brown, R. (1828). A brief account of microscopical observations made in the months of June, july, and August, 1827, on the particles contained in the pollen of plants ; and on the general existence of active moleculues in organic and inorganic bodies, Philosophical Magazine. *Ann. Philos. New Series*, 4:161–178.

[Brun et al., 2004] Brun, A., Knutsson, H., Park, H.-J., Shenton, M. E., and Westin, C.-F. (2004). Clustering fiber traces using normalized cuts. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004*, pages 368–375. Springer.

[Bullmore and Sporns, 2009] Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10(3):186–198.

[Bullock et al., 2019] Bullock, D., Takemura, H., Caiafa, C. F., Kitchell, L., McPherson, B., Caron, B., and Pestilli, F. (2019). Associative white matter connecting the dorsal and ventral posterior human cortex. *Brain Structure and Function.*

[Button et al., 2013] Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews. Neuroscience*, 14(5):365–376.

[Buxton et al., 1998] Buxton, R. B., Frank, L. R., Wong, E. C., Siewert, B., Warach, S., and Edelman, R. R. (1998). A general kinetic model for quantitative perfusion imaging with arterial spin labeling. *Magnetic Resonance in Medicine*, 40(3):383–396.

[Caiafa and Pestilli, 2017] Caiafa, C. F. and Pestilli, F. (2017). Multidimensional encoding of brain connectomes. *Scientific Reports*, 7(1):1–13.

[Calamante, 2019] Calamante, F. (2019). The Seven Deadly Sins of Measuring Brain Structural Connectivity Using Diffusion MRI Streamlines Fibre-Tracking. *Diagnostics*, 9(3):115.

[Caruyer, 2014] Caruyer, E. (2014). Phantomas: a flexible software library to simulate diffusion MR phantoms.

[Catani and De Schotten, 2008] Catani, M. and De Schotten, M. T. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex*, 44(8):1105–1132.

[Catani et al., 2002] Catani, M., Howard, R. J., Pajevic, S., and Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *NeuroImage*, 17(1):77–94.

[Charon and Trouvé, 2013] Charon, N. and Trouvé, A. (2013). The Varifold Representation of Nonoriented Shapes for Diffeomorphic Registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580.

[Christiaens et al., 2015] Christiaens, D., Reisert, M., Dhollander, T., Sunaert, S., Suetens, P., and Maes, F. (2015). Global tractography of multi-shell diffusion-weighted imaging data using a multi-tissue model. *NeuroImage*, 123:89–101.

[Clayden et al., 2007] Clayden, J. D., Storkey, A. J., and Bastin, M. E. (2007). A probabilistic model-based approach to consistent white matter tract segmentation. *Medical Imaging, IEEE Transactions on*, 26(11):1555–1561.

[Close et al., 2009] Close, T. G., Tournier, J.-D., Calamante, F., Johnston, L. A., Mareels, I., and Connelly, A. (2009). A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. *NeuroImage*, 47(4):1288–1300.

[Colby et al., 2012] Colby, J. B., Soderberg, L., Lebel, C., Dinov, I. D., Thompson, P. M., and Sowell, E. R. (2012). Along-tract statistics allow for enhanced tractography analysis. *NeuroImage*, 59(4):3227–3242.

[Corouge et al., 2006] Corouge, I., Fletcher, P., Joshi, S., Gouttard, S., and Gerig, G. (2006). Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. *Medical Image Analysis*, 10(5):786–798.

[Corouge et al., 2004] Corouge, I., Gouttard, S., and Gerig, G. (2004). Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pages 344–347 Vol. 1. IEEE.

[Côté et al., 2013] Côté, M.-A., Girard, G., Boré, A., Garyfallidis, E., Houde, J.-C., and Descoteaux, M. (2013). Tractometer: towards validation of tractography pipelines. *Medical Image Analysis*, 17(7):844–857.

[Daducci et al., 2015] Daducci, A., Dal Palù, A., Lemkaddem, A., and Thiran, J.-P. (2015). COMMIT: convex optimization modeling for microstructure informed tractography. *IEEE transactions on medical imaging*, 34(1):246–257.

[De Benedictis et al., 2018] De Benedictis, A., Nocerino, E., Menna, F., Remondino, F., Barbareschi, M., Rozzanigo, U., Corsini, F., Olivetti, E., Marras, C. E. E., Chioffi, F., Avesani, P., and Sarubbo, S. (2018). Photogrammetry of the Human Brain: A Novel Method for Three-Dimensional Quantitative Exploration of the Structural Connectivity in Neurosurgery and Neurosciences. *World neurosurgery*, 115.

[De Benedictis et al., 2016] De Benedictis, A., Petit, L., Descoteaux, M., Marras, C. E., Barbareschi, M., Corsini, F., Dallabona, M., Chioffi, F., and Sarubbo, S. (2016). New insights in the homotopic and heterotopic connectivity of the frontal portion of the human corpus callosum revealed by microdissection and diffusion tractography. *Human Brain Mapping*, 37(12):4718–4735.

