UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Elétrica e de Computação

Thais de Oliveira Caldeira

Corpus callosum studies in diffusion MRI: a web-based software for data processing, exploration and visualization

Estudos do corpo caloso em imagens de tensor de difusão: uma aplicação web para processamento, exploração e visualização de dados

Campinas

2021

Thais de Oliveira Caldeira

Corpus callosum studies in diffusion MRI: a web-based software for data processing, exploration and visualization

Estudos do corpo caloso em imagens de tensor de difusão: uma aplicação web para processamento, exploração e visualização de dados

Dissertation presented to the School of Electrical and Computer Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Electrical Engineering, in the area of Computing Engineering.

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica, na Área de Engenharia da Computação.

Orientadora: Prof. Dr. Leticia Rittner

Este trabalho corresponde à versão final da dissertação defendida pela aluna Thais de Oliveira Caldeira, e orientada pela Prof. Dr. Leticia Rittner.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Elizangela Aparecida dos Santos Souza - CRB 8/8098

C127c

Caldeira, Thais de Oliveira, 1995-
Corpus callosum studies in diffusion MRI : a web-based software for data processing, exploration and visualization / Thais de Oliveira Caldeira. – Campinas, SP : [s.n.], 2021.

Orientador: Leticia Rittner.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Corpo caloso. 2. Imagem de tensor de difusão. 3. Exploração de dados. 4. Visualização da informação. 5. Software livre. I. Rittner, Leticia, 1972-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estudos do corpo caloso em imagens de tensor de difusão : uma aplicação web para processamento, exploração e visualização de dados
Palavras-chave em inglês:
Corpus callosum
Diffusion tensor imaging
Data exploration
Data visualization
Open-source software
Área de concentração: Engenharia de Computação
Titulação: Mestra em Engenharia Elétrica
Banca examinadora:
Leticia Rittner [Orientador]
Paula Dornhofer Paro Costa
Brunno Machado de Campos
Data de defesa: 09-04-2021
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-0550-406X
- Currículo Lattes do autor: http://lattes.cnpq.br/8566640651571442

COMISSÃO JULGADORA: DISSERTAÇÃO DE MESTRADO

Candidato(a): Thais de Oliveira Caldeira **RA:** 229984
Data de defesa: 09 de ABRIL de 2021
Título da Dissertação: "Corpus callosum studies in diffusion MRI: a web-based software for data processing, exploration and visualization"

Comissão:
Profa. Dra. Leticia Rittner (Presidente)
Profa. Dra. Paula Dornhofer Paro Costa
Prof. Dr. Brunno Machado de Campos

A Ata de Defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

Agradecimentos

Agradeço primeiramente à minha orientadora, Leticia, sua gigantesca paciência e empolgação, e acertado direcionamento ao longo do desenvolvimento deste trabalho. Obrigada principalmente pela confiança e toda sua disposição.

À Universidade Estadual de Campinas e à minha querida *Alma Mater*, a Universidade Federal do Espírito Santo, seus professores e funcionários, que ao longo dos meus anos formativos me criaram na diversidade, me oferecendo acesso à diferentes mundos, oportunidades e, finalmente, educação pública, gratuita e de qualidade.

Aos meus chegados, minha fantástica família e amigos, eu amo vocês.

À banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Código de Financiamento 001.

diversonagens suspersas

Meu verso, temo, vem do berço.
Não versejo porque eu quero,
versejo quando converso
e converso por conversar.
Pra que sirvo senão pra isto,
pra ser vinte e pra ser visto,
pra ser versa e pra ser vice,
pra ser a supersuperfície
onde o verbo vem ser mais?

Não sirvo para observar.
Verso, persevero e conservo
um susto de quem se perde
no exato lugar onde está.

Onde estará meu verso?
Em algum lugar de um lugar,
onde o avesso do inverso
começa a ver e ficar.
Por mais prosas que eu perverta,
não permita Deus que eu perca
meu jeito de versejar.

(Paulo Leminski)

Resumo

O corpo caloso (CC) é a principal comissura interhemisférica do cérebro e, devido às suas fibras altamente organizadas, é frequentemente estudado em imagens de tensores de difusão (*diffusion tensor images*, ou DTI). DTI, uma modalidade de ressonância magnética, é um método não invasivo para mapear a difusão de água nos tecidos, o que o torna útil para mapear fibras, permitindo a quantificação de tratos *in vivo*.

Os estudos do CC em DTI geralmente dependem de sua segmentação - determinação das bordas da estrutura - e parcelamento - divisão da estrutura em diferentes partes, de acordo com as regiões corticais com as quais estão interligadas. No entanto, essas tarefas apresentam dificuldades que podem prejudicar a qualidade da pesquisa. As imagens de difusão geralmente apresentam baixa resolução espacial, baixa relação sinal-ruído e contraste, e imagens adquiridas em diferentes scanners também podem apresentar variabilidade de intensidade. Embora alguns métodos de segmentação e de parcelamento do CC em imagens de difusão tenham sido propostos, poucos são completamente automáticos e a maioria não tem implementação disponível.

Considerando este contexto, o presente trabalho propõe uma ferramenta de código aberto, portável e interativa para análise do corpo caloso em DTI individualmente ou em grupos, implementando diferentes técnicas disponíveis para segmentação e parcelamento e propondo métricas relevantes para comparação e avaliação da qualidade desses procedimentos em uma aplicação web. Também propõe diferentes tipos de visualizações interativas das informações extraídas do DTI, para permitir a exploração e visualização destes dados.

A ferramenta desenvolvida é uma aplicação web *open-source*, portável e pode ser usada nos principais sistemas operacionais – Linux, além de Windows e Mac usando Docker –, disponível em <https://github.com/thaiscaldeira/inCCsight>. Este trabalho apresenta também um estudo de caso aplicando a ferramenta a um estudo sobre Lupus Eritomastoso Sistêmico a partir de DTI, onde são discutidos os impactos da ferramenta proposta nas conclusões do estudo.

Palavras-chaves: Corpo Caloso; Imagem de tensor de difusão; Exploração de dados; Visualização de dados; Ferramenta de código aberto.

Abstract

The corpus callosum (CC) is the main interhemispheric commissure and, due to its highly organized fibers, is frequently studied in diffusion tensor images. Diffusion tensor imaging (DTI), a type of magnetic ressonance imaging, is a non-invasive method for mapping the diffusion of water in tissues, which makes it useful for mapping tracts, allowing in-vivo tract quantification.

Studies of the CC using DTI usually depend on its segmentation – determination of the CC borders – and parcellation – division of the structure in different parts, according with the cortical regions with which they are interconnected, but these tasks present difficulties that can jeopardise research quality. Diffusion images usually present low resolution, definition and contrast, and images acquired in different scanners might also present intensity variability. Although several diffusion CC segmentation and parcellation methods have been proposed, few are automatic and most have no implementation available.

Considering this context, the present work proposes an open-source, portable, tool for analysis of the corpus callosum in diffusion tensor images individually or in groups, implementing different techniques available for segmentation and parcellation and proposing relevant metrics for comparison and quality evaluation of these procedures in a graphical interface. It also proposes different types of interactive visualizations of extracted information from the DTI, to allow data exploration.

The developed tool is an *open-source*, web-based application, portable and can be used in the most used operating systems - Linux, in addition to Windows and Mac using Docker -, available at <https://github.com/thaiscaldeira/inCCsight>. This work also presents a case study applying the tool to a study on Systemic Lupus Erythemastous from DTI, where the impacts of the proposed tool on the study conclusions are discussed.

Keywords: Corpus Callosum; Diffusion tensor imaging; Data exploration; Data visualization; Open-source software.

List of Figures

List of Tables

List of abbreviations and acronyms

AD	Axial Diffusivity (Difusividade Axial)
CC	Corpus Callosum
CSF	Cerebral Spinal Fluid
DTI	Diffusion Tensor Image
DWI	Diffusion-Weighted Imaging
FA	Fractional Anisotropy
FMRIB	Oxford Centre for Functional MRI of the Brain
FSL	FMRIB Software Library
GM	Gray Matter
MD	Mean Diffusivity
MRI	Magnetic Resonance Image
RD	Radial Diffusivity
ROQS	Reproducible Objective Quantification Scheme
SNR	Signal-to-Noise Ratio
WM	White Matter

Contents

1 INTRODUCTION

1.1 Corpus callosum

The corpus callosum (CC) is the major interhemispheric commissure – a bundle of fibers that connects both brain hemispheres – that connects most of the cortical areas and is the largest white matter structure in the human brain (ABOITIZ; MONTIEL, 2003). Several studies have related CC characteristics with sex (MITCHELL et al., 2003), intelligence (LUDERS et al., 2007) and laterality (WITELSON, 1989). Furthermore, it has also been related to several conditions, including but not limited to: autism (EGAAS et al., 1995), epilepsy (O'DWYER et al., 2010), different psychiatric conditions such as depression (WALTERFANG et al., 2009) and schizophrenia (JOSHI et al., 2013) and different neurodegenerative diseases, including Alzheimer's (TEIPEL et al., 2002) and Parkinson's disease (WILTSHIRE et al., 2005).

1.2 Studies of the corpus callosum on diffusion tensor images

Diffusion Tensor Imaging (DTI), a modality of magnetic resonance imaging, is a non-invasive method for mapping water diffusion in tissues, making it useful for mapping tracts and allowing *in-vivo* tract exploration (BASSER et al., 2000). It is, therefore, susceptible to subtle differences in the architecture of white matter at the microstructural level (LEBEL; DEONI, 2018) and can be very useful in CC analysis. The diffusion tensor image is composed of tensors that define the magnitude, anisotropy (variation of the diffusion properties with direction), and orientation of water diffusion in biological tissues.

Studies of the CC based on DTI usually try to find and relate variations anatomically or in the diffusiveness parameters with a certain condition. Therefore, these studies depend on the segmentation of the corpus callosum – determination of the CC borders – and parcellation – division of the structure in different parts, according to the cortical regions with which they are interconnected. Still, these tasks present difficulties that can jeopardize research quality. Diffusion images usually present low spatial resolution, low signal-to-noise ratio (SNR) and contrast, and images acquired in different scanners

might also present intensity variability. The low definition can also lead to the "partial volume effect" that aggravates the definition of CC borders (HE et al., 2007). At the same time, the CC presents high variability among subjects, which limits region or shape-based segmentation methods. Additionally, proximity and similarity with other structures and thin areas at the CC central zone might lead to an undesired partition of the structure in the segmentation process (HERRERA et al., 2019).

Recent access to larger datasets has required the use of fully automatic segmentation methods, which are available for DTI studies of the CC but are often non-generalizable, case-specific, or lack quantitative and qualitative metrics that would allow results comparison. Also, most of the methods are described in the form of an algorithm in the literature, but there is no publicly available implementation (COVER et al., 2018). Our hypothesis is that by not enforcing the quality of the automatic segmentation methods or simply by using different methodologies, one might reach different conclusions concerning the analyzed data (PEREIRA et al., 2018).

Furthermore, since there are no anatomical landmarks indicating where the CC should be parcellated, Witelson (1989), and different authors have proposed parcellation techniques which are based on distinct premises, failing to meet a consensus (HOFER; FRAHM, 2006; WITELSON, 1989; HUANG et al., 2005). Moreover, the methods available propose different parcellation approaches, dividing the CC according to different associated cortical areas. Since there is no gold standard nor recognized metrics, choosing, evaluating, and comparing different parcellation methods is a challenge by itself (COVER et al., 2017).

1.2.1 The usual pipeline for studies of the CC with DTI

The usual pipeline a researcher has to follow to conduct DTI-based CC studies, from acquisition to statistical analysis, can be divided into the following steps (Fig. 1.1):

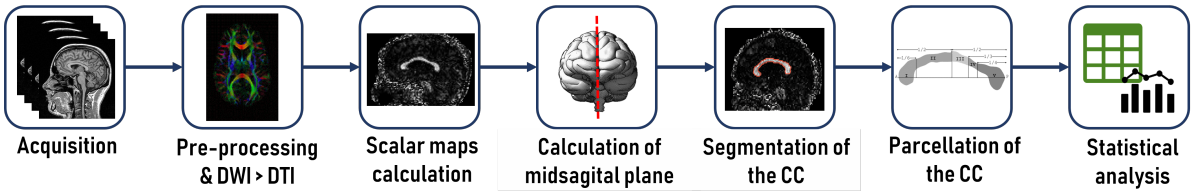


Figure 1.1 – Summarized scheme of the pipeline for DTI-based Corpus Callosum studies

Acquisition Selected subjects undergo an MRI scan, where we obtain the diffusion-weighted images (DWI);

Preprocessing DWI files are then anonymized and pre-processed, usually using software such as FSL (JENKINSON et al., 2012) for eddy currents correction, registration of the DWI volumes, and computation of the diffusion tensor images (DTI) and their eigenvectors/eigenvalues;

Scalar map calculation The eigenvectors and respective values are used to calculate scalar maps, reducing the dimensionality of the data and allowing a more practical analysis. The most commonly used are Fractional Anisotropy (FA), Mean Diffusivity (MD), Axial Diffusivity (AD), and Radial Diffusivity (RD);

Calculation of the midsagittal or midcallosal plane Most studies are concerned with the extreme central section of the corpus callosum, where its fibers are the most organized. Some studies find the midcallosal plane, considering only the CC, and others use the midsagittal plane, which considers the whole brain for its calculation;

Segmentation of the corpus callosum The borders of the structure are defined so that we can obtain statistical data from that specific region;

Parcellation of the corpus callosum The structure is divided into different parts according to the cortical regions with each they are interconnected. There are different possible divisions proposed in the literature;

Statistical analysis We obtain statistical values for each of the scalar maps considering the delimited regions. Later these values are used to compare different groups statistically, such as one control and case groups.

Based on previous projects, informal interviews, and discussions with collaborating domain experts, some major flaws within the current pipeline were identified. The first aspect is that it requires several software, scripts, libraries to be completed, most of them focusing on a few of the pipeline steps, as described on Sec. 2.5. Although the need of several different software does not imply in irreproducibility, it does add a complexity to the pipeline, making it prone to errors. Also, considering that end users are often health specialists, which might not have a background in technology, the need to install different packages, set up parameters before running studies and run commands from terminal can be another source of errors, jeopardizing reproducibility.

