UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA APLICADA À MEDICINA E BIOLOGIA

HOHANA GABRIELA KONELL

Avaliação da U-Net na segmentação de tratos de curta extensão:
transferência para rotina clínica de imagens em ressonância magnética

Assessment of U-Net in the segmentation of short tracts:
transferring to clinical MRI routine

RIBEIRÃO PRETO
2023

HOHANA GABRIELA KONELL

Avaliação da U-Net na segmentação de tratos de curta extensão:
transferência para rotina clínica de imagens em ressonância magnética

Assessment of U-Net in the segmentation of short tracts:
transferring to clinical MRI routine

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para obtenção do título de Mestre em Ciências.

Área de Concentração: Física Aplicada à Medicina e Biologia

Orientador: Prof. Dr. Carlos E. G. Salmon

RIBEIRÃO PRETO – SP

2023

Catalogação da Publicação
Serviço de Documentação da Universidade de São Paulo
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto

Konell, Hohana Gabriela

Avaliação da U-Net na segmentação de tratos de curta extensão: transferência para rotina clínica de imagens em ressonância magnética/ Hohana Gabriela Konell; orientador Carlos Ernesto Garrido Salmon. Ribeirão Preto, SP, 2023.

94 p.: il.

Dissertação (Mestre em Ciências) – Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2023.

Orientador: Prof. Dr. Carlos Ernesto Garrido Salmon

1. Segmentação da Substância Branca, 2. Tractografia, 3. Aprendizado Profundo, 4. Imagens ponderadas em difusão, 5. Tratos de Curta Extensão

Aos meus pais, Rosane e Valdemir, minha irmã Eloyse e minha sobrinha Lara.

AGRADECIMENTOS

Meus profundos agradecimentos a todos que tornaram a conclusão dessa dissertação possível.

Primeiramente, gostaria de agradecer ao meu orientador, Prof. Carlos Garrido, por todos os anos de aprendizado e por toda paciência para me ensinar a crescer, tanto profissionalmente quanto pessoalmente. Sempre me disseram que durante a nossa vida muitas pessoas passam, mas poucas marcam e realmente fazem a diferença nas nossas vidas, você é uma dessas poucas pessoas na minha.

Também quero agradecer a Prof. Renata Leoni, você sempre esteve presente em todos os momentos, seja para conversar e rir no laboratório ou para aconselhar nas mais diversas áreas. Obrigada por tornar o meu caminho na pesquisa e os dias no laboratório sempre mais doces. Meus agradecimentos também ao Prof. Carlinhos, ao Prof. Murta e ao técnico Adriano Holanda que estiveram presentes e tornaram essa pesquisa possível, seja com acesso as imagens do HC de Ribeirão Preto ou ao servidor para treinar minhas redes neurais. Sem vocês minhas redes ainda estariam treinando.

Aos meus queridos colegas de laboratório que me fizeram companhia e ouviram todas as minhas apresentações nas reuniões. Obrigada Eduardo, Tamires, Erick, Fábio e Henrique pelas nossas aventuras nos congressos, eventos e workshops. Meus agradecimentos especiais a Maíra e ao Bruno, por todo conhecimento passado, sem vocês eu não teria feito a minha primeira tractografia e nem minha primeira rede neural. Obrigada por sempre responderem todas as minhas perguntas e me darem suporte mesmo estando do outro lado do oceano.

Quero agradecer também aos meus amados amigos, que estiveram ao meu lado em todos os momentos, tanto nos bons quanto nos não tão bons assim. Obrigada por me escutarem e estarem sempre ao meu lado me incentivando. República Vai da Nada e Gruta, obrigada por me darem uma casa para nunca me sentir sozinha em Ribeirão. Obrigada aos apartamentos 34 e 13, que estiveram ao meu lado nesses dois anos que se passaram e tornaram tudo mais divertido. Meus agradecimentos especiais ao meu casal Mister e Subway, e também a Sheilinha e a Ace, nossas noites de fofoca no corredor serão eternas. Mister, você me ensinou que literalmente tudo pode virar uma piada, e as vezes é exatamente isso que a gente precisa para sorrir todos os dias. E sou eternamente grata pelos meus amigos de longa data, Zevita e Furão, que nunca deixaram os quilômetros fazerem diferença na nossa amizade.

Por último, a minha família. Palavras nunca serão o suficiente para expressar o quanto sou grata a vocês. Vocês sempre me apoiaram em todos os meus sonhos, e apesar dos momentos de dificuldade, nunca duvidaram de mim. Sem vocês, mãe e pai, Rosane e Valdemir, nada disso teria se tornado realidade, pois sei que vocês enfrentaram tudo para que eu chegasse até aqui. Acho que a palavra que define isso realmente é amor. E sem o teu exemplo todos os dias, Dudi, eu não seria quem sou hoje. Você sempre soube o que me falar mesmo quando eu não conseguia expor o que estava sentindo em palavras. Menção honrosa para o meu cunhado, Israel, que sempre trouxe conversas que me impulsionaram para frente e me traziam diferentes perspectivas, e que junto com minha irmã me deram a honra de poder ser tia, e me tornar oficialmente a Tia Lilo. Lara, minha criaturinha, você me mostra todos os dias que não existe melhor físico que uma criança, obrigada por me mostrar um amor que eu não sabia que existia.

Obrigada aos membros da banca por aceitarem o convite e contribuírem para a conclusão deste mestrado.

A realização desse trabalho foi possível por meio do financiamento da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES; 88887.621189/2021-00).

“New directions in science are launched by new tools much more often than by new concepts. The effect of concept-driven revolution is to explain old things in new ways. The effect of tool-driven revolution is to discover new things that have to be explained.”

- Freeman Dysson

RESUMO

KONELL, H.G. **Avaliação da U-Net na Segmentação de Tratos de Curta Extensão: Transferência para Rotina Clínica de Imagens em Ressonância Magnética.** 2023. Dissertação (Mestrado em Física Aplicada à Medicina e Biologia) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto – SP, 2023.

Estudos em conectividade estrutural cerebral requerem estratégias de segmentação de tratos precisas. A rede neural U-Net é altamente reconhecida pela sua capacidade em tarefas de segmentação de imagens, em especial no delineamento de tratos de longa extensão utilizando dados de alta qualidade de Imagens Ponderadas em Difusão (*DWI*). Contudo, tratos de curta extensão, associados a diversas doenças neurológicas, colocam desafios específicos à essas redes, especialmente com aquisições de dados em ambientes clínicos.

O objetivo deste trabalho foi avaliar a capacidade da rede U-Net na segmentação de tratos de curta extensão utilizando dados de *DWI* adquiridos em diferentes condições experimentais. Para isso, foram conduzidos três treinamentos diferentes com um total de 350 indivíduos saudáveis e 11 tratos da substância branca, incluindo comissura anterior, posterior e hipocampal, fórnix e fascículo uncinado. No primeiro experimento, o modelo foi treinado exclusivamente com dados de alta-qualidade do Projeto de Conectoma Humano (*HCP*). O segundo experimento foi focado em imagens de indivíduos saudáveis adquiridos em um hospital local, representando uma típica aquisição de rotina clínica. No último experimento, uma abordagem híbrida foi empregada, combinando imagens de ambos os conjuntos de dados. Por fim, o melhor modelo foi testado em 10 pacientes com epilepsia do hospital local e em 10 indivíduos saudáveis adquiridos em um scanner de uma empresa diferente.

Os resultados do terceiro experimento mostraram notável aumento na performance do modelo em comparação com os outros experimentos. Especificamente, os tratos curtos do conjunto de dados do hospital local alcançaram pontuações Dice entre 0.60 e 0.75. Intervalos similares foram obtidos para os dados do HCP no primeiro experimento, e um aumento substancial para os dados do hospital local que nesse experimento apresentaram pontuações entre 0.37 e 0.50. Esse progresso se manteve mesmo aplicando o método em dados de pacientes com epilepsia e em scanners com diferentes aquisições.

Esses resultados indicam que utilizar conjuntos de dados de diferentes fontes, junto com uma padronização das imagens, aumenta significativamente a capacidade de generalização da rede neural. É importante, contudo, reconhecer que essa performance está intrinsicamente ligada à composição dos conjuntos de dados de treinamento, validação e teste. Além de que, tratos pequenos e com maior nível de curvatura adicionam maior complexidade devido sua estrutura particular. Apesar dos resultados promissores, é necessário ter precaução ao extrapolar essa aplicação a dados adquiridos em circunstâncias distintas, seja em dados de maior qualidade ou analisando tratos de curta ou longa extensão.

Palavras-chave: 1. Segmentação da Substância Branca, 2. Tractografia, 3. Aprendizado Profundo, 4. Imagens ponderadas em difusão, 5. Tratos de Curta Extensão

ABSTRACT

KONELL, H. G. **Assessment of U-Net in the Segmentation of Short Tracts: Transferring to clinical MRI routine.** 2023. Dissertation (M. Sc. In Physics Applied to Medicine and Biology) – School of Philosophy, Scienced, and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto – SP, 2023.

Accurately studying structural connectivity requires precise tract segmentation strategies. The U-Net network has been widely recognized for its exceptional capacity in image segmentation tasks. It has demonstrated remarkable results in segmenting large tracts using high-quality diffusion-weighted imaging (DWI) data. However, short tracts, which are associated with various neurological diseases, pose specific challenges, particularly when considering the DWI data acquisition within clinical settings.

The objective of this work was to evaluate the capability of the U-Net network in segmenting short tracts using DWI data acquired in different experimental conditions. To accomplish this, we conducted three different types of training experiments with a total of 350 healthy subjects and 11 white matter tracts, including anterior, posterior, and hippocampal commissure, fornix, and uncinated fasciculus. In the first experiment, the model was exclusively trained using high-quality data from the Human Connectome Project (HCP) dataset. The second experiment focused on images of healthy subjects acquired from a local hospital dataset, representing a typical clinical routine acquisition. In the third experiment, a hybrid training approach was employed, combining images from the HCP and local hospital datasets. Finally, the best model was also tested in unseen DWIs of 10 epilepsy patients of the local hospital and 10 subjects acquired on a scanner from another company.

The outcomes of the third experiment demonstrated a notable enhancement in performance when contrasted with the preceding trials. Specifically, the short tracts within the local hospital dataset achieved dice scores ranging between 0.60 and 0.75. Similar intervals were obtained with HCP data in the first experiment and a substantial improvement compared to the scores of 0.37 and 0.50 obtained with the local hospital dataset at the same experiment. This improvement persisted when the method was applied to diverse scenarios, including different scanner acquisitions and epilepsy patients.

This outcome strongly indicates that the fusion of datasets from various sources, coupled with resolution standardization, significantly fortifies the neural network's capacity to generalize predictions across a spectrum of datasets. It's crucial, however, to recognize that the performance of short tract segmentation is intricately linked to the composition of the training, validation, and testing data. Moreover, the segmentation of shorter and intricately curved tracts introduces added complexities due to their intricate structural nature. Although this approach has shown promising results, caution is essential when extrapolating its application to datasets acquired under distinct experimental conditions, even when dealing with higher-quality data or analyzing long or short tracts.