[Dell'Acqua et al., 2007] Dell'Acqua, F., Rizzo, G., Scifo, P., Clarke, R. A., Scotti, G., and Fazio, F. (2007). A model-based deconvolution approach to solve fiber crossing in diffusion-weighted MR imaging. *IEEE transactions on bio-medical engineering*, 54(3):462–472.

[Descoteaux, 2015] Descoteaux, M. (2015). High Angular Resolution Diffusion Imaging (HARDI). In *Wiley Encyclopedia of Electrical and Electronics Engineering*, pages 1–25. American Cancer Society.

[Descoteaux and Deriche, 2009] Descoteaux, M. and Deriche, R. (2009). High Angular Resolution Diffusion MRI Segmentation Using Region-Based Statistical Surface Evolution. *Journal of Mathematical Imaging and Vision*, 33(2):239–252.

[Descoteaux et al., 2009] Descoteaux, M., Deriche, R., Knosche, T. R., and Anwander, A. (2009). Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *Medical Imaging, IEEE Transactions on*, 28(2):269–286.

[Dice, 1945] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

[Ding et al., 2003] Ding, Z., Gore, J. C., and Anderson, A. W. (2003). Classification and quantification of neuronal fiber pathways using diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(4):716–721.

[Dumoulin and Turski, 2007] Dumoulin, C. L. and Turski, P. A. (2007). Phase Contrast MRA. In *eMagRes*. American Cancer Society.

[Durrleman et al., 2011] Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., and Ayache, N. (2011). Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 55(3):1073–1090.

[Eckstein et al., 2009] Eckstein, I., Shattuck, D. W., Stein, J. L., McMahon, K. L., de Zubicaray, G., Wright, M. J., Thompson, P. M., and Toga, A. W. (2009). Active fibers: Matching deformable tract templates to diffusion tensor images. *NeuroImage*, 47:T82–T89.

[Edinger, 1885] Edinger, L. (1885). Über den Verlauf der centralen Hirnnervenbahnen mit Demonstrationen von Präparaten. *Arch Psychiat Nervenkr*, 16:858–859.

[Einstein, 1905] Einstein, A. B. (1905). On the Motion of Small Particles Suspended in Liquids at Rest Required by the Molecular-Kinetic Theory of Heat *.

[Esteban et al., 2007] Esteban, F. J., Sepulcre, J., de Mendizábal, N. V., Goñi, J., Navas, J., de Miras, J. R., Bejarano, B., Masdeu, J. C., and Villoslada, P. (2007). Fractal dimension and white matter changes in multiple sclerosis. *NeuroImage*, 36(3):543–549.

[Falconer, 2014] Falconer, K. J. (2014). *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons Inc, Hoboken, third edition edition.

[Farquharson et al., 2013] Farquharson, S., Tournier, J.-D., Calamante, F., Fabinyi, G., Schneider-Kolsky, M., Jackson, G. D., and Connelly, A. (2013). White matter fiber tractography: why we need to move beyond DTI: Clinical article. *Journal of Neurosurgery*, 118(6):1367–1377.

[Fonov et al., 2011] Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., and Brain Development Cooperative Group (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1):313–327.

[Forkel et al., 2014] Forkel, S. J., Thiebaut de Schotten, M., Kawadler, J. M., Dell'Acqua, F., Danek, A., and Catani, M. (2014). The anatomy of fronto-occipital connections from early blunt dissections to contemporary tractography. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 56:73–84.

[Fowlkes et al., 2004] Fowlkes, C., Belongie, S., Chung, F., and Malik, J. (2004). Spectral grouping using the Nystrom method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225.

[Garyfallidis, 2013] Garyfallidis, E. (2013). *Towards an accurate brain tractography.* PhD Thesis, University of Cambridge.

[Garyfallidis et al., 2014] Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., Nimmo-Smith, I., and Contributors, D. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers in Neuroinformatics*, 8(8):1+.

[Garyfallidis et al., 2012] Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). QuickBundles, a Method for Tractography Simplification. *Frontiers in neuroscience*, 6.

[Garyfallidis et al., 2018] Garyfallidis, E., Côté, M.-A. A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., and Descoteaux, M. (2018). Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*, 170:283–295.

[Garyfallidis et al., 2015] Garyfallidis, E., Ocegueda, O., Wassermann, D., and Descoteaux, M. (2015). Robust and efficient linear registration of white-matter fascicles in the space of streamlines. *NeuroImage*, 117:124–140.

[George and M Das, 2019] George, K. and M Das, J. (2019). Neuroanatomy, Thalamocortical Radiations. In *StatPearls*. StatPearls Publishing, Treasure Island (FL).

[Girard et al., 2014] Girard, G., Whittingstall, K., Deriche, R., and Descoteaux, M. (2014). Towards quantitative connectivity analysis: reducing tractography biases. *NeuroImage*, 98:266–278.

[Glasser et al., 2013] Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124.

[Gorgolewski, 2016] Gorgolewski, K. (2016). FreeSurfer reconstruction of the MNI152 ICBM2009c asymmetrical non-linear atlas. figshare. Dataset. https://doi.org/10.6084/m9.figshare.4223811.v1.