Considering the CC segmentation step, although several methods can be found (COVER et al., 2018), a lot are not easily reproducible or facilitates validation – such as the works from Nazem-Zadeh et al. (2012), Luis-García et al. (2011) and Niogi et al. (2007). Furthermore, when using automatic methods, ideally an expert should check the results one by one to guarantee that there are no silent errors among them. This is however, not a reality when considering large datasets, and errors should be mitigated with equally automated quality control methods, currently not available as a tool nor embedded in one for the segmentation assessment. All these aspects cited build a prevalent scenario in this line of research: works with low to none reproducibility characteristics.

Focusing specifically in the parcellation step, there is no gold-standard to be followed as there are no anatomical landmarks dividing the CC, and this obliges researchers to choose one of the proposed methods in the literature (see Sec. 2.4). All of those methods aim to divide the CC in regions (often in genu, body, isthmus and splenium parts), which depending on the chosen method will correspond to different interconnected cortex areas. This means that, when researchers draw conclusions relating characteristics of a parcel with a certain condition, they have to be sure of what the parcellation method used meant.

Within the multiple tools available for DTI processing and visualization, none can be applied to this specific type of region-based studies of the corpus callosum. Most available are focused in different applications such as multi-modal studies (PIEPER et al., 2004; GOEBEL, 2012; TOUSSAINT et al., 2007), tractography (WANG et al., 2007; TOURNIER et al., 2019; BAKHSHMAND et al., 2017) or atlas constructions (ZHANG et al., 2007). Others, however, have a different scope (considering the pipeline discussed previously) and focus for example on the preprocessing of DTI images, or on different ways to visualize the tensorial maps. Some, such as Freesurfer (FISCHL, 2012), can perform segmentation and parcellation of the corpus callosum, but does not perform CC segmentation quality control, implements different parcellation methods, nor allows specific data exploration and visualization tools for DTI data. Finally, although there are several well-established tools for diffusion tensor imaging exploration, still there is a lack on tools that apply specifically to DTI statistical analysis of regions, specially for corpus callosum studies.

1.3 Objectives

Considering the difficulties related to studies of the corpus callosum on diffusion tensor imaging, the purpose of the present work is to develop an open-source application for DTI studies of the corpus callosum, comprising all steps for this type of study after obtaining the diffusion tensor images. The specific goals were:

- to include in the developed tool: (1) midcallosal calculation, (2) corpus callosum segmentation, (3) segmentation quality control, (4) corpus callosum parcellation, (5) data exploration and visualization;
- to implement and provide access to currently unavailable corpus callosum segmentation and parcellation methods;
- to integrate a step of automated quality assessment of resulting segmentations, allowing large datasets studies and avoiding the propagation of errors through the pipeline;
- to provide an easy-to-install software, that does not require further library installation nor complex set-up of any kind;
- to allow medical and health professionals and researchers to perform CC analysis in DTI with confidence, even when working with large datasets, by providing an intuitive graphical user interface;
- to propose new ways for visualizing DTI-based metrics within group analysis
- to introduce a new platform for DTI-based CC study, where researchers can find relations and patterns in data in an exploratory way.

1.4 Main contributions

The main contributions of this work are:

- Open-source implementation of different segmentation and parcellation methods, which were only described in literature;

- An updated review of tools for DTI processing and data visualization. The review is included in a manuscript submitted to the Computer & Graphics Journal, and it is already on a second round of review;
- Complete automation of an originally semi-automatic segmentation method (ROQS), eliminating the need to place initial seeds;
- An open-source, portable, web-based software for interactive processing, quality control, exploration and visualization of DTI data of the corpus callosum, supporting reproducibility of such studies. The tool is available on github at <https://github.com/MICLab-Unicamp/inCCsight>;

1.5 Work structure

The present work is organized as described:

Chapter 2: Background A theoretical background is presented, including to familiarize the reader with concepts required to fully understand the posterior chapters. This includes information about the structure studied, the corpus callosum; also about the imaging technique utilized, DTI; and finally, relating with the priors, the methods for segmentation and parcellation of the structure in DTI. This chapter also summarizes the related literature on segmentation and parcellation methods and DTI tools;

Chapter 3: inCCsight The developed tool is then presented, first by discussing the requirements used to conceive it, then elaborating in each of its features and parts, and concluding with technical considerations regarding its format and structure;

Chapter 4: Case study To better understand how the tool might be used, a case study conducted with inCCsight is presented;

Chapter 5: Discussion This section will discuss at which length the proposed objectives were achieved considering the proposed solution and its application on the case study;

Chapter 6: Conclusions Finally, the conclusions, publications and final remarks, considering possible improvements and future works, are presented.

2 BACKGROUND

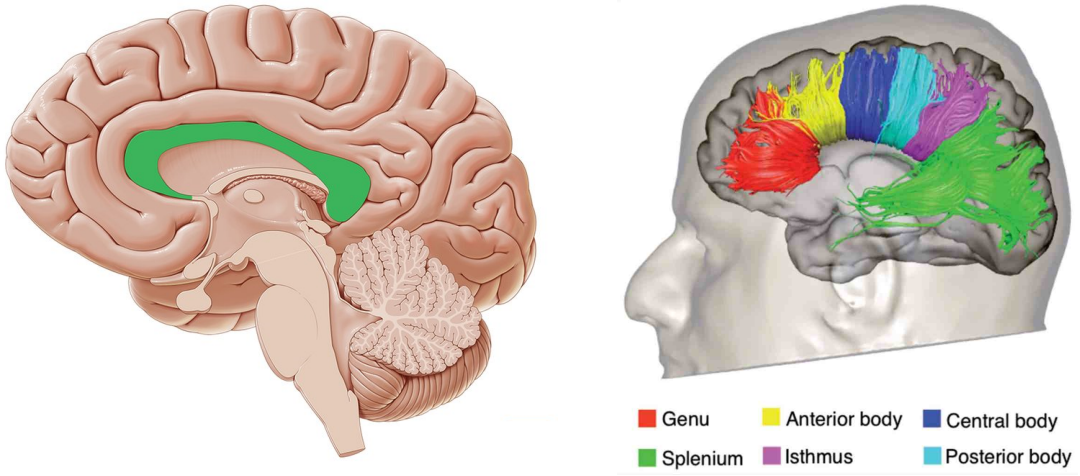
This chapter addresses the theoretical concepts necessary to understand the following chapters, and the relevant literature considered in the development of this work: first, the corpus callosum (CC), the structure of interest (Sec. 2.1); then, diffusion tensor imaging (MRI), the technique used to visualize and study the CC (Sec. 2.2); Sections 2.3 and 2.4 respectively, the CC segmentation and parcellation methods used will be studied; finally, different tools available for DTI and CC exploration will be discussed in Section 2.5.

2.1 Corpus callosum

The corpus callosum, which means *tough body* in Latin, is a bundle of fibers located underneath the cerebral cortex. It contains an average of about 200 million fibers that make it the largest white matter structure of the brain. It is also the largest commisure – fiber tracts that connect the two cerebral hemispheres and span the longitudinal fissure – and presents highly organized fibers (ABOITIZ; MONTIEL, 2003), that reflects in microstructural characteristics that can be used in the study of the structure.

There are several studies on the relationship between the size and shape of the CC and subject characteristics such as sex, age, numerical and mathematical skills, or even handedness. From a clinical point of view, the CC is affected by illness such as Alzheimer, autism, schizophrenia, dyslexia, epilepsy, multiple sclerosis, depression, among several other diseases. Also, the literature relates it with alcoholism, obesity and smoking (COVER et al., 2018).

It can be divided in regions according with the respective connected cortical regions (WITELSON, 1989): the most rostral region – in direction of the nose – of the corpus callosum, the genu and the rostrum, has connections between prefrontal brain regions (Fig. 2.1). The midsections between the genu and splenium are the body and isthmus, and connect the respectively zones responsible for motor, auditive and somatosensorial projections. The most caudal region – in direction to the back of the neck –, the splenium, contains connections between occipital, temporal and parietal regions.



(a) Corpus callosum highlighted in green in the sagittal plane. Source: Paul Kim, from Kenhub Gmgb[1]

(b) Regions of the corpus callosum and the connected cortical areas. Source: The Human Memory website[2]

Figure 2.1 – Visualizations of the corpus callosum and its connections

2.2 Diffusion tensor imaging

Diffusion weighted imaging (shortened as DWI, and also called diffusion MRI or simply diffusion) is a modality of MRI where contrast is based on differences in diffusion in the water molecules within the brain. The diffusion represents the random movement of the molecules (Brownian movement), and depends of several factors such as molecule type, temperature and micro-environment (BAMMER, 2003). In structures with highly oriented fibers such as brain tracts, the diffusion along the fibers is greater than the diffusion in any other orthogonal direction to them. This diffusion is known as anisotropic (HUISMAN, 2003) (opposed to isotropic diffusion where the molecules diffuse equally in all directions) .

In DWI, the Brownian movement of the molecules which affects the magnectic field applied, causes signal loss and is used to infer how water is moving. In order to figure the orientation of diffusion, gradients of the magnetic field are applied in various directions and combined to build a 3D diffusion map for each direction. This multi-dimensional map is complex, hard to be interpreted and rarely used in the medical practice. The diffusion tensor imaging (DTI) model was introduced to simplify the DWI acquisition (BAMMER, 2003). The diffusion values in all directions are combined, and every voxel (minimum volumetric element in a 3D image) is represented by the second-order tensor D on Eq. 2.1

[1] Available at: <https://www.kenhub.com/en/library/anatomy/corpus-callosum>. Accessed on: 1st march, 2021

[2] Available at: <https://human-memory.net/corpus-callosum/>. Accessed on: 1st march, 2021

describing the spatial diffusion of the volume (BIHAN et al., 2001).

$$\underline{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \tag{2.1}$$

The diagonal elements are the diffusion variances along the x, y and z axes, and the elements outside the diagonal are the covariance terms, with $D_{ij} = D_{ji}$. The diagonalization of the diffusion tensor produces the eigenvalues $(\lambda_1, \lambda_2, \lambda_3)$ and corresponding eigenvectors (e_1, e_2, e_3) of the diffusion tensor, which describe the apparent directions and diffusivities along the main axes. Diffusion tensor images are therefore images whose voxels store tensors, instead of scalar values, as in other imaging modalities. The diffusion tensor can be represented by an ellipsoid, with eigenvectors defining the directions of the main axes and the ellipsoidal radius defined by the eigenvalues (Fig. 2.2).

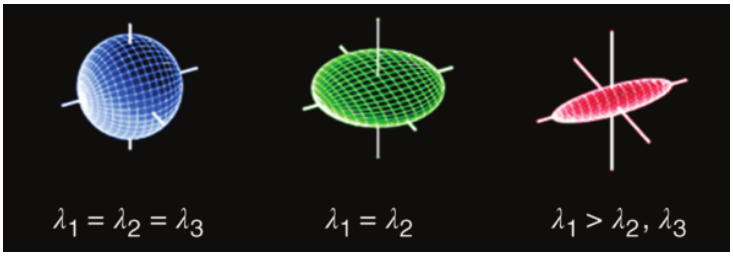


Figure 2.2 – Examples of different diffusion tensors represented by ellipsoids: isotropic (left); planar anisotropic (center); anisotropic (right). Source: Adapted from Descoteaux (2015)

The magnitudes of the eigenvalues can be affected by changes in the microstructure of the local tissue, such as many types of tissue damage, disease or normal physiological changes. Thus, the diffusion tensor is a useful parameter to characterize the normal and abnormal microstructure of the tissue.

2.2.1 DTI scalars: definition and interpretation

The major eigenvector (e_1), corresponding to the largest eigenvalue (λ_1) is the direction of fastest diffusivity and is generally assumed to be parallel to the direction of axon bundles in white matter. The medium and smallest eigenvalues (λ_2 and λ_3, respectively) are assumed to be perpendicular to the white matter tracts. From the eigenvalues one can extract the most common measures used in DTI studies, reducing the dimensionality of the data by

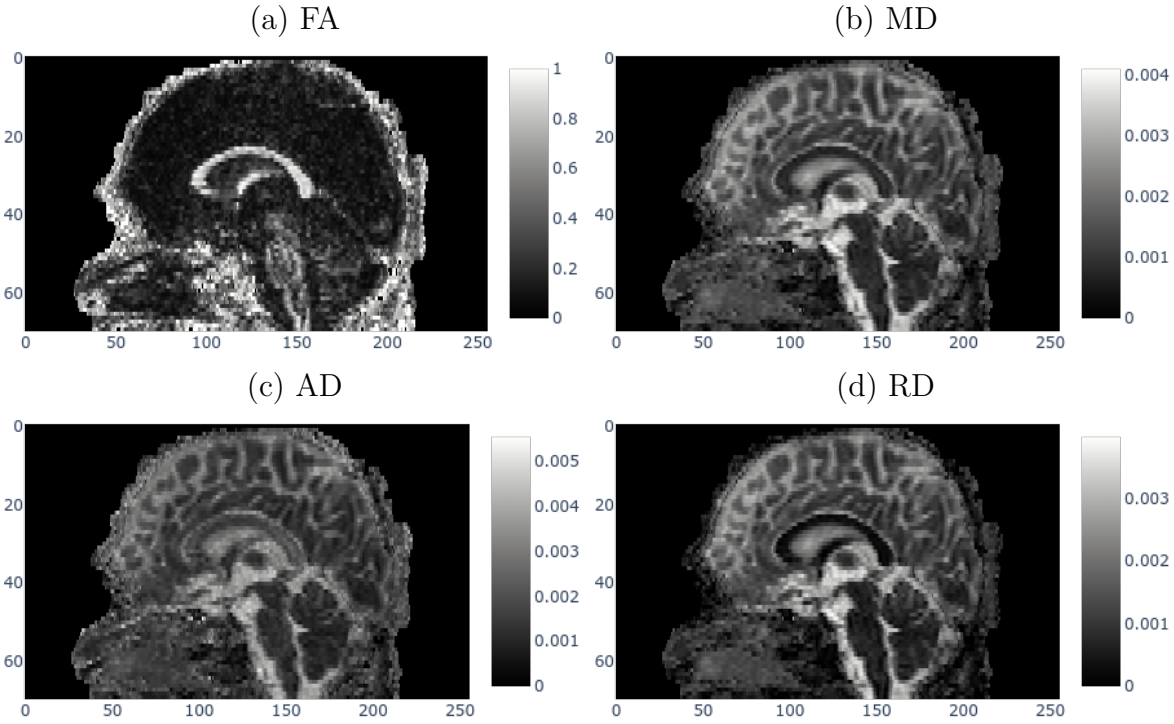


Figure 2.3 – DTI scalar maps of the midsagittal plane.

calculating scalar maps (Fig.2.3). A brief description of how each one is calculated and how they can be related to brain tissue characteristics is presented below (TROMP, 2013):

Fractional Anisotropy (FA) is the normalized version of the eigenvalues standard deviation and is a summary measure of microstructural integrity. While FA is highly sensitive to microstructural changes, it is less specific to the type of change (Fig. 2.3a).