Keywords: 1. White Matter Segmentation, 2. Tractography, 3. Deep Learning, 4. Diffusion Weighted Images, 5. Short Tracts

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC Anterior Commissure

AI Artificial Intelligence

ANN Artificial Neural Networks

CNN Convolutional Neural Networks

CSD Constrained Spherical Deconvolution

CSF Cerebrospinal Fluid

CST Corticospinal Tract

DL Deep Learning

DT Diffusion Tensor

DTI Diffusion Tensor Image

DWI Diffusion Weighted Images

FA Fractional Anisotropy

FOD Fiber Orientation Distribution

FX Fornix

GM Gray Matter

HC Hippocampal Commissure

HCP Human Connectome Project

IFOF Inferior Fronto-Occipital Fasciculus

ML Machine Learning

MRI Magnetic Resonance Imaging

NMR Nuclear Magnetic Resonance

PC Posterior Commissure

ROI(s) Region(s) of Interest

SNR Signal-to-Noise Ratio

UF Uncinate Fasciculus

WM White Matter

SUMMARY

1 INTRODUCTION

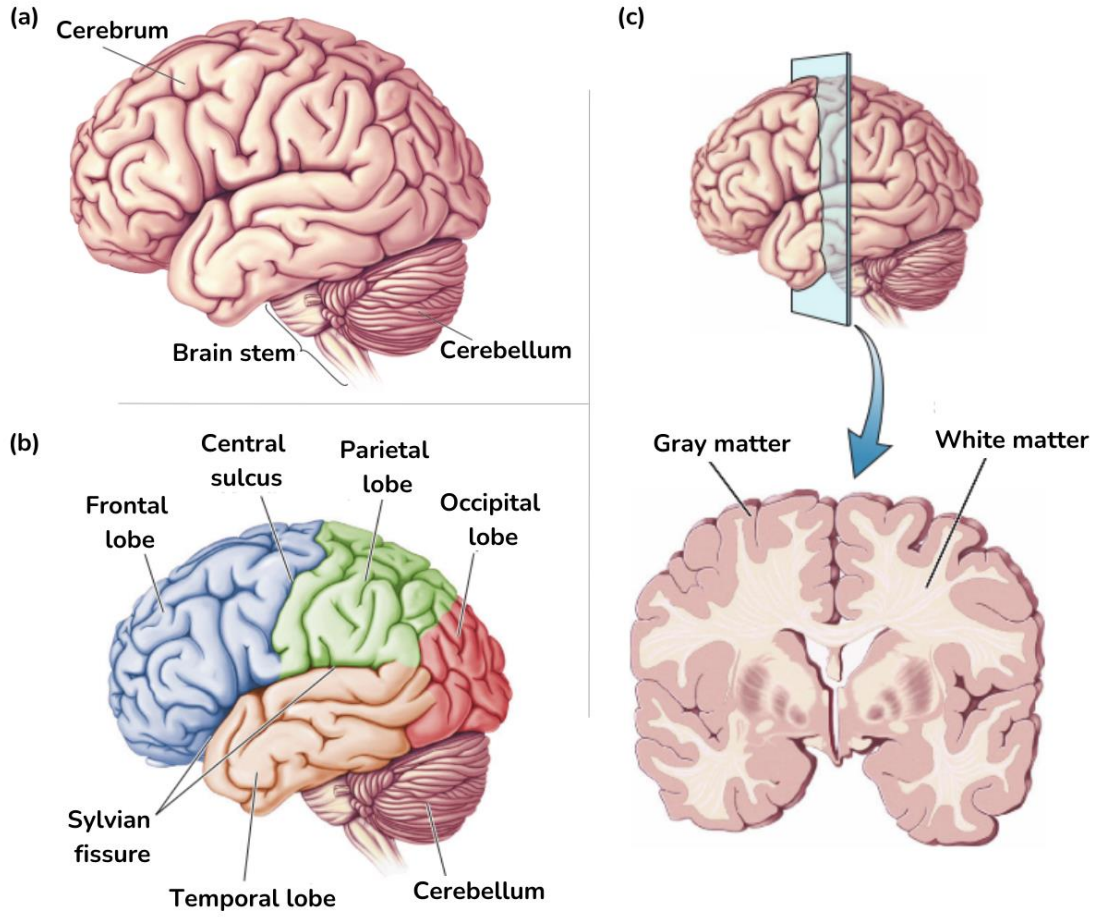
This section offers an overview of the essential concepts that underpin the development of this research. It begins by delving into brain anatomy, with a specific emphasis on the intricate structure of white matter (WM) and its fiber connections. We explore magnetic resonance imaging principles, focusing on diffusion-weighted imaging, detailing the acquisition process and the methods used to generate a representation of WM fibers through signal modeling, a technique known as tractography.

Furthermore, this chapter serves as a gateway to essential concepts within the realm of artificial intelligence, machine learning, and deep learning. It intricately explores the intersection of neural networks in both biological brains and computer systems, providing valuable insights into their operations. Particular attention is directed towards the convolutional neural networks, U-Net, which hold a central role in our research endeavors. This chapter concludes by framing the research problem and providing a final contextualization for our study. This comprehensive groundwork sets the stage for a deeper exploration of our research objectives and methodologies.

1.1 BRAIN AND WHITE MATTER

Our perception, cognition, interpretation, and response to the world around us are all orchestrated by the brain – an intricate masterpiece within our body. This complex structure can be partitioned into three fundamental units: the brainstem, the cerebellum, and the cerebrum (Figure 1 – a). The brainstem serves as a vital link connecting the spinal cord to the brain, and regulates many automatic body functions, including heart rate and breathing. Nestled in the lower posterior region of the brain, the cerebellum assumes the role of motor control, coordination, and balance. The cerebrum stands as the largest and most sophisticated component, intricately intertwined with intellectual activities and governing a wide spectrum of body functions (BEAR; CONNORS; PARADISO, 2015).

Figure 1 - (a) The brain divisions: cerebrum, brain steam and cerebellum. (b) The four lobes of the brain: frontal lobe, parietal lobe, occipital lobe, and temporal lobe. (c) Visualization of brain gray and white matter in coronal view.



Source: Adapted from (BEAR; CONNORS; PARADISO, 2015).

The cerebrum is partitioned into left and right hemispheres by a profound fissure known as the great longitudinal fissure. These hemispheres are interconnected through a specialized nerve fiber bundle called the corpus callosum. Within each hemisphere, four principal lobes—frontal, parietal, temporal, and occipital—carve out specific domains of functionality (Figure 1 – b).

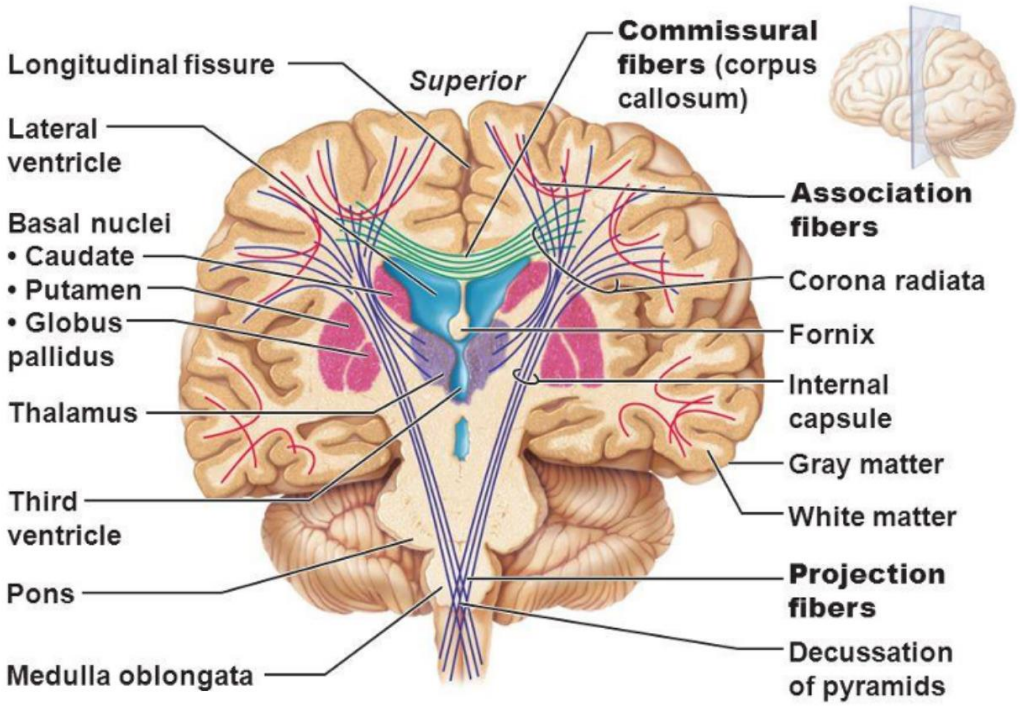
The frontal lobe takes the reins of voluntary movements, speech, and intellectual functions. Parietal lobes shoulder the responsibility of perception and interpretation of sensory inputs. Temporal lobes play a pivotal role in memory consolidation, auditory processing, and organizational tasks. Lastly, the occipital lobes diligently oversee all aspects related to visual processing. Also, in the inner brain, we have some structures, including the hippocampus, constituting the basal nuclei which coordinates the information exchange across various brain regions. Surrounding and protecting these intricate brain components, the cerebrospinal fluid (CSF) fills the cavities and ventricles (BEAR; CONNORS; PARADISO, 2015).

The brain's surface is a marvel of intricate wrinkles, displaying sulci, fissures, and elevated regions nestled within the troughs known as gyri. This remarkable expanse is the cerebral cortex, commonly referred to as gray matter, which encompasses neuron cell bodies, dendrites, myelinated and unmyelinated axons, glial cells, and capillaries.

Beneath this cerebral cortex, serving as an intricate connective web linking diverse cortical regions, resides a white-colored region aptly labeled the "white matter" (Figure 1 – c). While sharing a composition akin to that of gray matter, their distinction lies in the ratio of cell bodies to myelinated axons. Gray matter predominantly houses cell bodies, bestowing the tissue with its characteristic gray hue. In contrast, WM is rich in myelinated axons, which imparts the tissue its distinctive white coloration.

Enveloped in myelin, bundles of axons form the WM, assuming a central role as a vital conduit for transmitting neural signals across diverse regions of gray matter. Axons sharing similar destinations tend to converge, forming expansive bundles aptly referred to as WM tracts. These tracts can be classified according to their connectivity in three main groups of fibers: projection, association and commissural fibers as represented in Figure 2 (WAKANA et al., 2004; WYCOCO et al., 2013).

Figure 2 - Representation of the three main groups of white matter fibers: association fibers, commissural fibers, and projection fibers.



Source: (MARIEB; HOEHN, 2012).

Projection fibers serve as the bridges connecting the cortex with lower brain regions and the spinal cord. This class of fibers neatly divides into efferent fibers, responsible for conveying information away from the cortex, and afferent fibers, tasked with ferrying sensory data towards the cortex. Among the notable examples are the corticospinal tract, corticothalamic tract, auditory radiation, and thalamic radiation.

Association fibers, on the other hand, forge connections between discrete cortical regions within the same cerebral hemisphere. These fibers exhibit a spectrum of ranges, spanning both long-distance and short-range varieties, with the latter encompassing intriguing U-shaped configurations. Notable among the extensive pathways are the cingulum, arcuate fasciculus, superior longitudinal fasciculus, and inferior fronto-occipital fasciculus, while the shorter pathways comprise the fornix and uncinate fasciculus.