[Gori et al., 2016] Gori, P., Colliot, O., Marrakchi-Kacem, L., Worbe, Y., De Vico Fallani, F., Chavez, M., Poupon, C., Hartmann, A., Ayache, N., and Durrleman, S. (2016). Parsimonious Approximation of Streamline Trajectories in White Matter Fiber Bundles. *IEEE transactions on medical imaging.*

[Gray, 1918] Gray, H. (1918). *Anatomy of the Human Body. Philadelphia: Lea & Febiger; Bartleby. com, 2000.*

[Greene et al., 2017] Greene, C., Cieslak, M., and Grafton, S. T. (2017). Effect of different spatial normalization approaches on tractography and structural brain networks. *Network Neuroscience*, 2(3):362–380.

[Guevara et al., 2017] Guevara, M., Román, C., Houenou, J., Duclap, D., Poupon, C., Mangin, J. F., and Guevara, P. (2017). Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. *NeuroImage*, 147:703–725.

[Guevara et al., 2012] Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Leboyer, M., Houenou, J., and Mangin, J. F. (2012). Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *Neuroimage*, 61(4):1083–1099.

[Guevara et al., 2011] Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Houenou, J., Leboyer, M., and Mangin, J.-F. (2011). Segmentation of Short Association Bundles in Massive Tractography Datasets Using a Multi-subject Bundle Atlas. In San Martin, C. and Kim, S.-W., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Lecture Notes in Computer Science, pages 701–708. Springer Berlin Heidelberg.

[Guevara Alvez, 2011] Guevara Alvez, P. B. (2011). *Inference of a human brain fiber bundle atlas from high angular resolution diffusion imaging.* PhD Thesis, Paris 11.

[Guo et al., 2016] Guo, Z., Wang, Y., Lei, T., Fan, Y., and Zhang, X. (2016). DTI Image Registration under Probabilistic Fiber Bundles Tractography Learning. *BioMed Research International*, 2016.

[Hastie et al., 2017] Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2017). *The elements of statistical learning : data mining, inference, and prediction.* Springer, 2nd ed. 2009. corr. 7th printing 2013 edition. Published: Hardcover.

[Hau et al., 2017] Hau, J., Sarubbo, S., Houde, J., Corsini, F., Girard, G., Deledalle, C., Crivello, F., Zago, L., Mellet, E., Jobard, G., Joliot, M., Mazoyer, B., Tzourio-Mazoyer, N., Descoteaux, M., and Petit, L. (2017). Revisiting the human uncinate fasciculus, its subcomponents and asymmetries with stem-based tractography and microdissection validation. *Brain Structure and Function*, pages 1–18.

[Hau et al., 2016] Hau, J., Sarubbo, S., Perchey, G., Crivello, F., Zago, L., Mellet, E., Jobard, G., Joliot, M., Mazoyer, B. M., Tzourio-Mazoyer, N., and Petit, L. (2016). Cortical Terminations of the Inferior Fronto-Occipital and Uncinate Fasciculi: Anatomical Stem-Based Virtual Dissection. *Frontiers in Neuroanatomy*, 10.

[Heker et al., 2016] Heker, M., Amer, R., Alexandroni, G., and Greenspan, H. (2016). Automated supervised segmentation of anatomical fiber tracts using an AdaBoost framework. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5. IEEE. event-place: EILAT, Israel.

[Hua et al., 2008] Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C. M., and Mori, S. (2008). Tract Probability Maps in Stereotaxic Spaces: Analyses of White Matter Anatomy and Tract-Specific Quantification. *NeuroImage*, 39(1):336–347.

[Ioannidis, 2005] Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8):e124+.

[Jbabdi and Johansen-Berg, 2011] Jbabdi, S. and Johansen-Berg, H. (2011). Tractography: where do we go from here? *Brain Connectivity*, 1(3):169–183.

[Jbabdi et al., 2007] Jbabdi, S., Woolrich, M., Andersson, J., and Behrens, T. (2007). A Bayesian framework for global tractography. *NeuroImage*, 37:116–29.

[Jenkinson et al., 2002] 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(2):825–841.

[Jeurissen et al., 2019] Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2019). Diffusion MRI fiber tractography of the brain. *NMR in Biomedicine*, 32(4):e3785.

[Jeurissen et al., 2013] Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., and Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, 34(11):2747–2766.

[Jones, 2008] Jones, D. K. (2008). Studying connections in the living human brain with diffusion MRI. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 44(8):936–952.

[Jones and Nilsson, 2015] Jones, D. K. and Nilsson, M. (2015). Tractometry and the hunt for the missing link: a physicist perspective.

[Khan, 2014] Khan, F. A. (2014). *Biotechnology in Medical Sciences*. CRC Press.

[Kim et al., 2018] Kim, S., Bae, W. C., Masuda, K., Chung, C. B., and Hwang, D. (2018). Fine-Grain Segmentation of the Intervertebral Discs from MR Spine Images Using Deep Convolutional Neural Networks: BSU-Net. *Applied Sciences (Basel, Switzerland)*, 8(9).

[Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., and Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3):786–802.

[Klingler et al., 1935] Klingler, J., Klingler, J., and Klingler, J. P. (1935). Erleichterung der makroskopischen Präparation des Gehirn durch den Gefrierprozess.