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{2.2}$$

Mean Diffusivity (MD) is the average of the three eigenvalues. It is an inverse measure of the membrane density, is very similar for both gray matter and white matter and higher for cerebrospinal fluid. MD is sensitive to cellularity, edema, and necrosis (Fig. 2.3b).

$$MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{2.3}$$

Axial Diffusivity (AD) corresponds to the major eigenvector. It tends to vary with white matter changes and pathology, such as in axonal injury, where AD decreases. It also has been reported to increase with brain maturation (Fig. 2.3c).

$$AD = \lambda_1 \tag{2.4}$$

Radial Diffusivity (RD) is a measure of diffusivity in the perpendicular plane. It increases in white matter with demyelination. Also, changes in the axonal diameters or density may influence RD values (Fig. 2.3d).

$$RD = \frac{\lambda_2 + \lambda_3}{2} \tag{2.5}$$

Table 2.1 summarizes the relationship between the described scalars and several brain tissue properties, such as: (1) *Gray matter (GM)*, tissue consisting on neuronal cell bodies where the neuron nucleus are located; (2) *White matter (WM)*, areas of the brain that consist of myelinated axons that connect brain cells, and are whitish in color due to the myelin. It is usually distributed into bundles called tracts. (3) *Cerebral Spinal Fluid (CSF)*, a clear fluid that surrounds the brain and spinal cord. It cushions the brain and spinal cord from injury, serving as a nutrient delivery and waste removal system for the brain; (4) *Myelination*, refers to an increase in the fatty sheath surrounding neuronal fibers that increases the efficiency of electrical transmission; (5) *Density of axonal packing*, or the density of WM bundles; (6) *WM maturation*, or how it develops and modifies with human aging; and finally, (7) *Axonal degeneration*, loss of integrity of axonal processes (may occur following injury that directly affects axons, or due to degradation of the myelin sheaths); (8) *Demyelination*, damage to the myelin sheath around nerves, which causes neurological deficits, such as vision changes, weakness, altered sensation, and behavioral or cognitive problems;

Table 2.1 – Summary of the relation between DTI scalars and microstructural and brain tissues. Adapted from Tromp (2013)

	FA	MD	AD	RD
Gray matter	↓		↓	↑
White matter	↑		↑	↓
Cerebralspinal fluid	↓	↑	↑	↑
Myelination	↑	↓		↓
Dense axonal packing	↑	↓		↓
WM Maturation	↑	↓	↑	↓
Axonal degeneration	↓	↑	↓	↑
Demyelination	↓	↑		↑

Additionally to the above mentioned scalar maps, other maps were proposed in the literature, some of them for specific purposes such as segmentation (RITTNER et al., 2013). As the CC is a white matter structure with highly organized fibers connecting the two hemispheres of the brain, it has high FA and the diffusion in its fibers occurs mainly in the right-left direction. The FA map is then combined with the main eigenvector component, e_1 in the right-left direction to highlight the mentioned features and facilitate segmentation of the CC, creating the Weighted FA map (wFA) as described in Eq. 2.6 and depicted in Fig 2.4.

$$wFA = e_{1x} * FA \tag{2.6}$$

(a) Fractional anisotropy map

(b) Fractional anisotropy weighted by the e_{x1} eigenvector component map

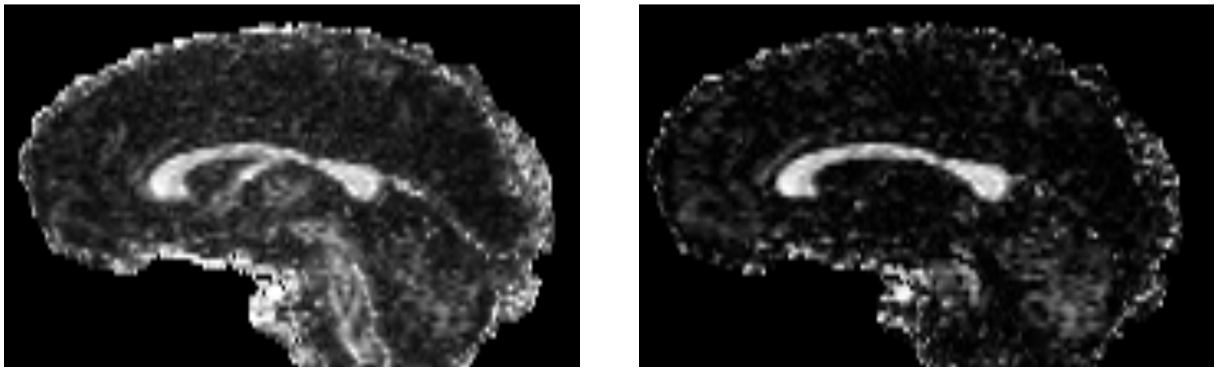


Figure 2.4 – Comparison between Fractional Anisotropy and Weighted Fractional Anisotropy maps.

2.3 CC segmentation

Studies of the CC on DTI usually try to find and relate variations anatomically or in the diffusiveness parameters with a certain condition. Therefore, these studies usually depend on the segmentation of the corpus callosum – determination of the CC borders – and parcellation – division of the structure in different parts, according with the cortical regions with which they are interconnected, but these tasks present difficulties that can jeopardise research quality. This section will explore published methods for segmentation and automatic quality assessment of these segmentations.

2.3.1 Segmentation methods available in the literature

There are multiple algorithms available in the literature for automatic or semi-automatic segmentation that performed quantitative evaluation (COVER et al., 2018), although few of them are DTI-based. Nazem-Zadeh et al. (2012), presented a 3D level-set method to CC segmentation based on a similarity measure used as a speed function. The similarity measure between every voxel on the propagating hypersurface and its neighbors is based on tensor and anisotropy values. The proposed segmentation starts by selecting the initial seed manually, initiating the hypersurface as the zero level-set of the Hamilton–Jacobi partial differential equation, running iteratively the algorithm for propagation of the level-set until convergence, and extracting the zero level-set as the segmented CC.

(a) Weighted fractional anisotropy method

(b) wFA map segmented using the ROQS method

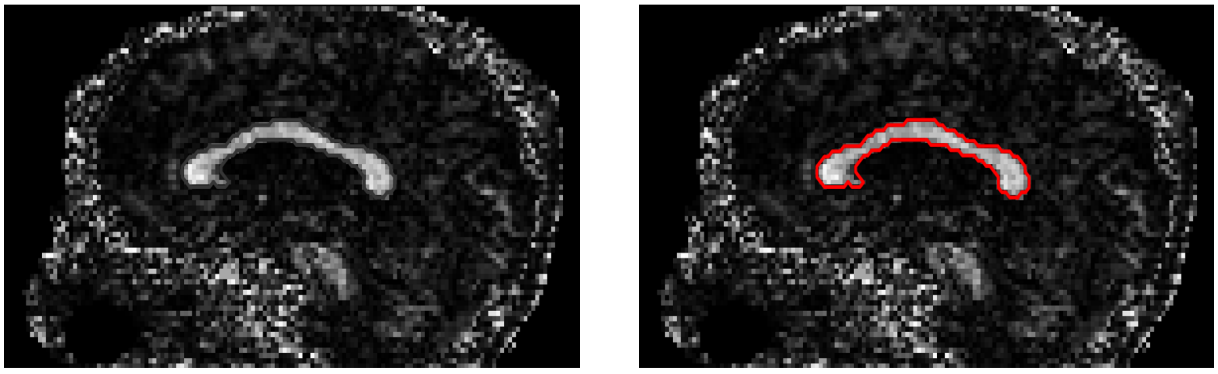


Figure 2.5 – wFA map before and after segmentation using the ROQS method.

Freitas et al. (2011) performed the watershed transform over the external morphological gradient of the fractional anisotropy (FA) map, using automated markers as local minima with highest extinction values. Additionally, the CC 3D extraction is implemented through watershed, using 2D segmentation as internal markers and brain mask as external markers (COVER et al., 2018). Similarly, Rittner et al. (2013) cropped the CC area so that the applied method would concentrate its effort on distinguishing the CC from its background. Then, the watershed transform is implemented over a tensorial morphological gradient map. Thus, when the computation of scalar maps is performed, a hierarchical segmentation is achieved by the application of the 3D watershed through automatically selected markers from the greatest volume extinction.

The threshold-based segmentation is a popular technique that creates a binary image associating pixels below or above a specified threshold. A global thresholding technique verges the entire image with a single value, whereas a local model partitions

a given image into subimages and determines a threshold for each of these partitions. Niogi et al. (2007) implemented a full-threshold CC segmentation by manually placing seeds inside the identified anatomical structure, automatically determining its thresholds through selection criteria based on tensorial seed properties, creating a binary mask, and consequently calculating the boundary structure. The method was named Reproducible Objective Quantification Scheme (ROQS), and one of its segmentations is shown on Fig. 2.5.

Luis-García et al. (2011) used geodesic active regions as a supervised segmentation method, which includes a statistical model for feature extraction, and level-set segmentation for the implementation of the curve evolution. To represent both the tensor and the intensity information, mixtures of Gaussians on tensor fields are used as a statistical model and following an expectation-maximization approach, the level-set evolution and the update of the parameters of the Gaussians mixtures are iterated until convergence in the final segmentation.

2.3.2 Quality assessment

Most CC studies in DTI and all of the studies that propose automatic and semi-automatic algorithms for segmentation cited in Section 2.3, rely on manual quality checking. When existent, the quality assessment usually comprises visual inspection, in an exploratory approach, making it subjective, prone to errors, and time-consuming (REEVES et al., 2016).

The work from Herrera et al. (2019) proposes a framework specifically for assessing the quality of corpus callosum segmentations. The proposed framework is based on a CC shape signature, a descriptor that measures curvature along the segmentation contour. The curvature (k) in one point ($p : x_p, y_p$) of the contour is given by:

$$k(x_p, y_p) = \arctan\left(\frac{y_{p+r} - y_p}{x_{p+r} - x_p}\right) - \arctan\left(\frac{y_p - y_{p+r}}{x_p - x_{p+r}}\right) \tag{2.7}$$

where k represents the angle at the p point and r determines the resolution of the signature. Later the shape signature is evaluated using a support vector machine (SVM) at a given resolution. That way, after being trained with ground truth images curated by experts, the proposed framework allows to evaluate and compare segmentations performed in different

MRI modalities with no manual segmentation nor registration required.

2.4 CC parcellation

In the parcellation task different approaches were proposed to divide the corpus callosum in regions associated particular areas of the cortex. A geometric CC parcellation method was proposed by Witelson (1989), based on postmortem connectivity analyses in primates and humans. Though, geometric approaches only divide the CC structure with the same proportion between all subjects, not considering the individual brain characteristics among different subjects.

Huang et al. (2005) combined the tractography and the cerebral connectivity analysis, using the CC fibers connecting the brain cortex to define CC parcellation regions. Six specific image planes were set as regions of interest (ROI), mapping the fibers that pass through the CC structure, performing the tractography to determine the fibers crossing each ROI and also the CC section. Cook et al. (2005b) proposed a parcellation method with cerebral connectivity in DTI using an atlas segmentation model to label the cortex and find distinct sub-regions in the CC according to their fibers connection.

In the work of Park et al. (2008), the determination of the ROI for tractography was even more specific, by choosing 47 cerebral cortex sub-regions. However, between DTI studies based on tractography, the CC parcellation performed by Hofer & Frahm (2006) was the only work that proposed a scheme for the CC subdivision from an average behavior observed through tractography in a population. The work proposed by Lebel et al. (2010) also used tractography to subdivide the CC into distinct regions, but it is based on target areas for the analysis of aging variation in the cerebral cortex of seven CC regions.

Unlike the CC parcellation methods that use the analysis of the cerebral connectivity by tractography, Rittner et al. (2014) proposed a data-driven method, using the specific information contained in the diffusion images of each subject. The CC parcellation method proposed by Rittner and colleagues was applied in the brain midsagittal plane and performed through the Watershed algorithm, finding five distinct regions. The method was latter improved by Cover et al. (2017) using k-means clustering in order to set the seeds for the Watershed algorithm.

Finally, the work performed by Chao et al. (2009) computes the CC parcellation according to the topography of its structure through the connection of the CC fibers made in a probabilistic way. It was based on an analyzed population, using the high angular resolution diffusion images (HARDI), where the anatomical tractography is estimated and incorporated into the final CC parcellation.

2.5 Related DTI tools

Given the complexity of DTI processing and analysis, several software tools were developed and are widely used by the scientific community. The existing tools are mainly focused on: preprocessing, visualization, tractography and connectivity studies. Just a few allows region based analysis, but often require manual segmentation. None of them incorporated automated DTI-based methods for segmentation and parcellation of the CC. Information about such tools summarized in this section was obtained in the *Neuroimaging Tools and Resources Collaboratory* (NITRC), an award-winning free web-based resource that offers comprehensive information on an ever expanding scope of neuroinformatics software and data [3].

The first group of available tools implement DTI preprocessing, such as artifacts correction and tensor estimation, alongside processing techniques such as connectivity calculations and fiber tracking. Three that are often used in the initial pipeline steps and are well known in the community are: **Camino** (TOUSSAINT et al., 2007), a free, open-source, object-oriented software package for analysis and reconstruction of Diffusion MRI data, tractography and connectivity mapping. The toolkit implements standard techniques, such as diffusion tensor fitting, mapping fractional anisotropy and mean diffusivity, deterministic and probabilistic tractography; **TORTOISE** (PIERPAOLI et al., 2010), a software package for processing DTI data. It contains two main modules, `DIFF_PREP`, software for image resampling, motion, eddy current distortion and susceptibility induced EPI distortion corrections, and for re-orientation of data to a common space, and `DIFF_CALC`, software for tensor fitting, error analysis, color map visualization and ROI analysis; and finally, **FSL** (JENKINSON et al., 2012), a command-line based, comprehensive library of image analysis and statistical tools for FMRI, MRI and DTI brain imaging data.