Concluding this intricate neural network are the commissural fibers, often referred to as transverse fibers. This network of fibers plays an indispensable role in seamlessly connecting the two hemispheres of the brain. Noteworthy exemplars within this category include the corpus callosum, anterior commissure, posterior commissure, and hippocampal commissure (BULLOCK, 2022).

In general, the long-range connections are comprised of commissural, projection, and association fibers that connect distant brain regions and facilitate the transfer of multimodal sensory information from subcortical nuclei to sensory and association cortices. Whereas short-range connections are comprised of fibers that connect neighboring cortical regions within lobes, presenting higher levels of curvature than other tracts (SALADIN, 2012; OYEFIADE et al., 2018). In a more quantitative way, the differentiation between long, medium, and short fibers can be made by is length size, with short fibers presenting a length < 40 mm and long fibers >150 mm (BAJADA et al., 2019).

1.1.1 SHORT TRACTS AND NEUROLOGICAL DISORDERS

The short fibers connect brain areas in close physical proximity with each other, linking WM regions as the uncinate fasciculus, that connects the frontal and anterior temporal lobe, the anterior commissure, that links the two olfactory bulbs, the fornix, connecting the hippocampus to the mammillary bodies, and posterior commissure, linking the hemispheres of the brain in the mesencephalon (MAGRO et al., 2012; CATANI et al., 2002). The short-range fibers are the majority of WM connections and have been associated with several neurological and

psychiatric diseases, such as schizophrenia, Alzheimer's disease, epilepsy and multiple sclerosis (ANAND; DHIKAV, 2012; YOSHINO et al., 2020; WU et al., 2022).

The anterior commissure (AC) is a vital interhemispheric WM structure in the brain, located in the anterior region of the third ventricle. It facilitates communication between the left and right cerebral hemispheres, particularly in olfactory. Abnormalities or lesions in the AC have been associated with various neurological and psychiatric conditions (MOON et al., 2008). For example, in schizophrenia, it has been linked to symptoms like disorganized thinking and hallucinations, highlighting its role in interhemispheric communication deficits (ÇAVDAR et al., 2021). In epilepsy, structural issues in the AC can contribute to an increase of seizures (FELON et al., 2021).

The posterior commissure (PC), located at the epithalamus, is related to visual and pupillary reflexes, including the pupillary light reflex and near reflex (OZDEMIR; 2014). Clinical studies and animal models have highlighted its significance, particularly in cases of abnormalities or lesions leading to conditions like light-near dissociation in pupils and visual impairments due to midbrain lesions. Beyond its physiological role, the PC also holds importance as a landmark in neuroanatomy, used for quantitative assessment of brain morphology (PRAKASH; NOWINSKI, 2006; CHOI et al., 2013). Its precise location serves as a valuable reference point for neurosurgeons and radiologists when navigating the brain, contributing to safer and more accurate surgical interventions and diagnostic procedures.

The hippocampal commissure (HC) connects the two fornixes. Its function is intimately linked to memory, emotional and learning processing. The hippocampus relies on this commissure to facilitate the exchange of information between its right and left hemispheric counterparts. Studies have revealed that disruptions or damage to the HC is related to memory disorders, like Alzheimer's disease, and neuro-immunological diseases, as Multiple Sclerosis (ANAND; DHIKAV, 2012; HEINE et al., 2020). In Alzheimer's, the degeneration of these connections can contribute to profound memory impairments and cognitive decline. Furthermore, abnormalities in the HC have been associated with epilepsy, where its role in synchronizing and regulating neural activity is compromised, potentially leading to seizure generation (SPENCER et al., 1987; KALAIVANI; SUNDARESWARAN; 2019).

The fornix (FX) is a bundle of WM fibers that connects the hippocampus to various regions of the brain. By facilitating the communication between these regions, FX takes part in the formation and retrieval of memories, as well as the regulation of emotions. Studies have unveiled its involvement in memory disorders such as Alzheimer's disease, where FX

degeneration can contribute to severe memory deficits and cognitive decline (THOMAS; KOUMELLIS; DINNEN, 2011). Similarly, abnormalities or damage to this tract have been associated with epilepsy, potentially leading to seizures by disrupting the synchronization of neural activity (KUZNIECKY et al., 1999; CONCHA et al., 2010).

The uncinate fasciculus (UF) serves as a WM tract that connects the frontal and temporal lobes of the brain, being associated to language and emotional processing. Alterations in the UF have been linked to language impairments, such as aphasia, which can significantly impact an individual's ability to communicate effectively (HARVEY et al., 2013). Additionally, disruptions in this tract have been associated with mood disorders, like schizophrenia, depression, and posttraumatic stress disorder (PTSD), and dementia diseases, such as Alzheimer's and Parkinson's (KUBICKI et al., 2002; YASMIN et al., 2008; BHATIA et al., 2018; KOCH et al., 2017; DI TELLA et al., 2020).

1.2 NUCLEAR MAGNETIC RESONANCE AND DIFFUSION IMAGING

Magnetic Resonance Imaging (MRI) stands as a remarkable imaging modality that grants us the capacity to visualize and delve into an array of soft tissue attributes within the body. The foundation of this potent technology lies at the intersection of three fundamental physics principles: nuclear, magnetic, and resonance.

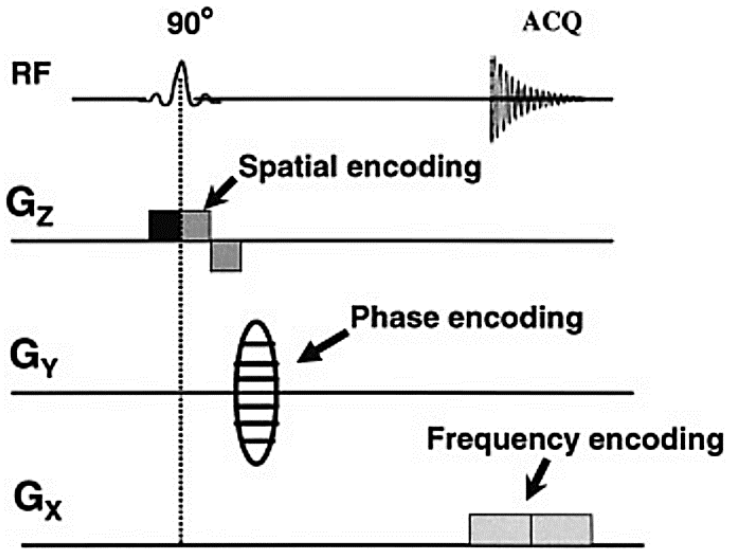
The "nuclear" facet refers to the manipulation of the nucleus in the hydrogen atoms, i.e., ^{1}H. This manipulation transpires within the framework of a high-intensity magnetic field, compared to earth magnetic field, typically 1.5 or 3 Tesla in the clinical scanners. The second facet, "magnetic," stems from the magnetic field itself, which exerts its influence upon the atomic nuclei. It's within this magnetic milieu that the protons of hydrogen atoms align and interact.

Resonance, the third key component, comes to the need to synchronize the radiofrequency of an oscillating magnetic field with the precessional frequency of certain spins nuclei within the tissue (Larmor frequency). These pulses perturb the protons from their initial equilibrium magnetization state. The protons will then return to their equilibrium magnetization state releasing energy that is detected by the receiver coil (MR signal) and compose the Nuclear Magnetic Resonance (NMR) phenomenon.

One last key component to be able to reconstruct the image is the magnetic field gradients. This is because, besides we obtain NMR signal, we still do not have the spatial information to allocate a position to the different signals. The gradients have a linear spatial

pattern to codify the magnetization frequency according to the spatial position. For that, pairs of coils are placed in each spatial direction (x, y, z) having identical properties, however they are applied in distinct moments and in different directions. With these gradients is possible to select the slice plane (selection gradient), alter the phase of spins (phase gradient) and encode the frequency signal (readout gradient) as demonstrated in Figure 3.

Figure 3 - Representation of spatial, phase, and frequency encoding in a typical MRI sequence.



Source: (MANDAL, 2006).

Local magnetic interactions are characteristic of each tissue, producing different rates of magnetization recovery. Commonly referred as relaxation times, this magnetization can be decomposed in longitudinal (T1) and transverse (T2) processes.

The T1 relaxation refers to the recovery of magnetization along the longitudinal direction, and the T2 relaxation characterize the loss of phase coherence in transverse plane. These relaxation times are characteristic for each tissue based on the local magnetic interactions. Tissues with fast molecular motion, such as CSF, blood and edemas, present longer T1 and T2 relaxation times. Soft tissues, for example gray and white matter, fat and muscle have shorter T1 and T2 values, producing the basic contrast mechanism of MR.

Not only are relaxation times mechanisms for contrast in MRI, but there are also other factors that affect the detected signal intensity. These factors include motion, occurring at various scales, from the macroscopic level, such as in angiography, to the microscopic level, as seen in diffusion-weighted imaging (DWI).

Water is present in all our body, but principally in our brain, and to traverse the intricate pathways of our tissues, it employs the diffusion. This phenomenon involves the molecular and

particulates motion within a medium, driven by the random energy of thermal agitation, often referred to as Brownian movement. In 1885, it was formulated the Fick's first law of diffusion, were states that the diffusive process drives from higher concentration to areas with lower concentration, defining the diffusion coefficient (equation 1).

$$F = -D\frac{\partial C}{\partial x} \quad (Eq.1)$$

With the negative sign highlighting that the flow is in the opposite direction to the increasing concentration and D is the diffusion coefficient. If there is no preferential diffusion direction (isotropic diffusion), the global behavior can be described as the derivate of the first law in time:

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right) = D\nabla^2 C \quad (Eq.2)$$

Which is equivalent to the Gaussian distribution, were P is the probability of a displacement X at time t:

$$P(X,t) = \frac{1}{\sqrt{D4\pi t}}\exp\left(-\frac{X^2}{4Dt}\right) \quad (Eq.3)$$

However, if instead of describing the process in terms of a concentration, we describe as a probability of finding a particle in a position in time, the mean square dynamic displacement will take the Gaussian form:

$$\langle R^2\rangle = \int_{-\infty}^{+\infty} R^2 P(R,t)dR = 6Dt \quad (Eq.4)$$

In which R is the net vector distance travelled by the molecule in time t. If we also consider the random nature of the process, we will obtain that for *n* random displacements of constant length ξ in an interval time τ:

$$\langle R^2\rangle = 6Dt = \frac{t}{\tau}\xi^2 \quad \rightarrow D = \frac{\xi^2}{6\tau} \quad (Eq.5)$$