[Labra et al., 2016] Labra, N., Guevara, P., Duclap, D., Houenou, J., Poupon, C., Mangin, J.-F., and Figueroa, M. (2016). Fast Automatic Segmentation of White Matter Streamlines Based on a Multi-Subject Bundle Atlas. *Neuroinformatics*, 15(1):71–86.

[Lazar, 2010] Lazar, M. (2010). MAPPING BRAIN ANATOMICAL CONNECTIVITY USING WHITE MATTER TRACTOGRAPHY. *NMR in biomedicine*, 23(7):821–835.

[Le Bihan, 1991] Le Bihan, D. (1991). Molecular diffusion nuclear magnetic resonance imaging. *Magnetic Resonance Quarterly*, 7(1):1–30.

[Le Bihan et al., 1986] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. (1986). MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology*, 161(2):401–407.

[Le Bihan et al., 1993] Le Bihan, D., Turner, R., and Douek, P. (1993). Is water diffusion restricted in human brain white matter? An echo-planar NMR imaging study. *Neuroreport*, 4(7):887–890.

[Maddah et al., 2005] Maddah, M., Mewes, A. U. J., Haker, S., Grimson, and Warfield, S. K. (2005). Automated atlas-based clustering of white matter fiber tracts from

DTMRI. In *Medical image computing and computer-assisted intervention–MICCAI 2005*, pages 188–195. Springer, Berlin, Heidelberg.

[Maddah et al., 2008] Maddah, M., Zollei, L., Grimson, Westin, C. F., and Wells, W. M. (2008). A mathematical framework for incorporating anatomical knowledge in DT-MRI analysis. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pages 105–108. IEEE. event-place: Paris, France.

[Maier-Hein et al., 2017] Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W. E., Glass, J. O., Chen, D. Q., Feng, Y., Gao, C., Wu, Y., Ma, J., Renjie, H., Li, Q., Westin, C.-F., Deslauriers-Gauthier, S., González, J. O. O., Paquette, M., St-Jean, S., Girard, G., Rheault, F., Sidhu, J., Tax, C. M. W., Guo, F., Mesri, H. Y., Dávid, S., Froeling, M., Heemskerk, A. M., Leemans, A., Boré, A., Pinsard, B., Bedetti, C., Desrosiers, M., Brambati, S., Doyon, J., Sarica, A., Vasta, R., Cerasa, A., Quattrone, A., Yeatman, J., Khan, A. R., Hodges, W., Alexander, S., Romascano, D., Barakovic, M., Auría, A., Esteban, O., Lemkaddem, A., Thiran, J.-P., Cetingul, H. E., Odry, B. L., Mailhe, B., Nadar, M. S., Pizzagalli, F., Prasad, G., Villalon-Reina, J. E., Galvis, J., Thompson, P. M., Requejo, F. D. S., Laguna, P. L., Lacerda, L. M., Barrett, R., Dell'Acqua, F., Catani, M., Petit, L., Caruyer, E., Daducci, A., Dyrby, T. B., Holland-Letz, T., Hilgetag, C. C., Stieltjes, B., and Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8(1).

[Mancall and Brock, 2011] Mancall, E. L. and Brock, D. G. (2011). *Gray's Clinical Neuroanatomy E-Book*. Elsevier Health Sciences. Google-Books-ID: R6ARC5tdhtYC.

[Mandelbrot, 1982] Mandelbrot, B. B. (1982). *The fractal geometry of nature*. W.H. Freeman, San Francisco.

[Martino and De Lucas, 2014] Martino, J. and De Lucas, E. M. (2014). Subcortical anatomy of the lateral association fascicles of the brain: A review. *Clinical Anatomy (New York, N.Y.)*, 27(4):563–569.

[Mayer et al., 2011] Mayer, A., Zimmerman-Moreno, G., Shadmi, R., Batikoff, A., and Greenspan, H. (2011). A supervised framework for the registration and segmentation of white matter fiber tracts. *IEEE transactions on medical imaging*, 30(1):131–145.

[Meynert, 1885] Meynert, T. (1885). *Psychiatry: A Clinical Treatise on Diseases of the Fore-Brain Based Upon a Study of Its Structure, Functions, and Nutrition.* GP Putnam's Sons.

[Moberts et al., 2005] Moberts, B., Vilanova, A., and van Wijk, J. J. (2005). Evaluation of Fiber Clustering Methods for Diffusion Tensor Imaging. In *VIS 05. IEEE Visualization, 2005.*, pages 65–72. IEEE. event-place: Minneapolis, MN, USA.

[Mori et al., 1999] Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of neurology*, 45(2):265–269.

[Mori et al., 2005] Mori, S., Wakana, S., Van Zijl, P. C. M., and Nagae-Poetscher, L. M. (2005). *MRI atlas of human white matter*, volume 16. Am Soc Neuroradiology, Amsterdam, The Netherlands.

[Neher et al., 2014] Neher, P. F., Laun, F. B., Stieltjes, B., and Maier-Hein, K. H. (2014). Fiberfox: facilitating the creation of realistic white matter software phantoms. *Magnetic resonance in medicine*, 72(5):1460–1470.