[3] Available at: https://www.nitrc.org/. Accessed on: 1st march, 2021

Given the complexity and high dimensionality of the data, there are tools that focus on the visualization of DTI data and manipulation of these images, such as: **ENIGMA-Viewer** (ZHANG et al., 2017), an interactive visualization tool for neuroscientists to examine DTI data and compare effective sizes and associated genetics information in meta-analysis; **MRIcron** (RORDEN, 2007), a cross-platform NIfTI (Neuroimaging Informatics Technology Initiative, an open-source neuroimaging format) image viewer that can load multiple layers of images, generate volume renderings and draw volumes of interest; **Papaya** (LANCASTER; MARTINEZ, 2018), a web-based medical research image viewer that supports overlays, atlases, GIFTI (Geometry format under the Neuroimaging Informatics Technology Initiative) surface data and DTI data, and finally; **BrainBrowser** (SHERIF et al., 2015), a web-enabled brain surface viewer that allows the user to explore in real time a 3D brain map expressed on a base surface and that can manipulate 3D fiber pathways derived from DTI.

Some of the software tools focus in multi-modal studies, and allow overlaying and data crossing such as: **3D Slicer** (PIEPER et al., 2004), a software platform for the analysis (including registration and interactive segmentation) and visualization (including volume rendering) of medical images and for research in image guided therapy. It is multi-organ and supports multi-modality imaging including, MRI, tomography, ultrasound, nuclear medicine, and microscopy; **BrainVoyager** (GOEBEL, 2012), a tool for analysis of anatomical, diffusion and functional MRI data sets, it allows analysis of Diffusion-Weighted Imaging including combined visualization of tracked fiber bundles with structural and functional MRI; and **MedINRIA** (TOUSSAINT et al., 2007), which allows to process and analyze a wide range of magnetic resonance (MR) images including anatomical MRI, functional MRI, and diffusion tensor MRI. It performs interactive manual image segmentation, and runs various image processing, segmentation and registration algorithms.

Another approach to DTI analysis is grouped brain normalization and the creation of atlases, which tries to mitigate anatomical differences in group studies. In this context, **DTI-TK** (ZHANG et al., 2007) is a spatial normalization & atlas construction toolkit, designed from ground up to support the manipulation of DTI. It allows resampling, smoothing, warping, registration and visualization, alongside spatial normalization and atlas construction for population-based studies. Similarly, but more general regarding MRI types, **Freesurfer** (FISCHL, 2012) is a set of automated tools for reconstruction of the

brain's cortical surface from structural MRI data, and overlay of functional MRI data onto the reconstructed surface.

Some of the tools available for DTI processing are dedicated to perform tractography. Some that can be cited into this category are the **Diffusion Toolkit (with TrackVis)** (WANG et al., 2007), a cross-platform software package that does reconstruction, fiber tracking, visualization and analysis on various diffusion imaging data; **MRtrix3** (TOURNIER et al., 2019), which provides a set of tools to perform various types of diffusion MRI analyses, from various forms of tractography through to next-generation group-level analyses; and **MultiXplore** (BAKHSHMAND et al., 2017) is a graphical user interface that serves to display corresponding set of cortical regions from functional connectivity matrix in an explorable 3D scene that represents brain anatomical environment. In addition to grey matter regions, MultiXplore automatically finds and extracts deterministic fiber bundles which exist between selected region(s) and adds them to the 3D environment.

Finally, some software comprehend several steps of the DTI processing pipeline and implement some of the techniques cited before, such as **MRI Studio** (JIANG et al., 2006), former DTI Studio, is an image processing program suitable for such tasks as tensor calculation, color mapping, fiber tracking, image registration and 3D visualization. It can also be used for Region of Interest drawing and statistics; **BioImage Suite** (PAPADEMETRIS et al., 2006) is a web-based medical image analysis software package with image processing, image registration and visualization capabilities. Overlay images to create image visualizations, connectivity visualizations and interactive manual image segmentation; **CMTK** (ROHLFING, 2011) is a software toolkit for computational morphometry of biomedical images and comprises a set of command line tools and a back-end general-purpose library for processing and I/O. It allows registration, image correction, processing (filters; combination of segmentations via voting and via Simultaneous Truth and Performance Level Estimation - STAPLE; shape-based averaging), statistics (t-tests; general linear model) and **DiPy** (GARYFALLIDIS et al., 2014), a package for DTI analysis in Python with tools for tracking, clustering, visualization, and statistical analysis of MRI data.

3 inCCsight: the developed web-application

This chapter will describe the application implemented. First the requirements considered for the development are presented (Sec. **??**). Then the software is described (Sec. 3.1): all scripts implemented and incorporated in the application for segmentation and parcellation are introduced; and, following, decisions regarding layout and all visual components of the application are introduced and discussed.

The review on related DTI tools (Section 2.5) evidences a long list of well implemented applications for DTI processing that, although having the same purpose, use different scripts and techniques and have expressively different ways to manipulate data. Furthermore, the extensive list of segmentation and parcellation methods and approaches builds a context where DTI studies might be performed in uncountable ways, which are rarely well explained or have their scripts available.

An extensive, although unstructured, research was done on recent, well cited papers – e.g the works from Tétreault et al. (2020), Alexander et al. (2007), Genc et al. (2018), Fan et al. (2019), among others – that considered analysis of the corpus callosum with the usual pipeline (Section 1.2.1), in order to understand the type of visualization and statistical analysis their results relied on. Also, to meet the requirements and the demands of the target audience, researchers that work with diffusion MRI data of the Corpus Callosum were consulted and the tools and literature cited in Section 2.5 were considered.

As a result, the challenges researchers face in this niche were identified, and the features and characteristics that would be required of a tool that proposed to mitigate them were defined. Following the nested model for visualization design by Munzner (Munzner, 2009), the problem domain was characterized and the requirements for the software were delimited as follows:.

R1 Extract and calculate diffusion MRI scalar maps;

R2 Perform segmentation with different methods, including a consensus method to improve quality and robustness, and allow comparison;

R3 Perform quality assessment of segmentations in visual and quantitative manner;

R4 Perform parcellation with different methods, and allow comparison;

R5 Perform statistical analysis (grouped mean and standard deviation, Pearson's correlation);

R6 Import external numerical and categorical data about the subjects, such as age or sex;

R7 Visualize and cross-reference processed and imported data in interactive plots;

R8 Implement Input/Output formats compatible with most used software tools: Import data format: NifTi (eigenvalues/eigenvectors) and Exported data: tables (.xlsx), images (.png), interactive plots (.html).

3.1 Software description

In this section, the features and characteristics of inCCsight are presented, also describing how it works, its inputs and outputs. This software was developed to be simple, intuitive, open-source, and portable, so it was developed as a web-based application, and can run on the browser of the user's choice. It is avaliable on github at <https://github.com/thaiscaldeira/inCCsight> and can be installed and built on Linux from source and run on Windows or Mac using a Docker image (ANDERSON, 2015). Instructions on installation and use can be found on the Appendix A.1.

3.1.1 Data and configuration input

Following DTI preprocessing software research, the user must input the eigenvectors and eigenvalues, and with them, the software generates the diffusion scalar maps (FA, MD, RD, AD) using in-house scripts developed with the NiPy (GORGOLEWSKI et al., 2011) package. The eigenvectors and values are input in the FSL format, which separates each vector and value in a Nifti file. FSL is frequently used for DTI preprocessing, and other available DTI software either use their file format or allow conversion to it (e.g CAMINO (COOK et al., 2005a), TORTOISE (PIERPAOLI et al., 2010)).

The software also accepts input of segmentation masks in Nifti (Neuroimaging Informatics Technology Initiative, an open-source neuroimaging format – `.nii`) format, allowing quality assessment and exploration of external algorithms or manual segmentations.

The user must indicate a string contained in the masks' name to be input, and the program will search for `.nii` (and compacted `.nii.gz`) files in the selected folders. In this case, the program will input the eigenvectors, eigenvalues, and the mask, all of which must be aligned.

At the moment, data is inputted via Terminal/Command Prompt, by indicating the individual subjects folders. This might be done folder by folder or by passing the parent folder of a group of subjects. If more than one parent folder is provided, the software allows group analysis considering the folder tree organization. Similarly, by using flags the user can select which segmentation methods will be applied, although there is a default mode corresponding to both available methods.

After the segmentation process is done one time, the program saves the mask and the extracted information so that it can be retrieved in a following execution. That way, the processing time (of about 3s for each segmentation) and the computational cost are only required at the first time execution.

3.1.2 Segmentation

Two fully-automatic DTI-based CC segmentation methods were implemented: an adaptation of the ROQS algorithm and the Watershed segmentation.

Watershed Rittner et al. (2014) segmented the CC in 2D and 3D using watershed transform implemented over a tensorial morphological gradient map. A hierarchical segmentation is achieved by the application of the 3D watershed through automatically selected markers from the greatest volume extinction;

ROQS Niogi et al. (2007) implemented a full-threshold CC segmentation by manually placing seeds inside the identified anatomical structure, automatically determining its thresholds through selection criteria based on tensorial seed properties, creating a binary mask, and consequently calculating the boundary structure;

STAPLE Warfield et al. (2004) proposed an algorithm that computes probabilistic estimate of the true segmentation from other segmentations employing an expectation-maximization approach.

These were the methods chosen and integrated due to the difficulty of finding publicly available implementations of existing segmentation methods. The Watershed-based method was integrated because it was proposed within the author's research group the original implementation was available.

The ROQS algorithm was the only one found in the literature with a description detailed enough to be implemented. An adaptation was, however, necessary as the method is originally semi-automatic and depends on the manual selection of a seed point inside the structure. Considering that the CC usually presents highest FA values, the need for manual seed placement was eliminated by placing the seed in the voxel with highest FA in the of the midsagittal slice. For robustness the neck region was filtered off before the seed placing.

Considering that not many methods were available, the software was designed in a modular way so that other methods might be added in the future. The program also allows the user to input its own segmentation masks and perform all following steps, including quality assessment of segmentations, parcellation and statistical analysis. Additionally, using the developed software, the user can apply the STAPLE (WARFIELD et al., 2004) algorithm even with the imported masks, combining available segmentations in a consensus for more confidence. The STAPLE result can be then used as starting point for the DTI analysis.

3.1.3 Quality assessment

Considering the pipeline steps comprised by the scope of the software (from scalar map calculation to statistical analysis), the most prone to errors is the segmentation step. The errors in the parcellation step are mostly – if not exclusively – due to segmentation errors, considering the geometrical nature of the algorithm.

With that in mind, the quality assessment of segmentations framework proposed by Herrera et al. (2020) was implemented, which uses the CC shape signature built by measuring the curvature along the CC contour (HERRERA et al., 2019) and an ensamble of trained support vector machines as a classifier. In addition, the average FA value for the whole CC is also considered and outliers (defined as samples below the 10th and above the 90th percentile) are also flagged for visual inspection. All possible errors detected by both

criteria are presented to the user in a panel (see Section 3.1.5 for more details), where the images can be checked and, if it was an actual error, be removed from further analysis.

3.1.4 Parcellation

Five different parcellation schemes were implemented, all of them sectioning the corpus callosum in five parts. Four of these schemes – the first, similar to Freesurfer's, that divides in five equal parts. alongside the schemes from (WITELSON, 1989), Hofer & Frahm (HOFER; FRAHM, 2006), Chao (CHAO et al., 2009) and one similar – are of the geometric type, which follows defined proportions with respect to the line connecting the extreme anterior and posterior points of the CC. Following, a description of each method is presented as well as the schemes for the geometric methods (Fig. 3.1).

Witelson (1989) Defined partition proportions are defined from postmortem studies dividing the CC in 7 regions reflecting connections to determined cortical regions. Frequently the first three regions are combined resulting in a 5-region parcellation (Fig. 3.1a).

Hofer & Frahm (2006) By classifying tracts according to their cortical projections using tractography, 5 partition proportions associated with cortical regions were defined (Fig. 3.1b).

Chao et al. (2009) By using high angular resolution diffusion imaging (HARDI) tractography and cytoarchitectonic information, twenty-eight Broadmann's areas on the surface of cortical cortex were traced and used to divide the CC in 5 partition proportions (Fig. 3.1c).

Cover et al. (2017) Considered a data-driven approache. By applying the Watershed transform from seeds determined using k-Means clustering over the segmented corpus callosum FA map, 5 partitions are obtained which reflect the CC internal tract organization.

Freesurfer (FISCHL, 2012) The software divides the Corpus Callosum in five equal parcels regarding the main axis (Fig. 3.1d).

They were chosen considering some aspects: first was reproducibility, as some parcellation techniques do not present clear steps to find the divisions that could be

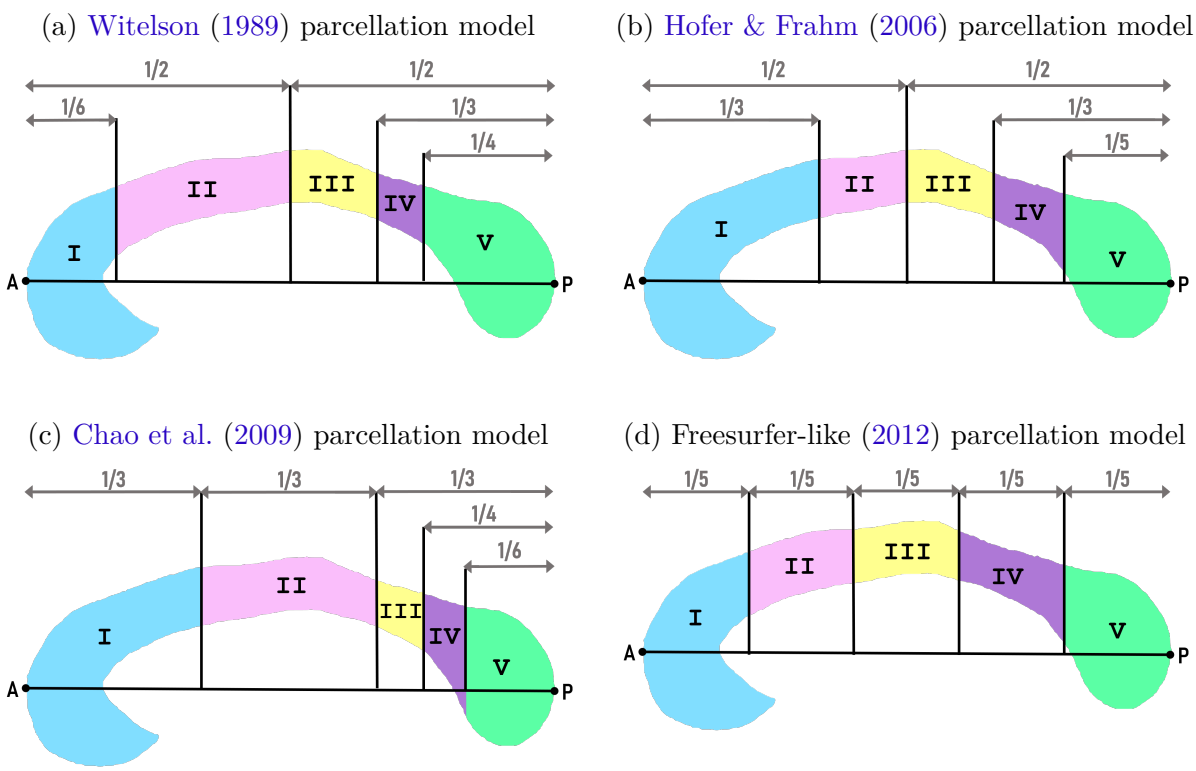


Figure 3.1 – Geometrical parcellation methods available in the literature.

reproduced; furthermore, all elected methods section the corpus callosum in five parts and, therefore, can be compared between themselves, even being based in different study approaches and premises (e.g. *in-vivo* studies, tractography, image processing); finally, those were the most cited at the moment the software started being developed. Similarly as for the segmentation methods, the program is modular regarding the parcellation methods and a new one can be easily integrated.