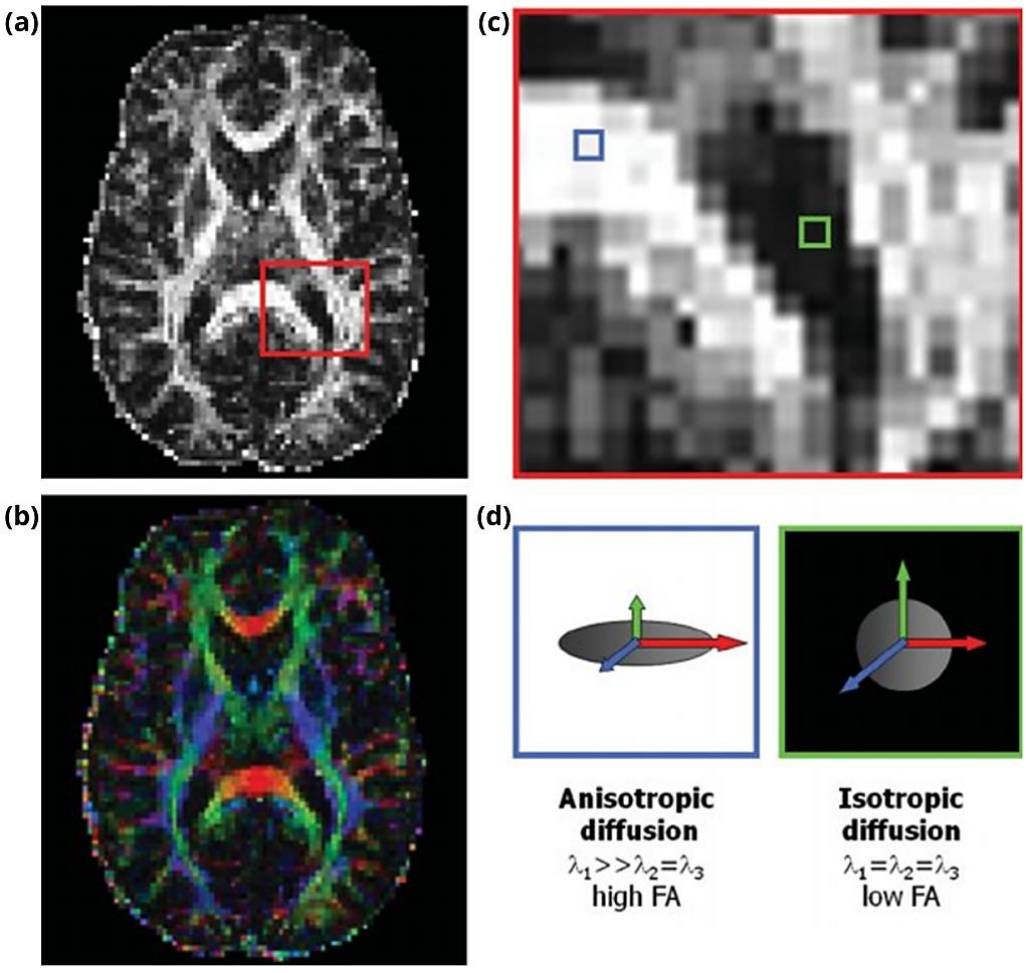
This description explains the relationship of the diffusion coefficient and the displacement of the molecule at a given time *t* (TOFTS, 2018). However, for molecules within a living tissue other factors has to be taken in count in order to interpret these measurements.

Through the Stokes-Einstein relation (equation 6) we understand that the diffusion also depends on temperature (T), viscosity of the solution (η), and hydrodynamic radius of the molecule (R), as depicted in equation 6 (SEIBERLICH et al., 2020).

$$D = \frac{k_b T}{6\eta R} \qquad (Eq. 6)$$

k_b is the Boltzmann's constant. Within our brain, the isotropic diffusion characterizes the behavior of CSF, a relatively homogenous fluid. However, the brain's architecture comprises several elements—such as membranes and axonal fibers—that introduce constraints to molecular movement. As a result, molecules develop directional preferences, leading to a phenomenon where the diffusion becomes favored in specific directions. This variant is known as anisotropic diffusion and can be characterized by a fractional anisotropic coefficient (FA) (Figure 4).

Figure 4 - Demonstration of regions with anisotropic diffusion and isotropic diffusion. (a) FA map, (b) color-coding FA map, (c) zoom region in FA map with blue square in fiber region and green square in CSF region, (d) representation of the signal in blue and green highlighted regions.



Source: (GEVA; CORREIA; WARBURTON, 2011).

In an anisotropic media, the probability of the molecular displacements still follows a multivariate Gaussian distribution over time. However, in that case the diffusion is described as a 3x3 tensor matrix, proportional to the variance of the Gaussian distribution. The mobility constraints can take on varying degrees of restriction. When water molecules encounter a singular direction for movement, their displacement becomes confined by barriers within the environment. On the other hand, hindered movement allows diffusion but with a favored direction, albeit without complete confinement.

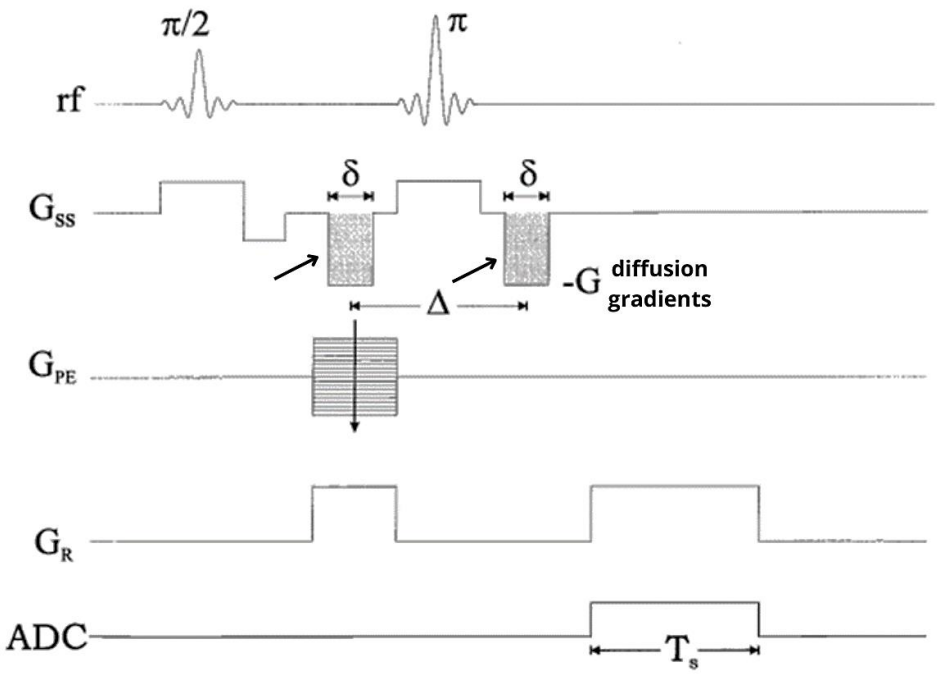
In the brain, the random movement of water molecules depends on many factors including limitations imposed by cellular membranes, neural fibers, and macromolecules. Notably, the presence of white matter exacerbates the constriction and directionality of water movement, owing to the aligned orientation of nerve fiber tracts. Armed with this understanding, it becomes feasible to design MR scanner sequences tailored to capture the distinctive diffusion traits of water within tissues. This tailored approach facilitates the investigation of WM characteristics and properties.

1.2.1 DWI ACQUISTION

DWI is a technique that detects and amplifies the signal reduction resulting from the erratic motion of water protons within tissues (BASSER; JONES, 2002). This enhancement of diffusive attenuation is achieved through the utilization of pulsed magnetic field gradients. These gradients introduce a linear magnetic field disparity in a specific direction, which allows the differentiation of molecules with isotropic and anisotropic diffusion.

Even before MRI scanners, Stejskal and Tanner proposed the pulsed gradient spin echo (PGSE) sequence that introduced the diffusion weighting in NMR acquisition in 1965 (STEJSKAL; TANNER, 1965). In its simplest instantiation, this sequence incorporates short gradient pulses (with a duration denoted as δ, a certain amplitude labeled as G and separated by a diffusion time interval Δ). These gradient pulses, referred to as diffusion gradients, are administered prior to and after the application of the refocusing RF pulse in a spin echo sequence. Figure 5 represents the idea of Stejskal and Tanner in a current sequence.

Figure 5 - Conventional 'Stejskal-Tanner' spin echo pulse sequence for DWI acquisition, with slice selection and diffusion sensitive gradients (shaded) applied along one physical dimension.

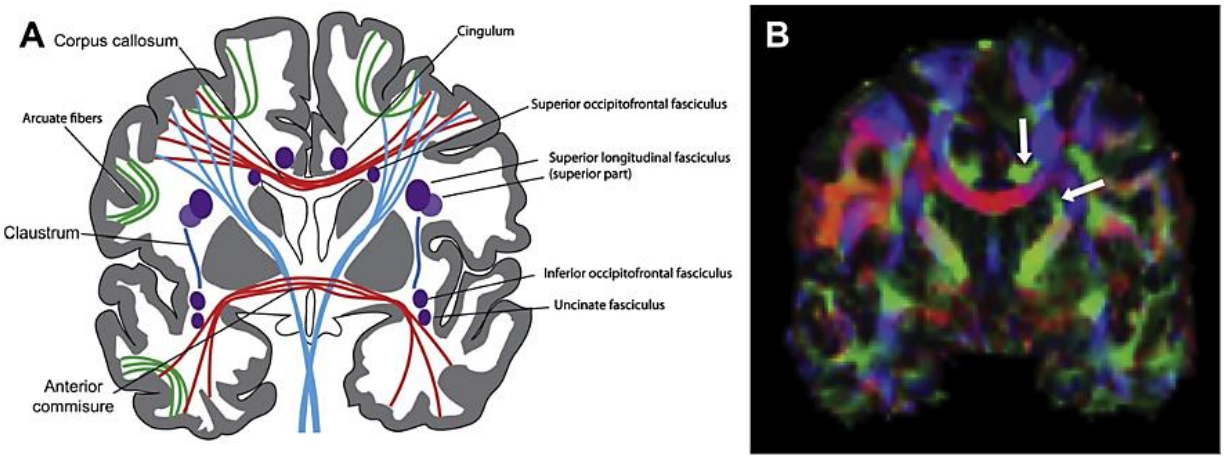


Source: Adapted from (HAACKE, et al., 1999).

Upon implementing this sequence, the initial application of gradients results in the accumulation of a phase shift, determined by the positions and movements of water molecules over the course of its duration. Following this, a 180º pulse serves to invert the phase, while the second gradient prompts another phase shift equal in magnitude to the initial one. Consequently, if the water molecules remain static their phase shifts would bear opposing signals, effectively nullifying each other. Yet, if the water molecules are in motion within the time interval Δ, these shifts diverge, leading to a reduction in echo amplitude due to partial refocusing (HAACKE et al., 1999).

This property is quantified through the diffusion coefficient, representing the extent of diffusion occurring within the tissue, specifically in the gradient direction employed during the acquisition. The intricate architecture of brain fibers gives rise to anisotropic and impeded water diffusion within white matter tissues. Consequently, depending on the orientation of the diffusion gradient, the directionality of fiber bundles, fasciculi, and tracts, i.e. WM regions can be observed (Figure 6).

Figure 6 - (a) Anatomical representation of the fiber regions. (b) FA map showing regions of known white matter tracts. Colors code the fiber directionality.



Source: (WYCOCO et al., 2013).

The mathematical representation of the DWI signal is outlined in equation 7. Besides the DWI image is strongly dependent on diffusion, the signal also depends on the transverse relaxation time, T2, of the underlying tissue. So, if the aim is to measure the diffusion coefficient, a minimum of two acquisitions with different diffusion weightings is necessary, one acquired without diffusion weighting and the other acquired with diffusion weighting.

$$S = S_0 \cdot e^{-bD} \qquad (Eq.7)$$

Where S is the measured diffusion-weighted signal and S_0 is the measured signal without the diffusion gradients application, in other words, with b=0. The exponential attenuation is a result of the interaction between the b-factor defined by the pulse sequence and the diffusion coefficient D as a property of the sample, achieved through their multiplication. The b value is related to the strength, interval and duration of the diffusion weighted gradients, as defined in equation 8.