[O'Donnell et al., 2002] O'Donnell, L., Haker, S., and Westin, C.-F. (2002). New Approaches to Estimation of White Matter Connectivity in Diffusion Tensor MRI: Elliptic PDEs and Geodesics in a Tensor-Warped Space. In Dohi, T. and Kikinis, R., editors, *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002*, Lecture Notes in Computer Science, pages 459–466, Berlin, Heidelberg. Springer.

[O'Donnell et al., 2013] O'Donnell, L. J., Golby, A. J., and Westin, C.-F. (2013). Fiber clustering versus the parcellation-based connectome. *NeuroImage*, 80:283–289.

[O'Donnell et al., 2006] O'Donnell, L. J., Kubicki, M., Shenton, M. E., Dreusicke, M. H., Grimson, and Westin, C.-F. (2006). A method for clustering white matter fiber tracts. *American Journal of Neuroradiology*, 27(5):1032–1036.

[O'Donnell et al., 2012] O'Donnell, L. J., Wells, W. M., Golby, A. J., and Westin, C.-F. F. (2012). Unbiased groupwise registration of white matter tractography. *Medical image computing and computer-assisted intervention : MICCAI - International Conference on Medical Image Computing and Computer-Assisted Intervention*, 15(Pt 3):123–130.

[O'Donnell and Westin, 2011] O'Donnell, L. J. and Westin, C.-F. (2011). An introduction to diffusion tensor image analysis. *Neurosurgery Clinics of North America*, 22(2):185–196, viii.

[O'Donnell and Westin, 2007] O'Donnell, L. J. and Westin, C.-F. F. (2007). Automatic tractography segmentation using a high-dimensional white matter atlas. *Medical Imaging, IEEE Transactions on*, 26(11):1562–1575.

[Ogawa et al., 1990] Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872.

[Oishi et al., 2008] Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., Akhter, K., Hua, K., Woods, R., Toga, A. W., Pike, G. B., Rosa-Neto, P., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P. C. M., Mazziotta, J., and Mori, S. (2008). Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. *NeuroImage*, 43(3):447–457.

[Olivetti and Avesani, 2011] Olivetti, E. and Avesani, P. (2011). Supervised segmentation of fiber tracts. In *Proceedings of the First international conference on Similarity-based pattern recognition*, SIMBAD'11, pages 261–274, Berlin, Heidelberg. Springer-Verlag. event-place: Venice, Italy.

[Olivetti et al., 2017] Olivetti, E., Bertò, G., Gori, P., Sharmin, N., and Avesani, P. (2017). Comparison of distances for supervised segmentation of white matter tractography. In *2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, pages 1–4. IEEE. event-place: Toronto, ON, Canada.

[Olivetti et al., 2012] Olivetti, E., Nguyen, T. B., and Garyfallidis, E. (2012). The Approximation of the Dissimilarity Projection. *IEEE International Workshop on Pattern Recognition in NeuroImaging*, 0:85–88.

[Olivetti et al., 2016] Olivetti, E., Sharmin, N., and Avesani, P. (2016). Alignment of Tractograms As Graph Matching. *Frontiers in Neuroscience*, 10.

[Pandit, 2013] Pandit, A. S. (2013). Diffusion magnetic resonance imaging in preterm brain injury | SpringerLink.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Pekalska and Duin, 2005] Pekalska, E. and Duin, R. P. W. (2005). *The Dissimilarity Representation for Pattern Recognition: Foundations And Applications (Machine Perception and Artificial Intelligence).* World Scientific Publishing Company, Singapore. Published: Hardcover.

[Pernet and Poline, 2015] Pernet, C. and Poline, J.-B. (2015). Improving functional magnetic resonance imaging reproducibility. *GigaScience*, 4(1).

[Pestilli, 2018] Pestilli, F. (2018). Human white matter and knowledge representation. *PLoS Biology*, 16(4).

[Pestilli et al., 2014] Pestilli, F., Yeatman, J. D., Rokem, A., Kay, K. N., and Wandell, B. A. (2014). Evaluation and statistical inference for human connectomes. *Nature Methods*, 11(10):1058–1063.

[Petit et al., 2019] Petit, L., Rheault, F., Descoteaux, M., and Tzourio-Mazoyer, N. (2019). Half of the streamlines built in a whole human brain tractogram is anatomically uninterpretable. OHBM abstract.

[Pierpaoli et al., 1996] Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A., and Di Chiro, G. (1996). Diffusion tensor MR imaging of the human brain. *Radiology*, 201(3):637–648.

[Porro-Muñoz et al., 2015] Porro-Muñoz, D., Olivetti, E., Sharmin, N., Nguyen, T., Garyfallidis, E., and Avesani, P. (2015). Tractome: a visual data mining tool for brain connectivity analysis. *Data Mining and Knowledge Discovery*, 29(5):1258–1279.

[Poupon et al., 2008] Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., and Mangin, J.-F. (2008). New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (HARDI) models. *Magnetic Resonance in Medicine*, 60(6):1276–1283.

[Quionero-Candela et al., 2009] Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., and Lawrence, N. D. (2009). *Dataset Shift in Machine Learning.* The MIT Press.