Considering all implemented parcellation schemes are based on different assumptions concerning the connections between the CC and the cortical areas, thus, choosing the most adequate for a study requires mindful consideration.

3.1.5 Layout and visual components

This section will present different visual components of the developed application and will discuss the prerogatives considered to make certain decisions regarding layout, design and interaction with the user.

Starting with the software layout, presented on Fig. 3.2, the main information

Figure 3.2 – Screenshot of the software inCCsight with highlighted components.

is exhibited in a dashboard-like layout because there are naturally several ways to group and compare the extracted information, as even if the user does not input any extra

data we already have to work with the combination of up to three segmentation and five parcellation methods. Furthermore, the proposed tables and plots are complementary and having them side by side facilitates data exploration and correlation.

(A) Individual subject view

In the list menu (highlighted in red in Fig. 3.2) an specific subject can be selected to check individual data. By clicking in one of the subjects displayed in the list a panel (Fig. 3.3) opens showing information such as the subject's scalar maps, resulting segmentations and tabular data. This feature is specially important after detecting outliers in other plots available, and you might use it to check why an individual is presenting abnormal values of DTI indices.

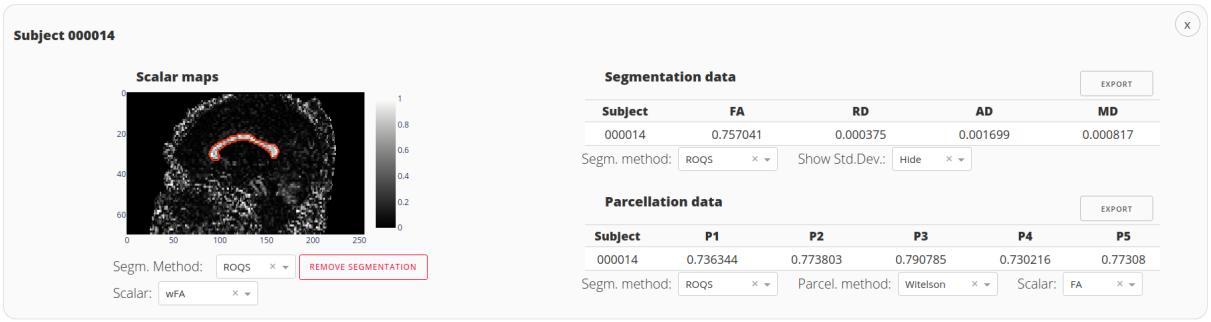


Figure 3.3 – Individual subject panel displaying: segmentation result overlapped to a scalar map (left) and mean values of DTI indices (FA, RD, AD, MD) for the whole CC and for each parcel, in tabular form (right). Users can select the scalar map, the segmentation and the parcellation methods to be shown.

(B) Categories and segmentation method selection

With inCCsight users can import numerical data, such as age, and categorical data, such as sex, using data sheets where each row refers to a single subject. That way, in the Category and segmentation method selection menu (highlighted in yellow in Fig. 3.2), the tables and plots can be stratified by dividing them by one of the native categories (by 'Folder' or by 'Segmentation Method'), or by one of the imported categorical data, such as sex. Furthermore, the imported numerical data can be observed and cross-referenced in the plots available as an extra dimension.

(C) Segmentation quality assessment

To assess if the chosen segmentation algorithm worked two checking procedures were implemented, a statistical one and a feature one. The statistical one detects outliers with regard to the mean FA over the corpus callosum, which is computationally effective, but prone to errors when working with few subjects or conditions that alter the anisotropy of the corpus callosum severely.

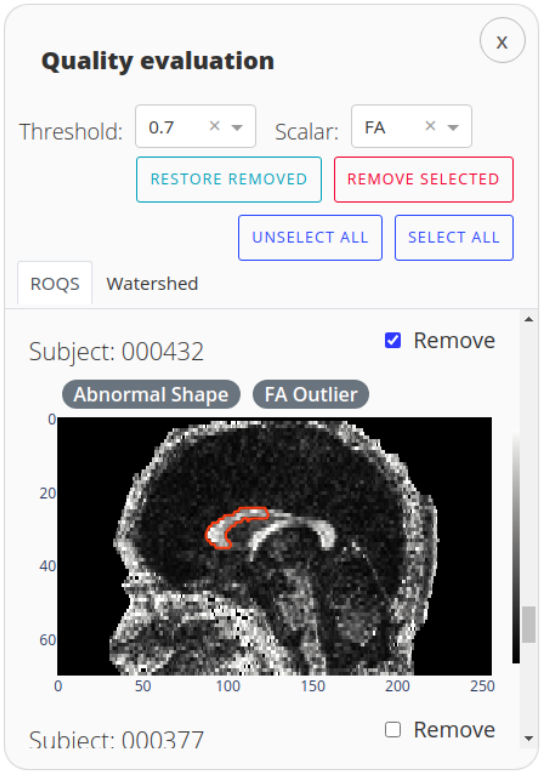


Figure 3.4 – Quality control panel: allows users to check segmentations that were pointed as possible defects according to the quality measurement criteria. Users can perform a visual check for all segmentation methods available, including imported masks.

The feature check is based on the shape signature of the CC (HERRERA et al., 2019), which is extracted for each segmentation applied and classified using a previously tested support vector machine ensamble. All possible errors detected by both checks are exhibited in a panel (Fig. 3.4) that, when opened, is showed side by side with the plots. The user can check if it was an actual error, exclude it from further analysis, or if it was a false positive, and keep it.

(D) Tabular data export

Considering that researchers working with DTI data of the Corpus Callosum are used to deal with tabular data for their analysis, even though a very intuitive platform is

presented for data exploration, it is also organized and exported in the usual tabular format for their appreciation.

A button in the dashboard (highlighted in pink in Fig. 3.2) opens a menu for selecton of which type and parts of data to export in a spreadsheet (.xslx). Users can also include columns with standard deviation, and group the row data averaging by the selected category.

(E) Tabular data view

Numerical values of each subject's segmentation and parcellation or averaged by the chosen category for each scalar obtained are displayed in tables (highlighted in orange in Fig. 3.2). These tables can be manipulated in different ways and their results can be exported and saved.

(F) Interactive plots

Several interactive and customizable plots to visualize and explore the processed data were proposed:

Box Plots For quartile analysis box-plots were implemented for both segmentation and parcellation data, where users can compare groups and discover outliers. Outliers can be pinpointed, showing information about the scalar in question and revealing the subject (Fig. 3.5). The tool allows zooming, panning and saving the plots as images.

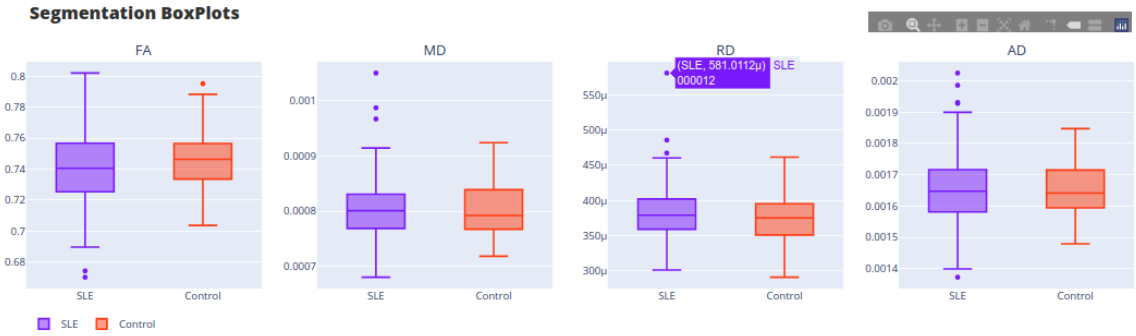


Figure 3.5 – Segmentation box-plot with outlier highlight. The menu in the top right corner allows plot zooming, panning and saving.

Scatter Plots For cross-reference of scalar data a scatter plot was implemented, in which users can select the dimensions to observe and find relations and trends in data (Fig 3.6). Additional histogram and violin plots enrich the analysis.

A Scatter Plot Matrix was also include, which provides a visualization of the relationship between all pairs of DTI indices. This overview might help identify trends in data usually not checked for relationships.

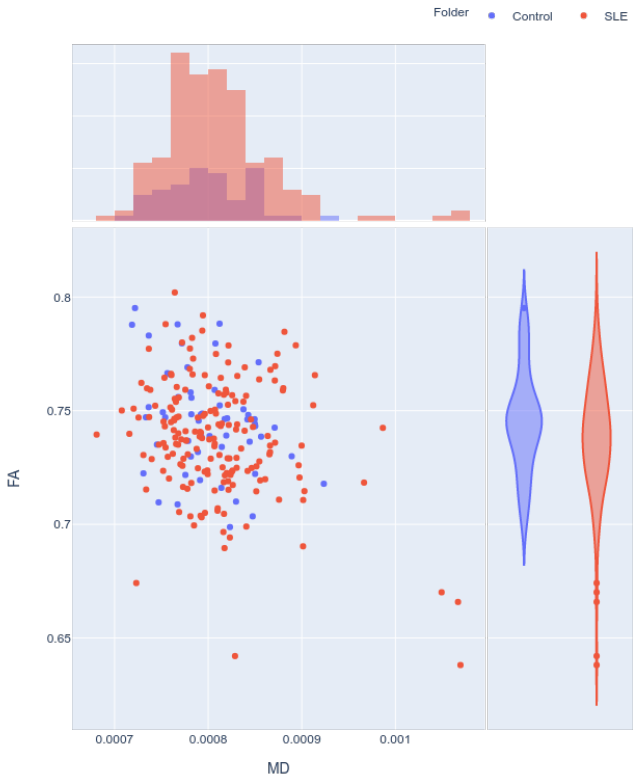


Figure 3.6 – FA x MD scatter plot, with additional histogram and violin plots.

Midline Plots The Midline Plot was implemented for analysis along the Corpus Callosum body, by averaging values for each DTI scalar along the midline of the CC for all subjects in a category (Fig. 3.7). Information about the thickness along the CC body was also included in this plot (Fig. 3.8).

Bubble Plots An alternative to the midline plot consists of an illustrative CC shape composed by bubbles with sizes and colors that vary according with the scalar value around that point for that group (Fig.3.9). Considering that the CC is a curved shape and the midline plot linearizes it, the *bubble plot* might help users interpret this kind of information.

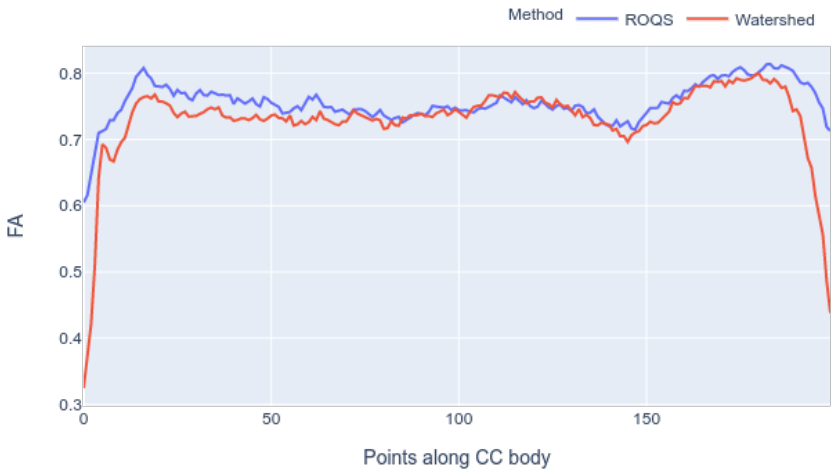


Figure 3.7 – FA along the CC body: mean value from subjects of a chosen category.

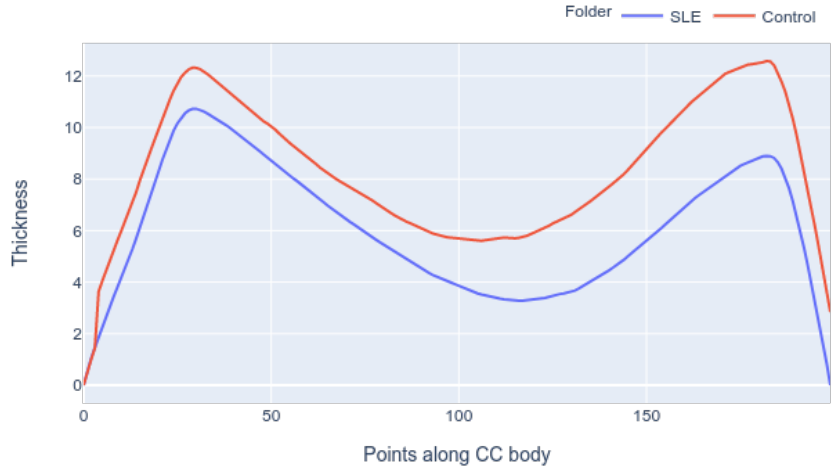


Figure 3.8 – Thickness along the CC body: mean value from subjects of a chosen category.