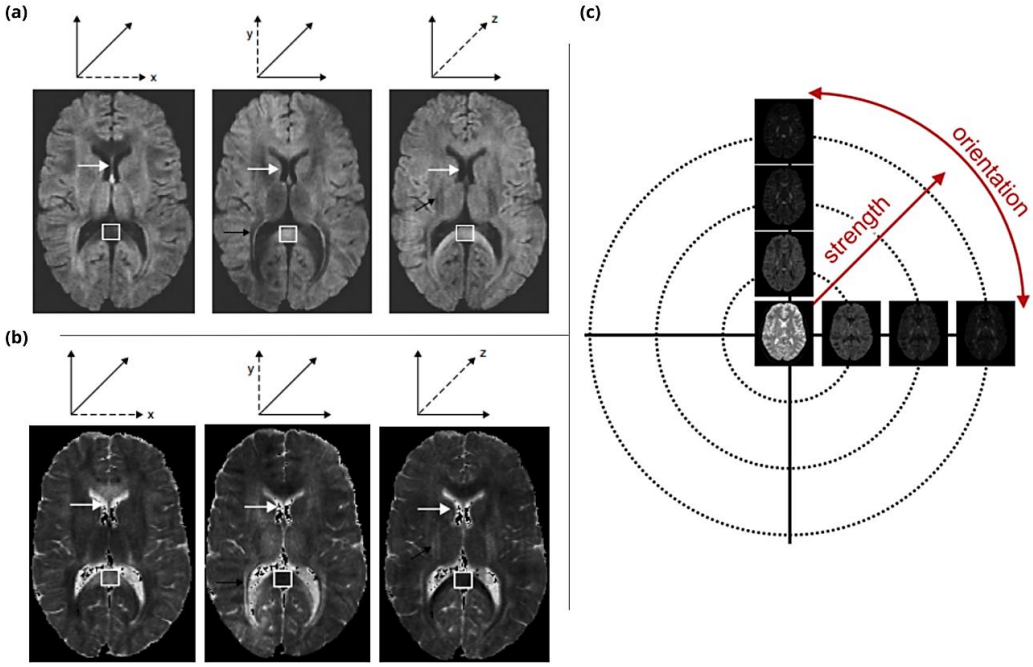
$$b = \delta^2 G^2 \Delta^2 \left(\Delta - \frac{\delta}{3}\right) \qquad (Eq.8)$$

When employing a higher b value, the resulting image exhibits heightened contrast, accentuating the distinctions between regions with higher diffusivity, for example the ventricles, resulting in darker regions in the DW image, and lower diffusivity. With these equations, we can obtain the diffusion coefficient for a specific direction as follows:

$$D = \frac{\log\left(\frac{S_0}{S}\right)}{b} \qquad (Eq. 9)$$

If S, S_0 are known for every voxel in the image, applying a voxel-wise calculation using equation 7 yields a diffusion map (Figure 7). Within this map, each voxel's value represents the average diffusion coefficient of the enclosed tissue, evaluated along the diffusion gradient's applied direction. It is worth noting that the resulting diffusion map possesses an intensity scale that is flipped in comparison to the DWI employed for its generation. Regions characterized by increased mobility of water molecules along the direction of diffusion sensitization manifest as darker areas on the diffusion-weighted image, while they exhibit a vibrant brightness on the diffusion map (e.g., CSF).

Figure 7 - (a) DWIs highlighting the signal difference between CSF region and white matter region with different diffusion gradient orientations indicated by the dashed line in each case. (b) Diffusion maps in the same orientations as in (a), hypersignal can be observed in isotropic high diffusion areas, such as CSF. (c) Schematic representation of the diffusion signal through orientation and strength variations of the diffusion gradient in the multi-shell acquisition approach.



Source: Adaptation from (TOFTS; DOWELL; CERCIGNANI) and 3rd MRtrix3 Workshop (TOURNIER et al., 2019).

Diffusion, however, is a three-dimensional phenomenon encompassing both direction and shape. To delve deeper into the directional behavior of water molecules, we can expand our study by obtaining diffusion-weighted images with encoding gradients applied along various axes. This acquisition strategy, involving the collection of diffusion-weighted images from

different directions while maintaining a constant b-factor, is named a single shell acquisition. This simple method captures the essence of a Gaussian diffusion distribution. Typical b-factor value used for single shell approach in brain exams is around 1000 s/mm².

In order to obtain a better characterization of the microscopic diffusion a multi-shell acquisition is used (Figure 7 – c). In this acquisition approach, several b-values and diffusion gradient directions are used in a same long acquisition. Yet, as we venture into higher b-values, the diffusion pattern assumes a non-Gaussian quality. This alteration brings forth distinct signal representations that elucidate the intricate displacement patterns of water molecules within diverse cerebral contexts (TOURNIER; MORI; LEEMANS, 2012).

1.2.2 MODELING INTRAVOXEL DIFFUSION

A diverse array of models exists to expound upon intravoxel diffusion in living tissues. Diffusion Tensor (DT) is the first and most consecrated model in literature. Basser et al. (1994) introduced a method involving multivariate linear regression to compute the diffusion coefficient as a tensor (BASSER; MATTIELLO; LEBIHAN, 1994). This process entails utilizing a non-diffusion weighted image alongside six or more diffusion-weighted measurements along noncollinear directions. The introduction of diffusion weighting is accomplished by concurrently applying diffusion gradients along combinations of the three physical axes.

In scenarios where the minimum number of diffusion-weighted images are acquired to define D—namely along the x, y, and z axes, as well as xy, xz, and yz—Equation 7 transforms into:

$$S = S_0 \cdot e^{-b\, \boldsymbol{g}^{\boldsymbol{t}} \cdot \vec{D} \cdot \boldsymbol{g}} \qquad (Eq. 10)$$

The resulting signal (S) along the diffusion gradient applied ($\boldsymbol{g}$) is equal now to the non-DW signal (S_0) multiplied by the exponential of the b-value and the apparent diffusion tensor ($\vec{D}$). The diffusion tensor is a 3x3 matrix characterizing the displacement in three axes:

$$\vec{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \qquad (Eq. 11)$$

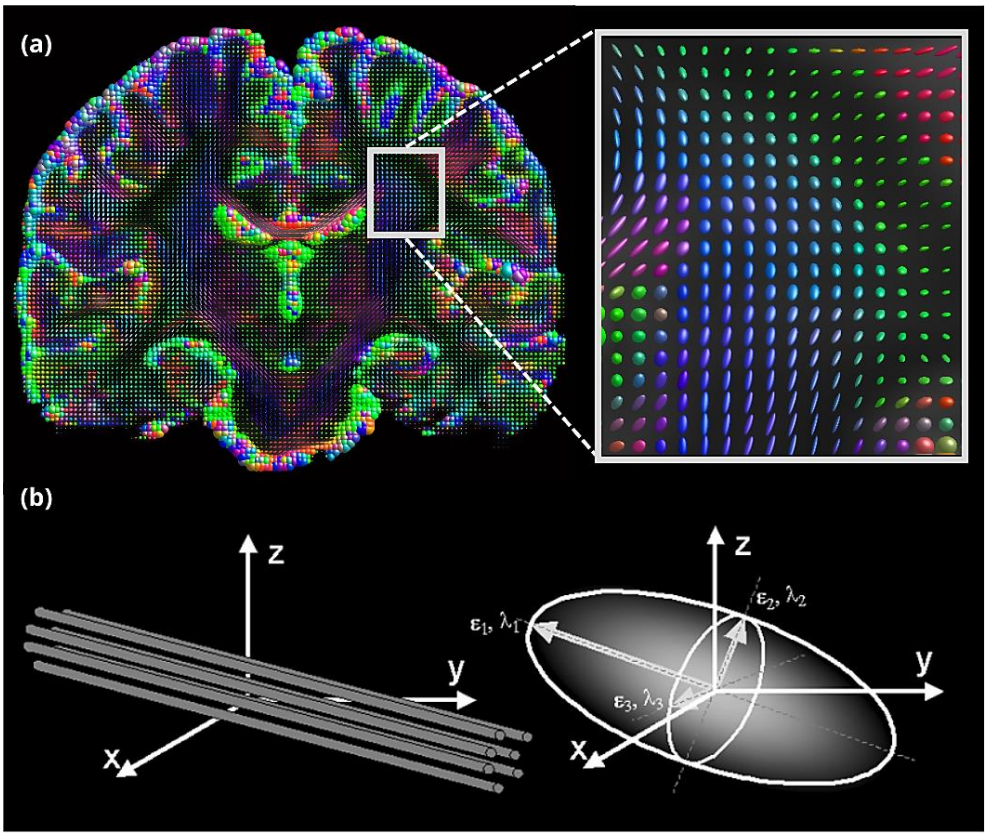
The diagonal elements signify diffusivities along these axes, and the off-diagonal elements denote correlations between them. This model requires a minimum of six DW image

acquisitions to derive tensor parameters. Consequently, the 'apparent diffusion tensor' name hinges on the specific directions employed during the experimental procedure, determined by the equipment used. From the diagonalization of the tensor, three eigenvalues ($\lambda 1$, $\lambda 2$, and $\lambda 3$) along with their corresponding eigenvectors ($\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$, and $\boldsymbol{\varepsilon}_3$) can be derived as demonstrated in equation 12. The eigenvalues are ordered with the decreasing of the eigenvector ($\lambda 1 > \lambda 2 > \lambda 3$) in way that $\boldsymbol{\varepsilon}_1$ represents the principal direction of diffusivity. The combination of eigenvalues and eigenvector is unique for each DT model, characterizing the diffusion properties of the tissue of interest.

$$\overleftrightarrow{D} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \end{pmatrix} \qquad (Eq. 12)$$

These parameters allow for the interpretation of the DT model as an ellipsoid, representing the diffusion pattern within each voxel (Figure 8). The principal axis of this ellipsoid and its configuration correspond to the primary diffusion direction and the extent of anisotropy within each voxel.

Figure 8 - (a) Intra-voxel diffusion modeling using the diffusion tensor model. (b) Representation of the fiber whiting the voxel as an ellipsoid.

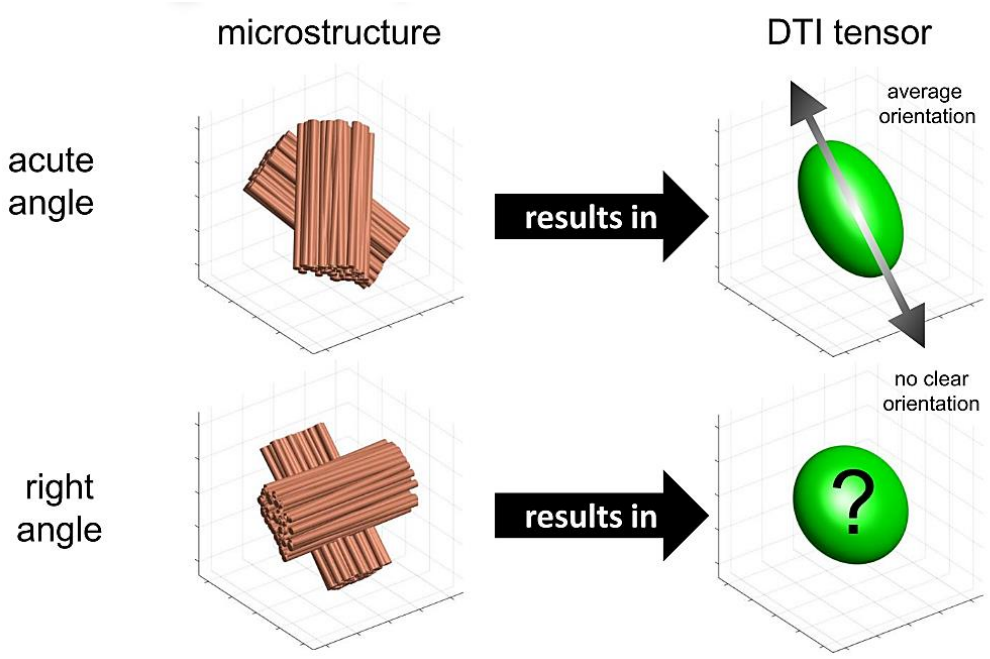


Source: a – Elaborated by the Author, b - (ALEXANDER et al. 2007).