[Raffelt et al., 2011] Raffelt, D., Tournier, Fripp, J., Crozier, S., Connelly, A., and Salvado, O. (2011). Symmetric Diffeomorphic Registration of Fibre Orientation Distributions. *NeuroImage.*

[Raffelt et al., 2012] Raffelt, D., Tournier, J.-D., Crozier, S., Connelly, A., and Salvado, O. (2012). Reorientation of fiber orientation distributions using apodized point spread functions. *Magnetic Resonance in Medicine*, 67(3):844–855.

[Reisert et al., 2011] Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., and Kiselev, V. (2011). Global fiber reconstruction becomes practical. *NeuroImage*, 54(2):955–962.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241. Springer International Publishing.

[Sabour et al., 2017] Sabour, S., Frosst, N., and Hinton, G. E. (2017). Dynamic Routing Between Capsules. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 3859–3869, USA. Curran Associates Inc. event-place: Long Beach, California, USA.

[Sani et al., 2019] Sani, I., McPherson, B. C., Stemmann, H., Pestilli, F., and Freiwald, W. A. (2019). Functionally defined white matter of the macaque monkey brain reveals a dorso-ventral attention network. *eLife*, 8:e40520.

[Sarubbo et al., 2016] Sarubbo, S., Benedictis, A. D., Merler, S., Mandonnet, E., Barbareschi, M., Dallabona, M., Chioffi, F., and Duffau, H. (2016). Structural and functional integration between dorsal and ventral language streams as revealed by blunt dissection and direct electrical stimulation. *Human Brain Mapping*, 37(11):3858–3872.

[Sarubbo et al., 2013] Sarubbo, S., De Benedictis, A., Maldonado, I., Basso, G., and Duffau, H. (2013). Frontal terminations for the inferior fronto-occipital fascicle: anatomical dissection, DTI study and functional considerations on a multi-component bundle. *Brain Structure and Function*, 218(1):21–37.

[Sarubbo et al., 2015] Sarubbo, S., De Benedictis, A., Milani, P., Paradiso, B., Barbareschi, M., Rozzanigo, U., Colarusso, E., Tugnoli, V., Farneti, M., Granieri, E., Duffau, H., and Chioffi, F. (2015). The course and the anatomo-functional relationships of the optic radiation: a combined study with 'post mortem' dissections and 'in vivo' direct electrical mapping. *Journal of Anatomy*, 226(1):47–59.

[Savadjiev et al., 2008] Savadjiev, P., Campbell, J. S. W., Descoteaux, M., Deriche, R., Pike, G. B., and Siddiqi, K. (2008). Labeling of ambiguous subvoxel fibre bundle configurations in high angular resolution diffusion MRI. *NeuroImage*, 41(1):58–68.

[Schmidt et al., 2017] Schmidt, M., Le Roux, N., and Bach, F. (2017). Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162(1):83–112.

[Sharmin, 2017] Sharmin, N. (2017). *CORRESPONDENCE AMONG CONNECTOMES AS COMBINATORIAL OPTIMIZATION*. PhD Thesis.

[Sharmin et al., 2016] Sharmin, N., Olivetti, E., and Avesani, P. (2016). Alignment of Tractograms as Linear Assignment Problem. In Fuster, A., Ghosh, A., Kaden, E., Rathi, Y., and Reisert, M., editors, *Computational Diffusion MRI*, Mathematics and Visualization, pages 109–120. Springer International Publishing.

[Sharmin et al., 2018] Sharmin, N., Olivetti, E., and Avesani, P. (2018). White Matter Tract Segmentation as Multiple Linear Assignment Problems. *Frontiers in Neuroscience*, 11.

[Siless et al., 2016] Siless, V., Chang, K., Fischl, B., and Yendiki, A. (2016). Hierarchical Clustering of Tractography Streamlines Based on Anatomical Similarity. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 184–191. Springer. event-place: Athens, Greece.

[Siless et al., 2018] Siless, V., Chang, K., Fischl, B., and Yendiki, A. (2018). AnatomiCuts: Hierarchical clustering of tractography streamlines based on anatomical similarity. *NeuroImage*, 166:32–45.

[Siless et al., 2013] Siless, V., Medina, S., Varoquaux, G., and Thirion, B. (2013). A Comparison of Metrics and Algorithms for Fiber Clustering. In *2013 International Workshop on Pattern Recognition in Neuroimaging*, pages 190–193. IEEE. event-place: Philadelphia, PA, USA.

[Smith et al., 2012] Smith, R. E., Tournier, J.-D., Calamante, F., and Connelly, A. (2012). Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage*, 62(3):1924–1938.

[Smith et al., 2013] Smith, R. E., Tournier, J.-D. D., Calamante, F., and Connelly, A. (2013). SIFT: Spherical-deconvolution informed filtering of tractograms. *NeuroImage*, 67:298–312.

[Sotiropoulos et al., 2013] Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., Glasser, M. F., Hernandez, M., Sapiro, G., Jenkinson, M., Feinberg, D. A., Yacoub, E., Lenglet, C., Van Essen, D. C., Ugurbil, K., Behrens, T. E., and WU-Minn HCP Consortium (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, 80:125–143.