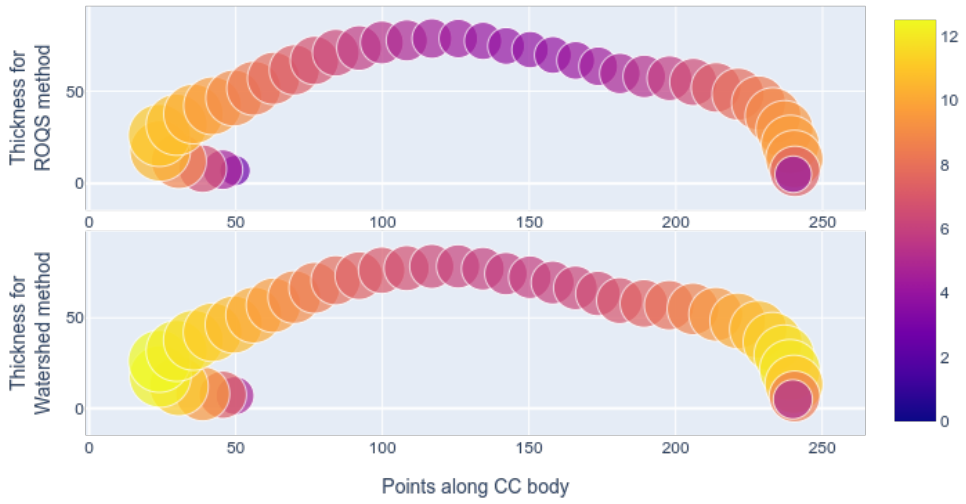


Figure 3.9 – Bubble Plot of thickness alongside the CC structure: size and color of bubbles are proportional to the thickness.

4 CASE STUDY

A case study was conducted in collaboration with domain experts. This is a demonstration of how inCCsight might help researchers find valuable information amidst their data, through data exploration and visualization. The analysis is based on the comparison between a Healthy Control (HC) group and subjects with Systemic Lupus Erythematosus (SLE) using inCCsight, and is focused to aspects where the software features revealed an aspect that might have had passed unnoticed.

The study exemplifies cases that would appear in a neuroscientific environment in this or a similar form and describes how using different automatic methods of segmentation might impact study results. The study compares data from 124 subjects with Systemic Lupus Erythematosus (SLE) and a control group with 55 subjects. Further information about subjects was not used, as the proposed analysis is interested in comparing the segmentation methods and not analysing the results from a medical perspective. Further information about this study were presented by Julio et al. (2020).

After the data analyzed is presented on Sec. 4.1, the following sections are divided to present how the choice of the segmentation method and quality control might impact study conclusions (Sec. 4.2), how the proposed bubble plot can be used to support midline plot analysis, and finally, how the choice of the parcellation .

4.1 Data acquisition and preprocessing

Diffusion Weighted Images (DWI) were acquired in a Philips Achieva 3T scanner with a 1×1 mm spatial resolution and $2mm$ slice thickness in the axial plane, along 32 directions ($b\text{-}value = 1000s/mm^2$) repetition time $8.5s$, time to echo $61ms$), dimensions of $256 \times 256 \times 70$, interpolated by the MRI scanner from a image of dimension $128 \times 128 \times 70$. All data was pre-processed using FSL (JENKINSON et al., 2012), for correction of eddy currents, registration of the DWI volumes, calculation of the Diffusion Tensor Image (DTI) and its eigenvalues and eigenvectors, and conversion to Nifti format. In order to choose the 2D slice in which the CC segmentation would be performed, the callosal fibers convergence plane was computed (PINHEIRO et al., 2018).

4.2 Segmentation methods and quality assessment

The mean values of DTI scalars considering the whole corpus callosum segmentation are presented on Tables 4.1 and 4.2, respectively obtained with the ROQS and Watershed segmentation methods. In the software this values are exhibited in interactive tables. As depicted in the first columns, the values are presented with and without quality assessment of segmentations. Table 4.1, reveals are a good example of how the lack of quality control can mislead us to wrongful conclusions. In this case the $p-value$ without quality assessment is much smaller, indicating statistically relevant differences between the groups, which are less acute after removing wrong segmentations (QC).

Similarly, quality assessment of segmentations has relevant impact on the $p-values$ of Table 4.2, however less important then on Table 4.1. Comparing all values between the two methods some relevant facts can be pointed out: the first is that choice of method leads to higher or lower value change depending on the scalar evaluated. For instance, regarding the mean values, FA is highly affected, while AD is fairly similar. Furthermore, and probably the most important difference between the two tables, is that the $p-values$, before or after quality assessment are lower when using the Watershed method, showing a much more prominent difference between the Control and SLE groups.

Table 4.1 – Grouped DTI scalar values (ROQS)

		FA	RD	AD	MD
Without QC	Control (mean)	0.74536	0.000374	0.001651	0.000799
	SLE (mean)	0.736912	0.000389	0.001662	0.000813
	Control x SLE (p-value)	0.047565	0.029428	0.512746	0.117178
With QC	Control (mean)	0.745572	0.000373	0.00165	0.000799
	SLE (mean)	0.74102	0.000383	0.001665	0.00081
	Control x SLE (p-value)	0.238569	0.102444	0.401954	0.175378

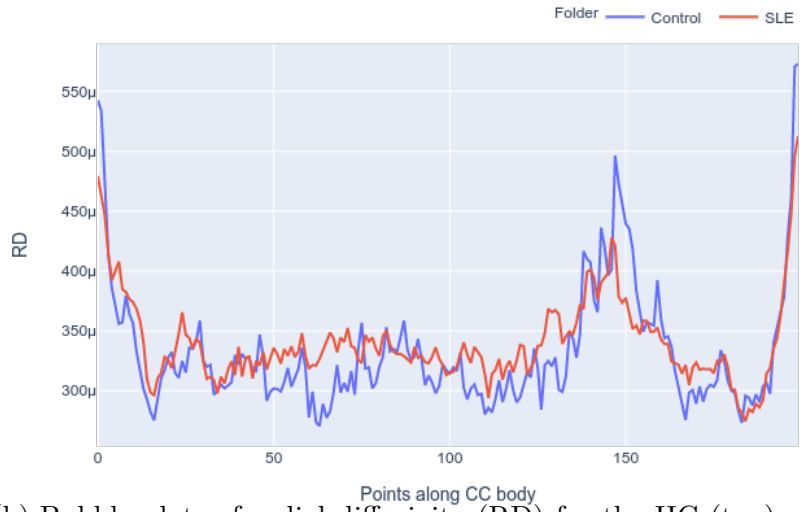
Table 4.2 – Grouped DTI scalar values (Watershed)

		FA	RD	AD	MD
Without QC	Control (mean)	0.649011	0.000551	0.00169	0.000931
	SLE (mean)	0.624276	0.000603	0.001714	0.000974
	Control x SLE (p-value)	0.001888	0.002167	0.268319	0.010496
With QC	Control (mean)	0.649542	0.00055	0.001688	0.000929
	SLE (mean)	0.637114	0.000586	0.001721	0.000964
	Control x SLE (p-value)	0.005726	0.002251	0.092205	0.008975

4.3 Midline vs bubble plot

Although information about the DTI scalars considering the whole CC are valuable, it is richer to evaluate what are the characteristics in details along the structure. The midline plot (Fig. 4.1a) is fairly common in such studies specially because it adds this new dimension and allows researchers to evaluate the inner regions of the CC. It is also useful comparing two groups, as the plotted lines can be easily overlapped.

(a) Midline plot of radial diffusivity (RD) for the HC (blue line) and SLE (red line) groups



(b) Bubble plots of radial diffusivity (RD) for the HC (top) and SLE (bottom) groups

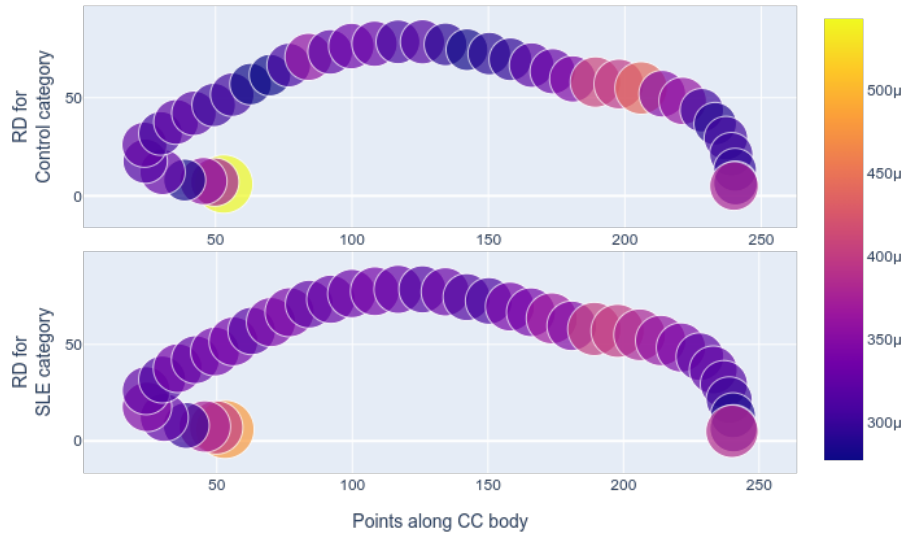


Figure 4.1 – Radial diffusivity plots.

Looking at the midline plot it is possible to evaluate the dimension of the difference between the two groups. It can be, however, misleading for untrained eyes

considering it flattens the information along the x axis, while it is gathered from a curvy structure as the CC is. This means that the plot causes an inherent loss of the notion of which of the regions of the CC each section of the midline plot refers to.

In order to mitigate this, the bubble plot was implemented to carry information about the scalar in question in the size and color of the glyphs, while also keeping the positional information. So, although in the midline plot it is not obvious which is the section of the CC affected, on the bubble plot (Fig.4.1b) this is quite clear. At the same time, in the bubble plot it is hard to find how large the difference is and one can recur to the midline plot for this quantification.

4.4 Parcellation methods

On Table 4.3 the $p - values$ comparing fractional anisotropy between the Control and SLE groups for each of the parcellation methods available (after applying the Watershed segmentation) are displayed. As discussed in the Section 2.4, and illustrated on Fig. 3.1, each of the parcels have different meanings or are connected to different regions of the cortex and, depending on the condition evaluated.

A possible strategy in this case would be choosing the parcellation method which results in the greater difference, and use it to trace back to the cortex region it is related to. The opposite way is also a possibility, and would mean choosing the parcellation method expecting a certain parcel to be affected because of its interconnected cortex region.

Table 4.3 – $p - values$ comparing the fractional anisotropy of the HC and SLE groups in all five parcels (P1 to P5) of the CC

	P1	P2	P3	P4	P5
Witelson	0.031316	0.052423	0.030946	0.042491	0.317473
Hofer & Frahm	0.039968	**0.799161**	**0.637933**	0.059134	0.256467
Chao et al.	0.031316	0.018457	0.064766	0.049969	0.362332
Cover et al.	**0.078624**	0.034163	0.086882	0.022032	**0.667209**
Fressurfer	0.068552	0.076477	0.019277	0.047309	0.184721

5 DISCUSSION

Among all several tools available for DTI exploration, considering those cited in this work, none can be used to perform region of interest (ROI) studies of the corpus callosum in DTI. Works such as those from Alexander et al. (2007), Fan et al. (2019), Tétreault et al. (2020) and Genc et al. (2018) rely on implemented in-house scripts, which cannot be neither easily validated nor reproduced.

In contrast with the available tools, inCCsight implement several of the steps of the described pipeline, eliminating the need to configure, use, and transfer data between different software. By implementing in one tool automated CC segmentation and parcellation, statistical analysis of DTI scalar maps within the CC and its parcels, visualization in interactive plots and exportation of results, the problem of needing several tools to complete DTI-based analysis of the Corpus Callosum is resolved. In fact, it is possible to use only one software (for instance, FSL) for corrections, registration and tensor computation, and then to use inCCsight to perform the rest of the analysis pipeline.

While some tools like DTI-TK and Freesurfer allow segmentation and visualization of the structure, they depend on atlases and segment the whole brain, making the process time and resource expensive. The developed tool, oppositely, focus only in the corpus callosum and takes about a second to perform the segmentation. At the same time, differently from most tools cited that are exclusive for DTI data, inCCsight allows importing external complementary information from the subjects. This is based on the premise that analysis of DTI-only data can become richer in group analysis when cross-referencing with categorical data such as sex, conditions, or numerical information such as age, or the dose of use of certain substance.

Although only the ROQS and Watershed segmentation methods have been implemented, which is justified by the lack of available or reproducible methods in the literature, the software still mitigates the lack of methods with the possibility of importing external segmentation masks. Additionally, the quality control step can be applied to all modalities of segmentations, internal or imported, as well as all following steps. Considering the parcellation methods, the importance of its choice seems to be underestimated in the

literature and, as seen on the Case Study, should not be neglected. With this in mind, the software was developed with features to allow comparison of results when applying different methods, but there is a lack of educational features to actually show the user what each type of parcellation means and how the results could be interpreted when applying one or other.

5.1 Case study

A case study was presented to demonstrate the usefulness of the software in this type of study: considering the first section which discusses the impact of segmentation and quality assessment, it was showed that, specially after the quality control, the *p-value* was significantly different between the two segmentation cases. Regarding only the fractional anisotropy, comparing the two groups while using Watershed lead to a greater statistical relevant difference between the two groups. Similarly, looking at the values for each segmentation method before and after the quality assessment, it is also completely different and would probably change the conclusions one might have taken from this study.

Regarding the use of the bubble plot discussed on Sec. 4.3, it was shown how the correlation between the midline and the bubble plot adds to the interpretation of both: while the midline is better to quantify the difference between the values plotted, the bubble plot is useful to locate where in the corpus callosum those differences are more acute.

Finally, the last section of the case study discussed how the choice of the parcellation method might impact the results of a study. In the presented case, for instance, each parcellation method lead to parcels with completely distinct *p-values* in comparison with the other methods. The software is the first to allow this type of comparison and draw attention of this aspect and its impact in the pipeline.