A compelling visualization method involves combining color-encoded anisotropy information with ellipsoidal representation, as depicted in Figure 8 – a. These colors mirror the principal direction of diffusion, aligned with the first eigenvector of the tensor ($\boldsymbol{\varepsilon}_1$). In a customary convention, the colors green, blue, and red correspond to the antero-posterior, cranio-caudal, and latero-lateral directions, respectively. The eccentricity of the ellipse conveys the degree of anisotropy inherent in the diffusion process.

Besides the DT establishment as the conventional model, it presents certain limitations and weaknesses. The model struggles to differentiate multiple fiber orientations within a single voxel (TOURNIER et al., 2004), leading to suboptimal fits of diffusion data, as exemplified in Figure 9. This becomes particularly relevant as about one-third of white matter voxels harbor crossing fibers (BEHRENS et al., 2007). Decreasing the reliability of the model and further structure integrity analysis.

Figure 9 - Representation of the crossing fiber problem in DT model.

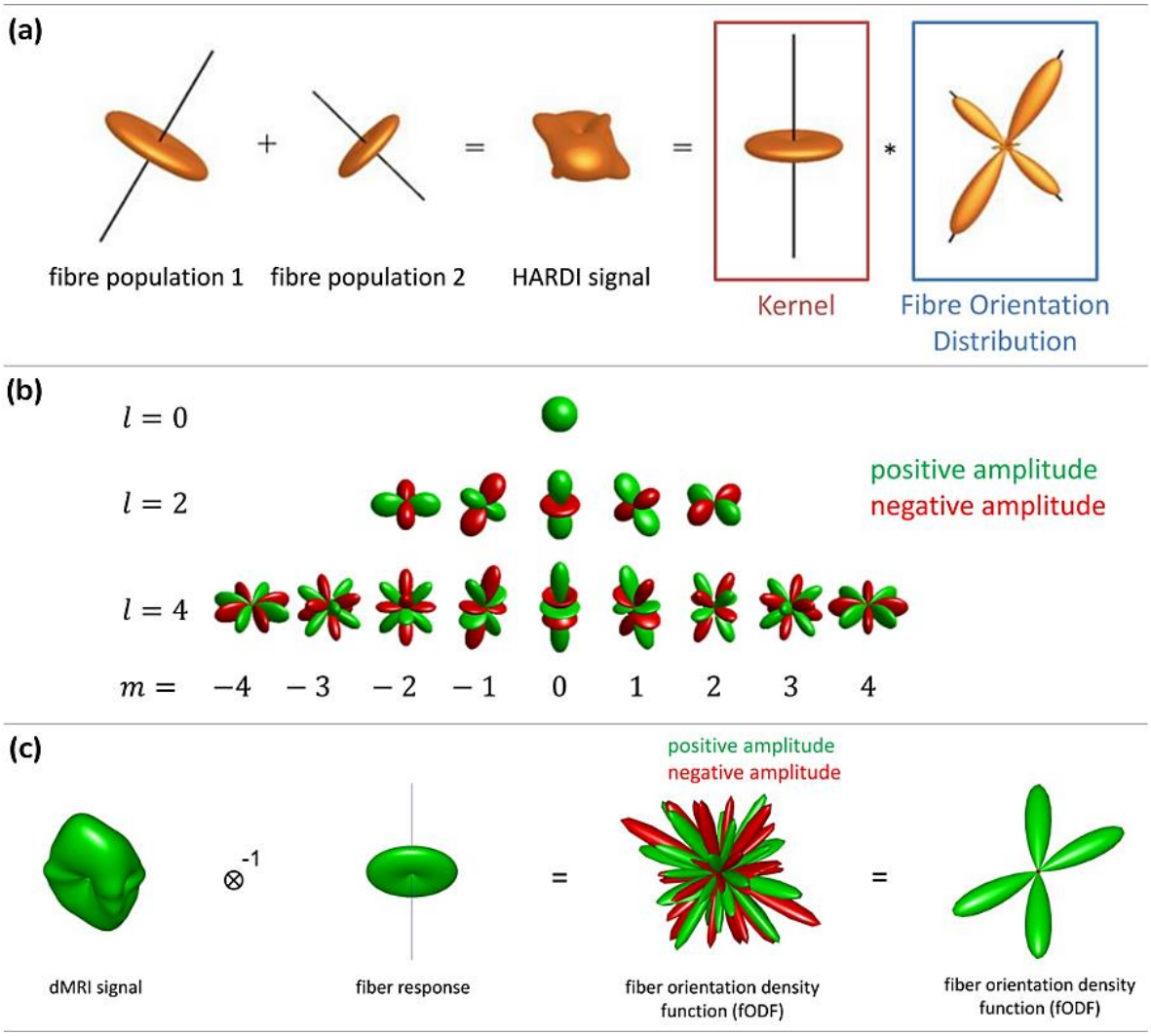


Source: 3rd MRtrx3 Workshop (TOURNIER et al., 2019).

In response to these limitations, several alternative models have been proposed being classified in two main groups: those that seek characterize the signal through q-space and those that attempt to estimate the fiber orientations. Each of these approaches presents their own limitations. Methods based in q-space, which suppose infinitesimally short diffusion pulses, requires a significant extension of the acquisition protocol, restricting their applicability for routine clinical practice.

In contrast, the strategies that attempts to recover the fiber orientations directly are more practical, aiming recover the Fiber Orientation Distribution (FOD) as a continuous distribution, instead of a discrete interpretation (TOURNIER, 2004). The idea behind this approach is that the signal in a voxel is composed by the sum of each fiber population, what can be expressed as a linear combination (Figure 10 – a).

Figure 10 - (a) How fiber crossing problem can be modeled. (b) Spherical harmonic representation. (c) Constrained Spherical Deconvolution solution.



Source: 3rd MRtrx3 Workshop (TOURNIER et al., 2019).

One way to deal with this issue is to use spherical harmonics (Figure 10 – b). A first approach would be considering the fiber populations, having one dedicated tensor to each of them. In this way, it becomes possible to resolve a voxel containing two distinct fiber populations through the utilization of a two-tensor model (equation 13 – first part). This strategy extends the multitensor concept, where each fiber population is denoted up its own diffusion tensor, by increasing the number of fiber populations to infinity. Mathematically speaking, the

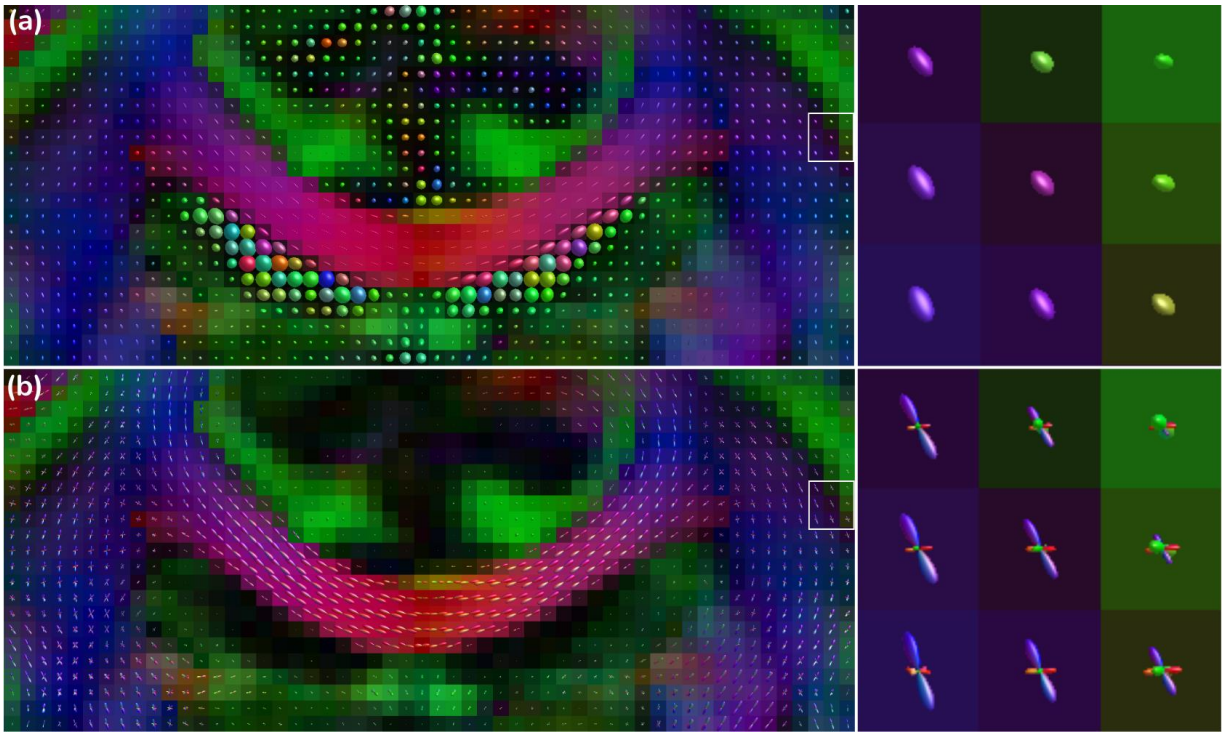
sum becomes an integral over the distribution of fiber orientations (equation 13 – second part), which allows the problem to be written as a spherical deconvolution operation, represented in Figure 10 – c.

$$S(\boldsymbol{g}, b) = \sum_{n=1}^{N} f_n e^{-b\boldsymbol{g}\boldsymbol{D}_n\boldsymbol{g}} \;\rightarrow\; S(\boldsymbol{g}) = \int_{\boldsymbol{u}} F(\boldsymbol{u}) R(\boldsymbol{g}\cdot\boldsymbol{u}) d\boldsymbol{u} \qquad (Eq.13)$$

Where $F(\boldsymbol{u})$ is the fODF and $R(\boldsymbol{g}\cdot\boldsymbol{u})$ is the fiber response containing the corresponding spherical harmonics. Yet, this results in a fiber orientation density function (fODF) with positive and negative amplitudes, with the negative values being an outcome of the susceptibility to noise. To address this, constrains based on prior knowledge about this type of distribution is applied leading to our final result, and to the name of this technique: Constrained Spherical Deconvolution (CSD) (JEURISSEN et al., 2014).