[Southall et al., 2000] Southall, B., Buxton, B. F., Marchant, J. A., and Hague, T. (2000). On the performance characterisation of image segmentation algorithms: a case study. In *European Conference on Computer Vision*, pages 351–365. Springer. event-place: Dublin, Ireland.

[Sporns et al., 2005] Sporns, O., Tononi, G., and Kötter, R. (2005). The Human Connectome: A Structural Description of the Human Brain. *PLoS Comput Biol*, 1(4):e42+.

[Stacke et al., 2019] Stacke, K., Eilertsen, G., Unger, J., Lundstr&#xF6, C., and m (2019). A closer look at domain shift for deep learning in histopathology.

[Stevens and Antigua, 2019] Stevens, E. and Antigua, L. (2019). Deep Learning with PyTorch.

[Takemura et al., 2016] Takemura, H., Caiafa, C. F., Wandell, B. A., and Pestilli, F. (2016). Ensemble Tractography. *PLoS computational biology*, 12(2).

[Thiebaut de Schotten et al., 2011] Thiebaut de Schotten, M., Ffytche, D. H., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., Murray, R., Williams, S. C., Murphy, D. G. M., and Catani, M. (2011). Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. *NeuroImage*, 54(1):49–59.

[Thomas et al., 2014] Thomas, C., Ye, F. Q., Irfanoglu, M. O., Modi, P., Saleem, K. S., Leopold, D. A., and Pierpaoli, C. (2014). Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proceedings of the National Academy of Sciences of the United States of America*, 111(46):16574–16579.

[Tournier et al., 2012] Tournier, J., Calamante, F., Connelly, A., and Others (2012). MRtrix: diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology*, 22(1):53–66.

[Tournier et al., 2007] Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472.

[Tournier et al., 2010] Tournier, J. D., Calamante, F., and Connelly, A. (2010). Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In *Proc. 18th Annual Meeting of the Intl. Soc. Mag. Reson. Med.(ISMRM)*, page 1670.

[Tournier et al., 2004] Tournier, J.-D. D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185.

[Tuch, 2004] Tuch, D. S. (2004). Q-ball imaging. *Magnetic resonance in medicine*, 52(6):1358–1372.

[Tunç et al., 2014] Tunç, B., Parker, W. A., Ingalhalikar, M., and Verma, R. (2014). Automated tract extraction via atlas based adaptive clustering. *Neuroimage*, 102:596–607.

[Van Essen et al., 2013] Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, 80:62–79.

[Van Essen et al., 2012] Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S. W., and Others (2012). The Human Connectome Project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231.

[Vercruysse et al., 2014] Vercruysse, D., Christiaens, D., Maes, F., Sunaert, S., and Suetens, P. (2014). Fiber Bundle Segmentation Using Spectral Embedding and Supervised Learning. In *Computational Diffusion MRI*, pages 103–114. Springer, Boston, MA.

[von Luxburg and Schoelkopf, 2008] von Luxburg, U. and Schoelkopf, B. (2008). *Statistical Learning Theory: Models, Concepts, and Results.*

[Vu et al., 2018] Vu, M.-A. T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V. S., Widge, A. S., Mayberg, H. S., Sapiro, G., and Dzirasa, K. (2018). A Shared Vision for Machine Learning in Neuroscience. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 38(7):1601–1607.

[Wakana et al., 2007] Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., and Others (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage*, 36(3):630–644.

[Wandell, 2016] Wandell, B. A. (2016). Clarifying Human White Matter. *Annual Review of Neuroscience*, 39(1):103–128.

[Wang et al., 2007] Wang, R., Benner, T., Sorensen, A. G., and Wedeen, V. J. (2007). Diffusion toolkit: a software package for diffusion imaging data processing and tractography. In *Proc Intl Soc Mag Reson Med*, volume 15. Berlin.

[Wang et al., 2011] Wang, X., Grimson, W. E., and Westin, C.-F. F. (2011). Tractography segmentation using a hierarchical Dirichlet processes mixture model. *NeuroImage*, 54(1):290–302.

[Wassermann et al., 2010] Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., and Deriche, R. (2010). Unsupervised white matter fiber clustering and tract probability map generation: applications of a Gaussian process framework for white matter fibers. *NeuroImage*, 51(1):228–241.

[Wassermann et al., 2013] Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., and Westin, C.-F. F. (2013). On describing human white matter anatomy: the white matter query language. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16(Pt 1):647–654.

[Wassermann et al., 2016] Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., and Westin, C.-F. F. (2016). The white matter query language: a novel approach for describing human white matter anatomy. *Brain structure & function*, 221:4705–4721.

[Wassermann et al., 2011] Wassermann, D., Rathi, Y., Bouix, S., Kubicki, M., Kikinis, R., Shenton, M., and Westin, C.-F. F. (2011). White matter bundle registration and population analysis based on Gaussian processes. *Information processing in medical imaging : proceedings of the ... conference*, 22:320–332.

[Wasserthal et al., 2018a] Wasserthal, J., Neher, P., and Maier-Hein, K. H. (2018a). TractSeg - Fast and accurate white matter tract segmentation. *NeuroImage*, 183:239–253.