5.2 User experience

Regarding user experience (UX), the software is very simple to install (or run if the user is using Docker), and although being web-based, runs on the users machine guaranteeing security of data. The dashboard layout has proven to be a good way to cross information

extracted from different plots, but at the moment the layout is fixed, meaning the user cannot reorganize the plots. Also, it would be interesting for users to be able to change plots sizes and positions freely and save a personalized layout, as this would help some specific data exploration line of thought. Additionally, some smaller personalizations such as the color palette and font styling would help those using the program to generate publication quality images.

Considering that the software was developed for medical and health researcher, who might not have a solid background on computational skills, the tool is simple and practical, and at the same time allows the more technical user to personalize the pipeline. Although manual interaction was not completely eliminated with data, as the user has to check the quality control and choose between the segmentation and parcellation methods, the tool educates the user to make this choice consciously by allowing the comparison between different methods. Furthermore, despite having several methods for segmentation and parcellation implemented, the program has a default mode that runs the analysis completely automatic, so that the user does not have to make choices before checking its impact in results. Once the analysis is ready, the interactive aspect of the tool offers the possibility of exploring and comparing the obtained results from the different methods, if desired.

One functionality not available at the moment and that should be prioritized for development is the ability to save the workspace, including the segmentations flagged as errors. At the moment this forces the user to redo the checking work every time the program is executed, which is time expensive and repetitive.

The application handles well larger amounts of data, but there is room for improvement by adopting solutions such as parallel processing and multi-threading. The input of data is another aspect that should be improved by implementing a dedicated user interface and avoiding the use of the Terminal/Command Prompt.

A prototype was developed considering a different structure for the program with a different layout, in a non-web based manner, with multi window and sequential mode. With this organization each step of the pipeline was separated in a different window, and the user would follow the pipeline step by step. The format was reconsidered because some feedback revealed that, while not being able to explore different combinations of

segmentation and parcellation strategies in the results, the user was forced to make those pipeline decisions.

The complementary visualizations were designed not to be displayed apart from the main plots. That way, the program was built in a way that the individual subject information panel and the quality assessment of segmentations panel could also be added to the main dashboard and related to the plots, or, if not necessary, hidden from the main view.

6 CONCLUSIONS

This work presents a new web-based open-source tool, inCCsight, for segmentation, quality control, parcellation, data visualization and exploration of the Corpus Callosum from diffusion MRI data. The motivation given for a new tool is the need for fully automatic methods for the segmentation and parcellation of the corpus callosum on DTI, considering the availability of large amounts of data, while still promoting transparency and reproducibility.

The proposed software, inCCsight, has proven to be a practical and useful tool for DTI-based analysis of the Corpus Callosum. It is the first publicly available tool that provides an automated analysis pipeline comprising: CC segmentation by two distinct methods; automated Quality Control (QC) of segmentation results; CC parcellation by five different methods; and results visualization on a rich, comprehensive dashboard, including several table and plots. The interactive aspect of it was designed to allow a broad spectrum of studies, ranging from single subject analysis to group comparison.

Considering the requirement analysis, the implemented software was able to: deal with input/output formats compatible with most used softwares; compute DTI maps; segment the CC using two different method; parcellate the CC using five different methods; perform quality assessment of the obtained segmentations; perform statistical analysis using, in addition to the DTI maps, imported categorical data; and visualize computed data in interactive plots. Additionally, it is open-source, portable and web-based, conceived to facilitate collaboration and to aggregate other methods to the pipeline, allowing comparison and cross analysis of the different options available.

The dashboard also opens the possibility for more exploratory studies, where patterns and correlations can be found in an interactive manner, expanding the possibilities of the typical DTI-based study, that usually ends up in simple tabular data. By providing a completely automated segmentation step, followed by an automated assessment of its quality results, the tool is an important step towards achieving reproducibility in large dataset studies. With the help of the software, inCCsight, the impact of methods choice for segmentation and parcellation of the CC was used in a study of Systemic Lupus

Erithematosus (SLE) patients. It was concluded that the differences might be relevant for the results of a study and that the developed software has an important role in revealing such differences.

6.1 Future works

The software presented in this dissertation, inCCsight, still has some aspects to be improved. Although the type of CC DTI-based studies discussed in this work is focused on the midsagittal section of the structure, it would be beneficial to integrate 3D automatic segmentation and volumetric visualization as this is a growing line of work. This is specially hard for the corpus callosum, as its fibers are not well delimited far from the midcallosal plane and those methods usually require manual fine tuning —. It would also be good to allow users to perform studies such as tractography and tract based spacial statistic (TBSS) based on the segmentation and parcellation masks obtained from the software.

There are several improvements that could be done regarding the user experience, as pointed on the Discussion (Sec. 5.2), namely: allowing dashboard personalization, including plot reorganization and styling changes such as font and color palette; letting users save their workspace, including the indication of whether a segmentation was marked as erroneous or not; processing and visualization performance could be improved by implementing solutions such as multi-threading or parallel processing; finally, and perhaps the most relevant, data inputting should be made more intuitive, eliminating the need to use the Terminal/Command Prompt by developing a graphical user interface for this.

Regarding the scripts implemented, a logical next step would also be to add other CC segmentation and parcellation algorithms to the software, although their scripts are usually not available and reproducing them is a hard task. Finally, the pipeline implemented and the visualization and data exploration strategy adopted could also be extended to other brain structures.

An important aspect that should be addressed carefully is the effort to inform the user regarding the segmentation and parcellation methods applied and how they work, what are the type of impacts expected from each of them and what are the premises they are based on. This would greatly impact in the objective of allowing users to make educated decisions regarding the pipeline.

6.2 Publications and presentations

These are the publications and presentations related to this work:

Journal paper: inCCsight: A software for exploration and visualization of diffusion MRI data of the Corpus Callosum. *Computers & Graphics, 2021 (second round of review).*

Conference paper: ***Awarded for best oral presentation (Masters category)*** - Automatic corpus callosum segmentation and parcellation in DTI: a tool for visualization and comparison of results from different methods. *7th Brazilian Institute of Neuroscience and Neurotechnology Congress, 2021.*

Registered computer program: inCCsight (process BR512021000281-4). *National Institute of Industrial Property, 2021.*

Technical workshop - live demonstrations: inCCsight. *SPIE Medical Imaging, 2021: Computer-Aided Diagnosis (Conference 11597).*

Conference abstract (co-author): Microstructural damage is associated with age at disease-onset and cognitive impairment in Systemic Lupus Erythematosus (JULIO et al., 2020). *American College of Rheumathology Convergence, 2020.*

6.3 Tool and materials

inCCsight, the tool developed and described in this document, is open-source and available online in GitHub: inCCsight[1] where the user can find an extensive documentation on how to use it. It was built in Python and Javascript using the Plotly/Dash framework. Intructions on how to install and how to use inCCsight can be found on the Appendix.

[1] https://github.com/thaiscaldeira/inCCsight

Bibliography

ABOITIZ, F.; MONTIEL, J. One hundred million years of interhemispheric communication: the history of the Corpus Callosum. *Braz. J. Med. Biol. Res.*, SciELO Brasil, v. 36, n. 4, p. 409–420, 2003. Cited 2 times on pages 14 and 20.

ALEXANDER, A. L. et al. Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage*, Elsevier, v. 34, n. 1, p. 61–73, 2007. Cited 2 times on pages 32 and 48.

ANDERSON, C. Docker. *Ieee Software*, IEEE, v. 32, n. 3, p. 102–c3, 2015. Cited on page 33.

BAKHSHMAND, S. M. et al. Multixplore: Visual exploration platform for multimodal neuroimaging data. *Journal of neuroscience methods*, Elsevier, v. 290, p. 1–12, 2017. Cited 2 times on pages 17 and 31.

BAMMER, R. Basic principles of diffusion-weighted imaging. *European journal of radiology*, Elsevier, v. 45, n. 3, p. 169–184, 2003. Cited on page 21.

BASSER, P. J. et al. In vivo fiber tractography using dt-mri data. *Magnetic resonance in medicine*, Wiley Online Library, v. 44, n. 4, p. 625–632, 2000. Cited on page 14.

BIHAN, D. L. et al. Diffusion tensor imaging: concepts and applications. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, Wiley Online Library, v. 13, n. 4, p. 534–546, 2001. Cited on page 22.

CHAO, Y. et al. Probabilistic topography of human Corpus Calossum using cytoarchitectural parcellation and high angular resolution diffusion imaging tractography. *Human brain mapping*, Wiley Online Library, v. 30, n. 10, p. 3172–3187, 2009. Cited 3 times on pages 29, 36, and 37.

COOK, P. A. et al. Camino: Diffusion mri reconstruction and processing. 2005. Cited on page 33.

COOK, P. A. et al. An automated approach to connectivity-based partitioning of brain structures. In: SPRINGER. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. [S.l.], 2005. p. 164–171. Cited on page 28.

COVER, G. et al. Computational methods for Corpus Calossum segmentation on MRI: A systematic literature review. *Comput.Meth.Prog.Bio.*, Elsevier, v. 154, p. 25–35, 2018. Cited 4 times on pages 15, 17, 20, and 26.

COVER, G. et al. Data-driven Corpus Calossum parcellation method through diffusion tensor imaging. *IEEE Access*, IEEE, v. 5, p. 22421–22432, 2017. Cited 3 times on pages 15, 28, and 36.

DESCOTEAUX, M. Wiley encyclopedia of electrical and electronics engineering. Wiley, 2015. Cited on page 22.

EGAAS, B.; COURCHESNE, E.; SAITOH, O. Reduced size of Corpus Calossum in autism. *Archives of neurology*, American Medical Association, v. 52, n. 8, p. 794–801, 1995. Cited on page 14.

FAN, Q. et al. Age-related alterations in axonal microstructure in the corpus callosum measured by high-gradient diffusion MRI. *Neuroimage*, Elsevier, v. 191, p. 325–336, 2019. Cited 2 times on pages 32 and 48.

FISCHL, B. FreeSurfer. *Neuroimage*, Elsevier, v. 62, n. 2, p. 774–781, 2012. Cited 4 times on pages 17, 30, 36, and 37.

FREITAS, P. et al. Watershed-based segmentation of the midsagittal section of the Corpus Calossum in diffusion MRI. In: IEEE. *2011 24th SIBGRAPI Conference on Graphics, Patterns and Images*. [S.l.], 2011. p. 274–280. Cited on page 26.

GARYFALLIDIS, E. et al. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, Frontiers, v. 8, p. 8, 2014. Cited on page 31.

GENC, S. et al. Age, sex, and puberty related development of the corpus callosum: a multi-technique diffusion MRI study. *Brain Structure and Function*, Springer, v. 223, n. 6, p. 2753–2765, 2018. Cited 2 times on pages 32 and 48.

GOEBEL, R. Brainvoyager—past, present, future. *Neuroimage*, Elsevier, v. 62, n. 2, p. 748–756, 2012. Cited 2 times on pages 17 and 30.

GORGOLEWSKI, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in neuroinformatics*, Frontiers, v. 5, p. 13, 2011. Cited on page 33.

HE, Q. et al. A context-sensitive active contour for 2d Corpus Calossum segmentation. *International journal of biomedical imaging*, Hindawi, v. 2007, 2007. Cited on page 15.

HERRERA, W.; BENTO, M.; RITTNER, L. Corpus Calossum shape signature for segmentation evaluation. In: SPRINGER. *XXVI Brazilian Congress on Biomedical Engineering*. [S.l.], 2019. p. 143–147. Cited 4 times on pages 15, 27, 35, and 40.

HERRERA, W. G. et al. A framework for quality control of corpus callosum segmentation in large-scale studies. *Journal of neuroscience methods*, Elsevier, v. 334, p. 108593, 2020. Cited on page 35.

HOFER, S.; FRAHM, J. Topography of the human Corpus Calossum revisited—comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage*, Elsevier, v. 32, n. 3, p. 989–994, 2006. Cited 4 times on pages 15, 28, 36, and 37.

HUANG, H. et al. DTI tractography based parcellation of white matter: application to the mid-sagittal morphology of Corpus Calossum. *Neuroimage*, Elsevier, v. 26, n. 1, p. 195–205, 2005. Cited 2 times on pages 15 and 28.

HUISMAN, T. A. Diffusion-weighted imaging: basic concepts and application in cerebral stroke and head trauma. *European radiology*, Springer, v. 13, n. 10, p. 2283–2297, 2003. Cited on page 21.

JENKINSON, M. et al. FSL. *Neuroimage*, Elsevier, v. 62, n. 2, p. 782–790, 2012. Cited 3 times on pages 16, 29, and 44.

JIANG, H. et al. Dtistudio: resource program for diffusion tensor computation and fiber bundle tracking. *Computer methods and programs in biomedicine*, Elsevier, v. 81, n. 2, p. 106–116, 2006. Cited on page 31.

JOSHI, S. H. et al. Statistical shape analysis of the Corpus Calossum in schizophrenia. *Neuroimage*, Elsevier, v. 64, p. 547–559, 2013. Cited on page 14.

JULIO, P. et al. Microstructural damage is associated with age at disease-onset and cognitive impairment in systemic lupus erythematosus. In: WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA. *ARTHRITIS & RHEUMATOLOGY*. [S.l.], 2020. v. 72. Cited 2 times on pages 44 and 54.

LANCASTER, J. L.; MARTINEZ, M. J. *Papaya*. 2018. Disponível em: <http://rii-mango.github.io/Papaya/>. Cited on page 30.

LEBEL, C.; CAVERHILL-GODKEWITSCH, S.; BEAULIEU, C. Age-related regional variations of the Corpus Calossum identified by diffusion tensor tractography. *Neuroimage*, Elsevier, v. 52, n. 1, p. 20–31, 2010. Cited on page 28.

LEBEL, C.; DEONI, S. The development of brain white matter microstructure. *Neuroimage*, Elsevier, v. 182, p. 207–218, 2018. Cited on page 14.

LUDERS, E. et al. Positive correlations between Corpus Calossum thickness and intelligence. *Neuroimage*, Elsevier, v. 37, n. 4, p. 1457–1464, 2007. Cited on page 14.

LUIS-GARCÍA, R. de; WESTIN, C.-F.; ALBEROLA-LÓPEZ, C. Gaussian mixtures on tensor fields for segmentation: Applications to medical imaging. *Computerized Medical Imaging and Graphics*, Elsevier, v. 35, n. 1, p. 16–30, 2011. Cited 2 times on pages 17 and 27.