Figure 11 compares the DT model and CSD, highlighting the refinement of signal interpretation obtained with the fiber orientation distribution, by solving the crossing fibers problem.

Figure 11 - Comparison between diffusion tensor model (a) and constrained spherical deconvolution model (b).



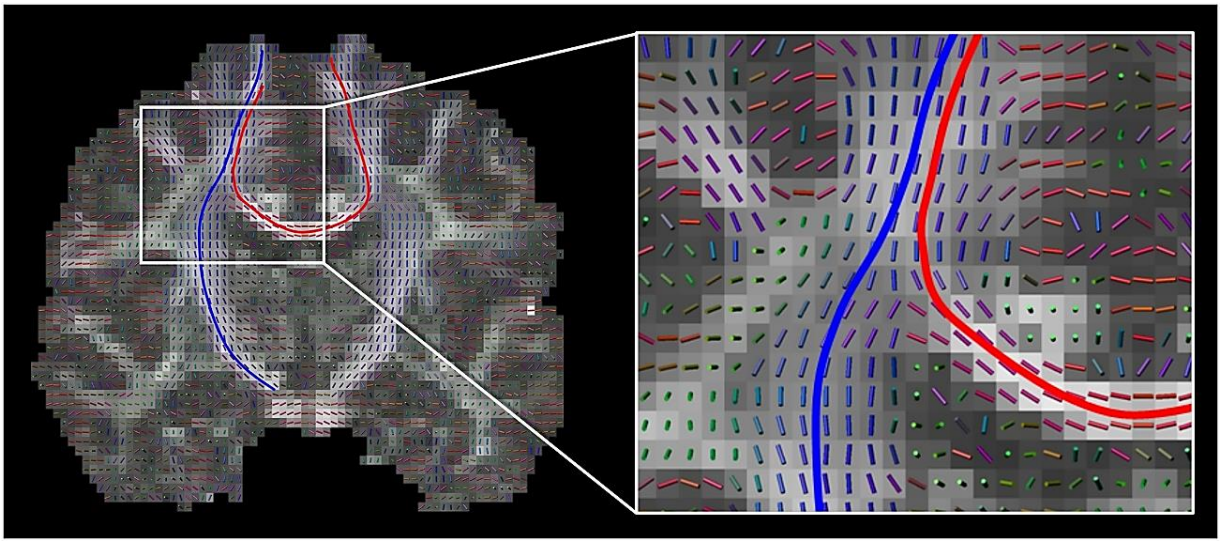
Source: Elaborated by the Author.

1.2.3 TRACTOGRAPHY

After the characterization of intravoxel diffusion is possible to delineate the path taking by axonal fibers in the brain generating a three-dimensional visualization of reconstructed brain fiber tracts. Fiber tracking or tractography is the only non-invasive technique for study brain structural connectivity *in vivo*. There is a range of algorithms to reconstruct the computational visualization of the fibers, also called streamlines, and they can be separated in deterministic or probabilistic groups (SEIBERLICH et al; 2020).

Deterministic algorithms are the simplest and basically involves tracing the supposed preferred directions of the fibers until a defined stopping point is reached (Figure 12). Together with diffusion tensor model, one of the widely adopted and robust tractography techniques is the fiber assignment by continuous tracking (FACT) method (MORI et al., 1999). This technique continuously connects voxels utilizing angle and anisotropy thresholds to determine the endpoints of the streamlines. The deterministic concept originates from the trajectory that is sole settled for the given seed point.

Figure 12 - Simplified representation of how tractography algorithms work.

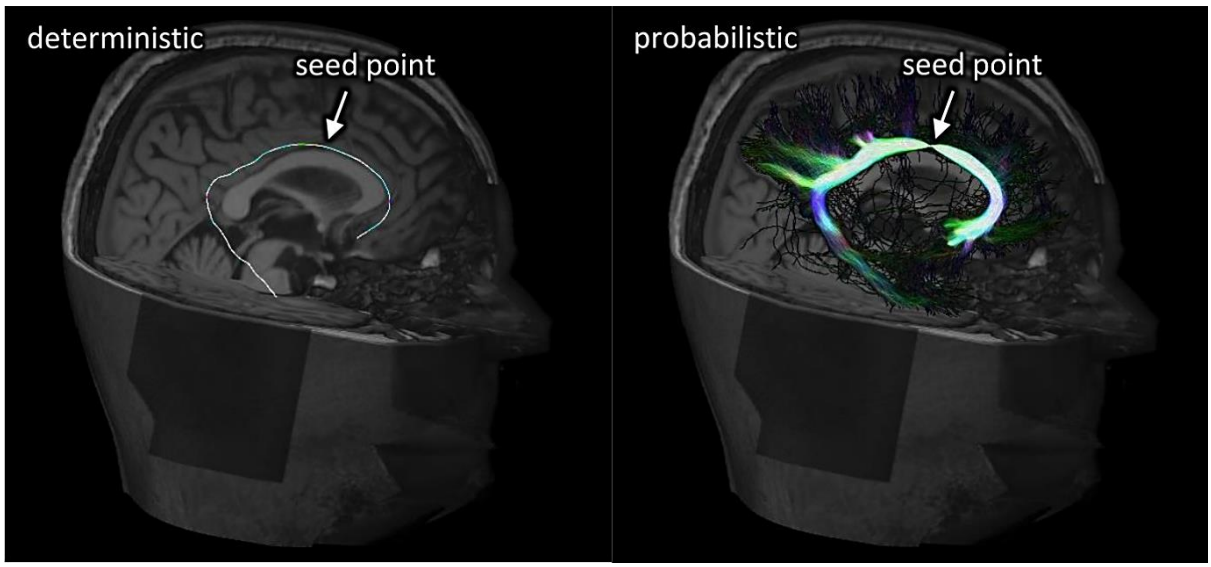


Source: (JEURISSEN et al., 2019).

However, the low signal-to-noise (SNR) of DWI leads to noise in response functions estimation and the tracking process propagate this error. Also, the fibers configurations, such as curving, diverging, and crossing can produce similar signals bringing ambiguity to tracking and decreasing the reliability of the tractography as the presence of spurious false positive connections becomes prominent. Probabilistic algorithms were developed with the objective to solve many of these issues through estimate a pathways distribution instead of a unique gauge.

The probabilistic approach relies on fODF, in which each step follows a random orientation conducted by the local fiber probability density function (PDF) resulting in a distribution of likely pathways, rather than only one streamline (Figure 13). In 2010, it was proposed the iFOD2 algorithm using a 2nd order integration strategy, being capable of tracking with high accuracy in fiber crossing and highly curved regions (TOURNIER; CALAMANTE; CONNELY, 2010). Instead of straight-line steps, the algorithm uses curved arcs in which the underlying FOD amplitudes are sampled through trilinear interpolation. In this concept, the more probable path is where the amplitudes are larger, but it may also follow the smalls ones, as long it remains above the FOD amplitude threshold all the path.

Figure 13 - Visual comparison of deterministic (a) and probabilistic (b) tractography algorithms executions from a seed point.



Source: 3rd MRtrx3 Workshop (TOURNIER et al., 2019).

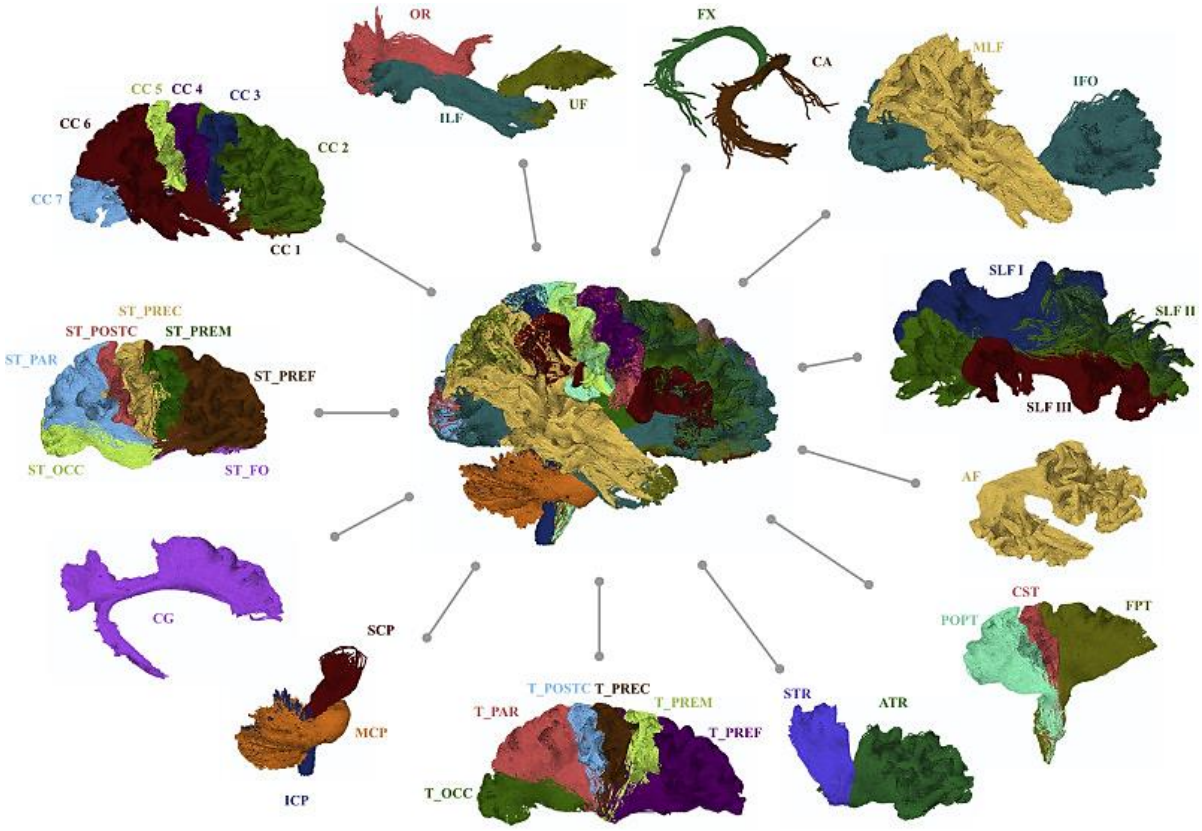
These algorithms enable to reconstruct comprehensive and more reliable whole-brain tractographies, which can subsequently be refined to isolate specific brain tracts and connections.

1.2.4 TRACT SEGMENTATION

Recognized brain tracts, based on anatomical information (WAKANA et al., 2007), can be isolated from whole-brain tractogram through track selection. These pathways can experience structural changes in different diseases and as part of the aging process, offering a wide range of clinical applications.

Tracts segmentation is useful to study specific brain connections and through diffusion MRI it is possible to analyze several distinct properties and characteristics of the WM, and also its tracts diversity (Figure 14). Fractional anisotropy, mean diffusivity, volumetry, number of streamlines, are some of the parameters that is possible to extract for WM integrity investigation.