[Wasserthal et al., 2018b] Wasserthal, J., Neher, P. F., and Maier-Hein, K. H. (2018b). Tract Orientation Mapping for Bundle-Specific Tractography. In Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C., and Fichtinger, G., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Lecture Notes in Computer Science, pages 36–44, Cham. Springer International Publishing.

[Wedeen et al., 2005] Wedeen, V. J., Hagmann, P., Tseng, W.-Y. I., Reese, T. G., and Weisskoff, R. M. (2005). Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic resonance in medicine*, 54(6):1377–1386.

[Wei et al., 2019] Wei, Z., Zhang, J., Liu, L., Zhu, F., Shen, F., Zhou, Y., Liu, S., Sun, Y., and Shao, L. (2019). Building Detail-Sensitive Semantic Segmentation Networks With Polynomial Pooling. pages 7115–7123.

[Wilkinson et al., 2016] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):1–9.

[Witelson, 1989] Witelson, S. F. (1989). Hand and sex differences in the isthmus and genu of the human corpus callosum. A postmortem morphological study. *Brain: A Journal of Neurology*, 112 (Pt 3):799–835.

[Wu et al., 2007] Wu, J.-S., Zhou, L.-F., Tang, W.-J., Mao, Y., Hu, J., Song, Y.-Y., Hong, X.-N., and Du, G.-H. (2007). Clinical evaluation and follow-up outcome of diffusion tensor imaging-based functional neuronavigation: a prospective, controlled study in patients with gliomas involving pyramidal tracts. *Neurosurgery*, 61(5):935–948; discussion 948–949.

[Wu et al., 2016a] Wu, Y., Sun, D., Wang, Y., and Wang, Y. (2016a). Subcomponents and Connectivity of the Inferior Fronto-Occipital Fasciculus Revealed by Diffusion Spectrum Imaging Fiber Tracking. *Frontiers in Neuroanatomy*, 10.

[Wu et al., 2016b] Wu, Y., Sun, D., Wang, Y., Wang, Y., and Wang, Y. (2016b). Tracing short connections of the temporo-parieto-occipital region in the human brain using diffusion spectrum imaging and fiber dissection - ScienceDirect. *Brain Research*, 1646(0006-8993):152–159.

[Yeatman et al., 2012] Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., and Feldman, H. M. (2012). Tract Profiles of White Matter Properties: Automating Fiber-Tract Quantification. *PLoS ONE*, 7(11):e49790+.

[Yeatman et al., 2018] Yeatman, J. D., Richie-Halford, A., Smith, J. K., Keshavan, A., and Rokem, A. (2018). A browser-based tool for visualization and analysis of diffusion MRI data. *Nature Communications*, 9(1):1–10.

[Yeh et al., 2018] Yeh, F.-C., Panesar, S., Fernandes, D., Meola, A., Yoshino, M., Fernandez-Miranda, J. C., Vettel, J. M., and Verstynen, T. (2018). Population-averaged atlas of the macroscale human structural connectome and its network topology. *NeuroImage*, 178:57–68.

[Yendiki et al., 2011] Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., Jbabdi, S., Gollub, R., and Fischl, B. (2011). Automated Probabilistic Reconstruction of White-Matter Pathways in Health and Disease Using an Atlas of the Underlying Anatomy. *Frontiers in Neuroinformatics*, 5.

[Yoo et al., 2015] Yoo, S. W., Guevara, P., Jeong, Y., Yoo, K., Shin, J. S., Mangin, J.-F., and Seong, J.-K. (2015). An Example-Based Multi-Atlas Approach to Automatic Labeling of White Matter Tracts. *PloS one*, 10(7).

[Zemmoura et al., 2016] Zemmoura, I., Blanchard, E., Raynal, P.-I., Rousselot-Denis, C., Destrieux, C., and Velut, S. (2016). How Klingler's dissection permits exploration of brain structural connectivity? An electron microscopy study of human white matter. *Brain Structure & Function*, 221(5):2477–2486.

[Zhang et al., 2006] Zhang, L., Liu, J. Z., Dean, D., Sahgal, V., and Yue, G. H. (2006). A three-dimensional fractal analysis method for quantifying white matter structure in human brain. *Journal of Neuroscience Methods*, 150(2):242–253.

[Zhang et al., 2008] Zhang, S., Correia, S., and Laidlaw, D. H. (2008). Identifying White-Matter Fiber Bundles in DTI Data Using an Automated Proximity-Based Fiber-Clustering Method. *Visualization and Computer Graphics, IEEE Transactions on*, 14(5):1044–1053.

[Zhang et al., 2019] Zhang, X., Feng, Y., Chen, W., Li, X., Faria, A. V., Feng, Q., and Mori, S. (2019). Linear Registration of Brain MRI Using Knowledge-Based Multiple Intermediator Libraries. *Frontiers in Neuroscience*, 13.

[Zhang et al., 2010] Zhang, Y., Zhang, J., Oishi, K., Faria, A. V., Jiang, H., Li, X., Akhter, K., Rosa-Neto, P., Pike, G. B., Evans, A., and Others (2010). Atlas-guided tract reconstruction for automated and comprehensive examination of the white matter anatomy. *Neuroimage*, 52(4):1289–1301.