MITCHELL, T. N. et al. Reliable callosal measurement: population normative data confirm sex-related differences. *American Journal of Neuroradiology*, Am Soc Neuroradiology, v. 24, n. 3, p. 410–418, 2003. Cited on page 14.

Munzner, T. A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, v. 15, n. 6, p. 921–928, 2009. Cited on page 32.

NAZEM-ZADEH, M.-R. et al. Segmentation of Corpus Calossum using diffusion tensor imaging: validation in patients with glioblastoma. *BMC medical imaging*, BioMed Central, v. 12, n. 1, p. 10, 2012. Cited 2 times on pages 17 and 26.

NIOGI, S.; MUKHERJEE, P.; MCCANDLISS, B. Diffusion tensor imaging segmentation of white matter structures using a reproducible objective quantification scheme (roqs). *Neuroimage*, Elsevier, v. 35, n. 1, p. 166–174, 2007. Cited 3 times on pages 17, 27, and 34.

O'DWYER, R. et al. Differences in Corpus Calossum volume and diffusivity between temporal and frontal lobe epilepsy. *Epilepsy & Behavior*, Elsevier, v. 19, n. 3, p. 376–382, 2010. Cited on page 14.

PAPADEMETRIS, X. et al. Bioimage suite: An integrated medical image analysis suite: An update. *The insight journal*, NIH Public Access, v. 2006, p. 209, 2006. Cited on page 31.

PARK, H.-J. et al. Corpus callosal connection mapping using cortical gray matter parcellation and dt-MRI. *Human brain mapping*, Wiley Online Library, v. 29, n. 5, p. 503–516, 2008. Cited on page 28.

PEREIRA, M. et al. Corpus callosum parcellation methods: a quantitative comparative study. In: GIMI, B.; KROL, A. (Ed.). *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*. [S.l.]: SPIE, 2018. v. 10578, p. 271 – 276. Cited on page 15.

PIEPER, S.; HALLE, M.; KIKINIS, R. 3d slicer. In: IEEE. *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*. [S.l.], 2004. p. 632–635. Cited 2 times on pages 17 and 30.

PIERPAOLI, C. et al. Tortoise: an integrated software package for processing of diffusion mri data. In: *ISMRM 18th annual meeting*. [S.l.: s.n.], 2010. v. 1597. Cited 2 times on pages 29 and 33.

PINHEIRO, G. et al. Automatic callosal fiber convergence plane computation through DTI-based divergence map. In: *Medical Imaging 2018*. [S.l.: s.n.], 2018. v. 10578, p. 1057815. Cited on page 44.

REEVES, A. P.; LIU, S.; XIE, Y. Image segmentation evaluation for very-large datasets. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Medical Imaging 2016: Computer-Aided Diagnosis*. [S.l.], 2016. v. 9785, p. 97853J. Cited on page 27.

RITTNER, L. et al. Analysis of scalar maps for the segmentation of the Corpus Callosum in diffusion tensor fields. *J. Math. Imaging Vis.*, Springer, v. 45, n. 3, p. 214–226, 2013. Cited 2 times on pages 25 and 26.

RITTNER, L. et al. Automatic DTI-based parcellation of the Corpus Calossum through the watershed transform. *Rev. Bras. Eng. Biom.*, SciELO Brasil, v. 30, n. 2, p. 132–143, 2014. Cited 2 times on pages 28 and 34.

ROHLFING, T. User guide to the computational morphometry toolkit (cmtk). *Neuroscience Program, SRI International, Menlo Park, CA*, v. 3, 2011. Cited on page 31.

RORDEN, C. Mricron. *Retrieved from*, 2007. Cited on page 30.

SHERIF, T. et al. Brainbrowser: distributed, web-based neurological data visualization. *Frontiers in neuroinformatics*, Frontiers, v. 8, p. 89, 2015. Cited on page 30.

TEIPEL, S. J. et al. Progression of Corpus Calossum atrophy in alzheimer disease. *Archives of Neurology*, American Medical Association, v. 59, n. 2, p. 243–248, 2002. Cited on page 14.

TÉTREAULT, P. et al. Diffusion time dependency along the human corpus callosum and exploration of age and sex differences as assessed by oscillating gradient spin-echo diffusion tensor imaging. *NeuroImage*, Elsevier, v. 210, p. 116533, 2020. Cited 2 times on pages 32 and 48.

TOURNIER, J.-D. et al. Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, Elsevier, v. 202, p. 116137, 2019. Cited 2 times on pages 17 and 31.

TOUSSAINT, N.; SOUPLET, J.-C.; FILLARD, P. Medinria: medical image navigation and research tool by inria. In: *Proc. of MICCAI'07 Workshop on Interaction in medical image analysis and visualization*. [S.l.: s.n.], 2007. Cited 3 times on pages 17, 29, and 30.

TROMP, D. *DTI Scalars (FA, MD, AD, RD) - How do they relate to brain structure?* 2013. Disponível em: <http://www.diffusion-imaging.com/2013/01/relation-between-neural-microstructure.html>. Cited 2 times on pages 23 and 24.

WALTERFANG, M. et al. Corpus Calossum size and shape in individuals with current and past depression. *Journal of affective disorders*, Elsevier, v. 115, n. 3, p. 411–420, 2009. Cited on page 14.

WANG, R. et al. Diffusion toolkit: a software package for diffusion imaging data processing and tractography. In: BERLIN. *Proc Intl Soc Mag Reson Med*. [S.l.], 2007. v. 15, n. 3720. Cited 2 times on pages 17 and 31.

WARFIELD, S.; ZOU, K.; WELLS, W. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE T. Med. Imaging*, NIH Public Access, v. 23, n. 7, p. 903, 2004. Cited 2 times on pages 34 and 35.

WILTSHIRE, K. et al. Corpus Calossum in neurodegenerative diseases: findings in parkinson's disease. *Dementia and geriatric cognitive disorders*, Karger Publishers, v. 20, n. 6, p. 345–351, 2005. Cited on page 14.

WITELSON, S. Hand and sex differences in the isthmus and genu of the human Corpus Calossum: a postmortem morphological study. *Brain*, Oxford University Press, v. 112, n. 3, p. 799–835, 1989. Cited 6 times on pages 14, 15, 20, 28, 36, and 37.

ZHANG, G. et al. Enigma-viewer: interactive visualization strategies for conveying effect sizes in meta-analysis. *BMC bioinformatics*, Springer, v. 18, n. 6, p. 17–29, 2017. Cited on page 30.

ZHANG, H. et al. High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: an example study using amyotrophic lateral sclerosis. *IEEE transactions on medical imaging*, IEEE, v. 26, n. 11, p. 1585–1597, 2007. Cited 2 times on pages 17 and 30.

Appendix

APPENDIX A – inCCsight

A.1 How to install

A.1.1 Building from source

Warning for Windows users: We failed to install one of the required libraries (siamxt[1]) on Windows, which chould be built from source. If you're a Windows user we suggest you use the Docker image provided, as explained in the next section.

We suggest you to create a separate virtual environment running Python 3 for this app, and install all of the required dependencies there. The installation steps below include the creation of the virtual environment.

You will need `Python3`, `pip`, `git` and `virtualenv` installed in your machine in order to install this app, please refer to the documentation [2,3] if you need help installing these tools. If you already have `Python3` and `pip` installed, you can install `virtualenv` by typing:

```
pip install virtualenv
```

Run in Terminal. Clone the git repository:

```
git clone https://github.com/thaiscaldeira/ccinsight/
cd ccinsight
```

Create a new virtual environment:

```
python3 -m virtualenv venv
source venv/bin/activate
```

Update pip and install app requirements

[1] https://github.com/rmsouza01/siamxt
[2] https://docs.python.org/3/using/unix.html
[3] https://git-scm.com/book/en/v2/Getting-Started-Installing-Git

```
python -m pip install -U pip
pip install -r requirements.txt
```

Install siamxt[4]:

```
cd siamxt
python setup.py install
```

A.1.2 Using Docker

Docker is an open platform for developing, shipping, and running applications. Docker enables you to separate your applications from your infrastructure so you can deliver software quickly. You can find more information, as well as download links and installation steps here: <https://www.docker.com/>

To use the docker image you'll only have to run it normally with the commands explained in the next sections. The first time you run it, it will automatically download the image.

If you want to download the image without running it do:

```
docker pull thaiscaldeira/inccsight
```

Using docker will require, however, that you use some docker flags. inCCsight runs on localhost port 8000 by default, so if you wish to change its port while using the docker image you can do it by mapping the port like in the example below where it will run on <https://localhost:8888>:

```
docker run -p 8888:8000 thaiscaldeira/inccsight [INCCSIGHT FLAGS]
```

Alongside, to allow the container to read and save information from your disk we have to map the folders we wanna work with. For example, if I have folders organised as showed below we can map the volume `ALL_DATA` to a folder /f/ by using the flag `-v`followed by the mapping configuration `./ALL_DATA/:/f/` and use it with with the `--parent` flag from inCCsight. We will include more examples in the following 'How to use inCCsight' section of how to use it with docker.

[4] https://github.com/rmsouza01/siamxt

```
|- ALL_DATA
   |- HEALTH_CONTROLS
      |- SUBJECT_000001
      |- SUBJECT_000002
      |- SUBJECT_000003
      |- ...
   |- CONDITION_X
      |- SUBJECT_000014
      |- SUBJECT_000015
      |- SUBJECT_000016
      |- ...
   |- CONDITION_Y
      |- SUBJECT_000029
      |- SUBJECT_000030
      |- SUBJECT_000031
      |- ...
```

```
docker run -v ./ALL_DATA:/f/ thaiscaldeira/inccsight --parent
    /f/HEALTH_CONTROLS/ /f/CONDITION_X/ /f/CONDITION_Y/
```

A.2 How to use inCCsight

inCCsight processes DTI data, segmenting the Corpus Callosum and parcellating it using most proeminent segmentation an parcellation techniques available and reproducible in literature. More information about the techniques available please refer to the Segmentation and Parcellation documents.

A.2.1 Input data

We input data by importing DTI eigenvectors and eigenvalues in the FSL format as outputted by FSL DTIFit [5], where you can inform the used basename by using the -b flag

[5] <https://users.fmrib.ox.ac.uk/~behrens/fdt_docs/fdt_dtifit.html>

(default is dti):

- `basename_V1.nii.gz` : 1st eigenvector
- `basename_V2.nii.gz` : 2nd eigenvector
- `basename_V3.nii.gz` : 3rd eigenvector
- `basename_L1.nii.gz` : 1st eigenvalue
- `basename_L2.nii.gz` : 2nd eigenvalue
- `basename_L3.nii.gz` : 3rd eigenvalue

For each subject there must be a folder with the indicated files. The name of each folder will be used in the program as reference to the subject analysed.

You can indicate the folders one by one (subject by subject) or the parent(s) folders that contain folders for each subject in a group.

For instance, using the flag `-f` you can indicate each subject folder, subch as `SUBJ_00001` and `SUBJ_00002` in the example below:

```
|- SUBJECT_000001
   |- dti_V1.nii.gz
   |- dti_V2.nii.gz
   |- dti_V3.nii.gz
   |- dti_L1.nii.gz
   |- dti_L2.nii.gz
   |- dti_L3.nii.gz
|- SUBJECT_000002
   |- dti_V1.nii.gz
   |- dti_V2.nii.gz
   |- dti_V3.nii.gz
   |- dti_L1.nii.gz
   |- dti_L2.nii.gz
   |- dti_L3.nii.gz
```

When working with several subjects and/or different categories of individual (such as *Health Control* x *Condition* or *Male* x *Female*) it is more practical to indicate the

parent folders of each group, such as `HEALTH_CONTROLS`, `CONDITION_X` and `CONDITION_Y` in the example below, where all `SUBJECT` folders contain the eigenvectors and eigenvalue files discussed previously:

```
|- HEALTH_CONTROLS
   |- SUBJECT_000001
   |- SUBJECT_000002
   |- SUBJECT_000003
   |- ...
|- CONDITION_X
   |- SUBJECT_000014
   |- SUBJECT_000015
   |- SUBJECT_000016
   |- ...
|- CONDITION_Y
   |- SUBJECT_000029
   |- SUBJECT_000030
   |- SUBJECT_000031
   |- ...
```

A.2.2 Run commands

To use inCCsight we simply have to call the path to `app.py` on the Terminal/Command Prompt, indicating the subject(s) folder(s) paths (using the flag `--folders` or `-f`) or parent folder(s) paths (using the flag `--parents` or `-p`) to be analysed, such as shown below.

```
python app.py -f ./SUBJECT_000001 ./SUBJECT_000002 ./SUBJECT_000034 ...
```

```
python app.py -p ./HEALTH_CONTROLS ./CONDITION_X ./CONDITION_Y ...
```

We can also indicate auxiliary flags:

- `-b`, or `--basename` : string indicating the basename used in the eigenvectors/eigenvalues files (default is `'dti'`). See the Input Data section for more information;

- `-s`, or `--segm` : segmentation methods to be performed on the data (default is all available: *ROQS* and *Watershed*); *
- `--staple` : if used will create a consensus using the STAPLE method between the segmentations available (including imported masks);
- `-d`, or `--extra-data` : path to sheet file (.xls, .xlsm or .csv) with additional information to be imported and visualized.
- `-m`, or `--maskname` : string contained in the file name of imported masks. See the Importing Masks section for more information;

After calling the initial command, data will be processed and your default browser will open, showing the interactive dashboard for data exploration and visualization. While the Terminal or Command Prompt where the program is running is kept open, you can access the dashboard on <http://127.0.0.1:5050/>.

A.2.3 Importing external data

We often would like to cross categorical data (such as sex or race), or numerical data (such as age), with information extracted through DTI processing. Using **inCCsight** we can import such types of data and visualize relations between them and the processed data by importing a sheet file, such as .xlsm, .xls or .csv, that must have a column called 'Subjects' with the names of the imported Subject folders.

Columns with categorical data will be listed as a View Category that, when selected, will make all graphs compare the groups in this category. For example, if you select the category 'Sex' and your columns divided subjects between 'M' and 'F', all graphs will compare these two groups.

To import external data you can use the flag `-d` or `--extra-data`, as shown:

```
python app.py -p ./HEALTH_CONTROLS ./CONDITION_Y -d
    ./subjects_informations.xls
```