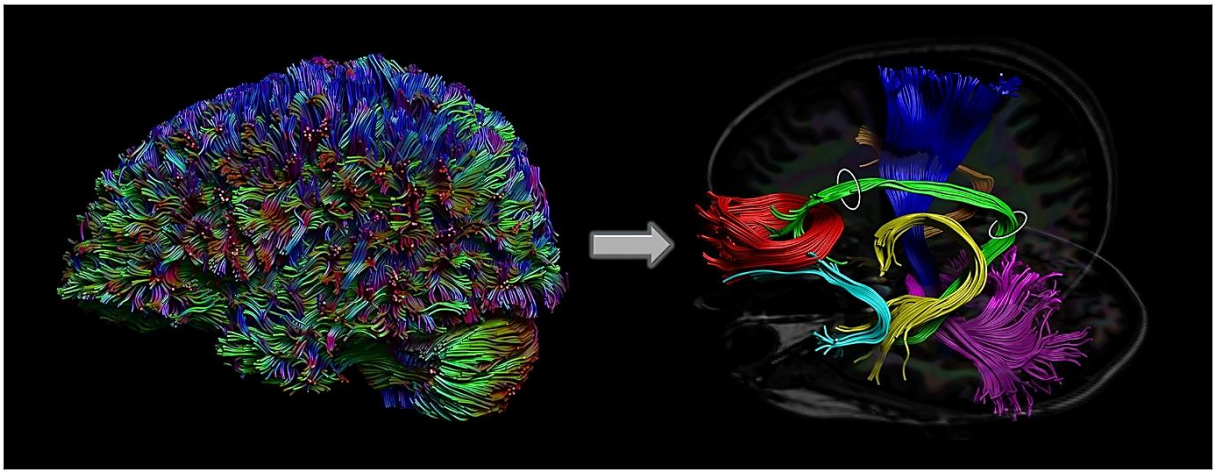
Figure 14 - Visualization of 72 tracts of white matter in the human brain.



Source: (WASSERTHAL; NEHER; MAIER-HEIN, 2018).

In practice, tracts segmentation is performed using regions-of-interest (ROIs) in which is known that the tract of interest is passing through a region as exemplified in Figure 15. These ROIs are defined as AND, OR and NOT operators, with the usage being dependent of the characteristic trajectory of each specific tract (JEURISSEN et al., 2019). For example, if the streamlines enter the inclusion regions, they are considered anatomically plausible, if not they are discarded. The exclusion ROIs can be used to discard not wanted fiber tracts. This strategy is called virtual dissection and is the more stablished method in literature for tracts segmentation. Although, there are several limitations with this approach since relies in prior anatomical knowledge.

Figure 15 - Multiple fiber bundle trajectories (right) are virtually dissected from a whole-brain tractogram (left), using tract selection. As an example, part of the cingulum bundle pathways (green) is dissected from the whole-brain tractogram using two circular “AND” gates (white).



Source: (JEURISSEN et al., 2019).

While virtual dissection remains the established gold standard, ongoing efforts are directed toward the exploration of automatic and semi-automatic approaches. These strategies aim to decrease the time-consuming nature of the process and reduce operator-dependent variability and can be classified into three main categories: ROI-based, clustering-based, and direct segmentation techniques.

ROI-based approaches employ anatomical priors to delineate regions of interest, which in turn assist in guiding the fiber segmentation process. An intuitive method for implementing this approach is through atlas-based tractogram filtering, which entails two key steps. First, the atlas containing streamline bundles is registered to the subject's data. Second, streamlines that do not entirely reside within a single bundle mask after registration are subsequently filtered out. As an example, Wassermann and his collaborators employed a Freesurfer parcellation method to delineate and characterize distinct WM tracts within the brain (WASSERMANN et al., 2016). However, atlas-based methods struggle to account for anatomical variability among subjects, potentially resulting in inaccuracies. Additionally, changes in the anatomical shape of fiber bundles, due to factors like brain tumors, hinder the effectiveness of standard atlases.

Clustering-based techniques in the context of WM tractography involve grouping individual fibers into coherent clusters by assigning labels to them (GARYFALLIDIS et al., 2012). These clusters serve as representations of the underlying WM structure, making the complex data more manageable. Fibers that either belong to small or insignificant clusters or do not share similar properties with the bundles of interest can be systematically removed. The key to employing standard clustering algorithms for streamline clustering lies in defining an

appropriate distance metric between the streamlines. While the concept of proposing distance metrics is relatively straightforward, the challenge arises in identifying the most suitable metric for streamlines, given the variety of WM structures and tractography data characteristics.

Recobundles tool is an atlas-based approach, where fiber clusters are registered in a simplified streamline bundle atlas and distant streamline are pruned (GARYFALLIDIS et al., 2018). However, the inherent limitations of atlas-based approaches can potentially impact the accuracy of such methods. Also, these methods rely on bundle similarity, consequently, only addressing certain tracts of interest.

On the other hand, direct methods aim to obtain the segmentation directly from the volumetric data, without involving intermediate processes. Notable examples of automated WM bundle segmentation, include TRACULA, which reconstructs 42 bundles using prior anatomical knowledge of brain structures (YENDIKI et al., 2011); AFQ, identifying 25 bundles through waypoint ROIs (YEATMAN et al., 2012); and TractSeg, a novel convolutional neural network-based approach directly segmenting 72 tracts using FOD peaks (WASSERTHAL et al., 2018).

The mentioned studies have highlighted the strengths of their approaches in evaluating brain disorders across different age groups. However, these tools have specific technical limitations. AFQ primarily analyzes central fiber tract data, potentially missing valuable marginal information and limiting quantitative measurements. TRACULA and TractSeg were trained on data from healthy young and middle-aged adults, requiring careful evaluation for accuracy in children and older patients. Also, the impact of data quality and acquisition variations on WM segmentation remains uncertain, as validation studies have typically used relatively small cross-sectional datasets (ANDICA et al; 2023).

Nevertheless, artificial intelligence (AI) can help to enhance automatic and direct segmentation methods used in medical imaging and neuroscience research. AI offers benefits such as automated feature extraction, reduced manual intervention, integration of multimodal data, handling variability across subjects and pathological conditions, enabling real-time processing, and efficiently scaling to process large datasets. These AI-driven strides can be used as a tool to elevate the accuracy and efficiency of WM segmentation, addressing some challenges in the field, such as the reproducibility of tractography segmentation.

1.3 ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DEEP LEARNING

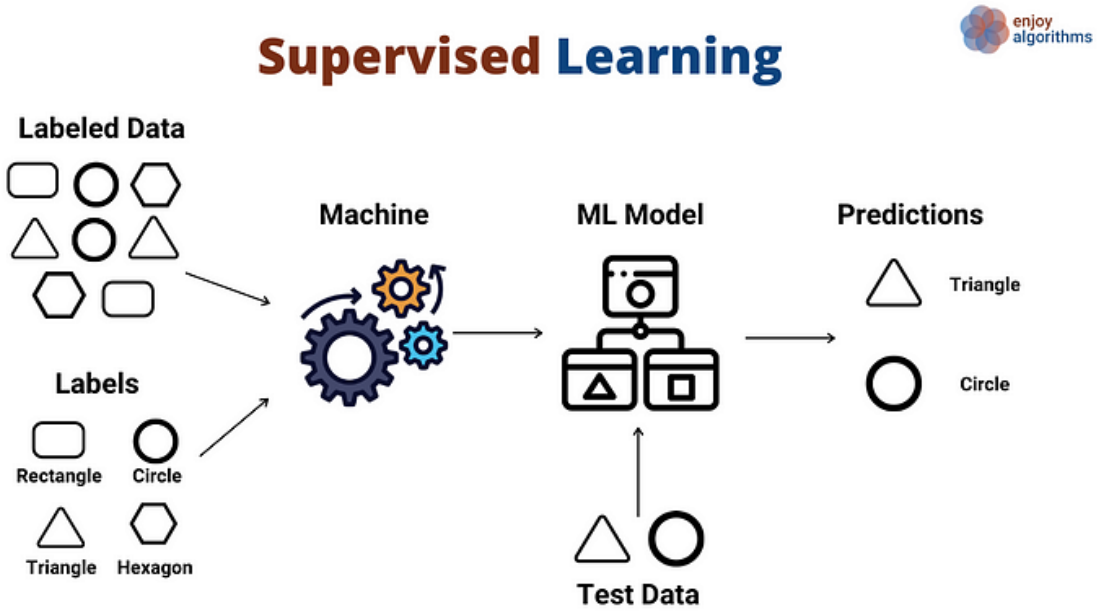
AI refers to the capabilities of the machine to perceive its environment and takes some actions in the tentative to succeed in specific objectives. Elaine Rich's definition succinctly captures this concept: "Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better" (ERTEL, 2017).

Planning, learning, natural language processing and perception are some central problems of AI research. For the machine be capable of problem solving and learning, such as human minds, approaches like statistical methods, mathematic optimization, logic, and probabilities are used. Thereby, AI encompasses any approach that empowers computers to mimic human behaviors or make decisions autonomously, tackling intricate tasks with minimal or no human intervention. This is reached by developing algorithms that iteratively learn from training data, where it can be found hidden insights and patterns without being explicitly programmed, given birth to ‘machine learning’ (ONGSULEE, 2017).

Machine Learning (ML) is a subfield of artificial intelligence. Basically, this is the result of pattern recognition studies and computational learning theory, in which algorithms are constructed in a way that they can learn and make predictions based on data exposition. ML methods encompass four primary categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Notably, supervised learning constitutes the majority of ML techniques, accounting for roughly 70 percent, with unsupervised learning comprising 10 to 20 percent (ONGSULEE, 2017; JANIESCH; ZSCHECH; HEINRICH, 2021).

Supervised learning algorithms are constituted by datasets with labeled examples, so given an input you already know the output desired (Figure 16). This knowledge enables the algorithm to learn by comparison. Through classification, regression, prediction or gradient boosting, the machine can search for patterns to predict the label on unseen data. This technique is normally used for data classification and prediction. Following the logic, unsupervised learning does not have historical labels, thus there is no template for the machine compare. For that situation, the purpose is to explore data and uncover inherent structures or characteristics, such as grouping similar data point.

Figure 16 – Example of supervised algorithms.



Source: Enjoy Algorithms.

In summary, we can divide the ML in three main layers: the input layer (comprising data for information extraction), the hidden layer (housing the learning algorithm itself), and the output layer (yielding predictions). These ML algorithms are often referred to as "shallow learning" due to their single hidden layers and linear processing units which constrain the learning.

If we deepen the ML algorithms, so they contain more than one hidden layer, we have the deep learning (DL). A key feature of DL is the stacking of numerous layers with nonlinear processing units, where the output of one-layer feeds into the next. This architecture allows for the extraction of multiple hierarchical levels of data features or representations.

In this way, for more complex tasks, such as object recognition and bioinformatics, deep learning super-pass the machine learning algorithms. For example, given a certain image a deep neural network can extract different characteristics of the image in each hidden layer. Using as input the image pixels, the first layer can identify the edges of the picture by comparing colors or brightness of the neighbors' pixels. In the second layer, the corners and contours can be detected using the edges description. More specific details that characterize certain object in the image can be found in the third layers. By the end, the objects in the input image can be recognized (HAO; ZHANG; MA, 2016).

Now, let's envision the initial layer as a composition of numerous processing units intricately connected to the subsequent layer via multiple connections. Each connection bears a weight, dictating its activation or deactivation, and the outcome of one layer becomes the

input for the next, repeating through several layers until a final output emerges. By renaming these processing units as neurons and the connections as synapses, we arrive at a simplified representation of the intricate network of connections within the brain (KROGH, 2008). Artificial Neural Networks (ANN) are one of the most successful methods of deep learning. Inspired by the biological model proposed by David H. Hubel and Torsten Wiesel in 1959, ANN can be conceptualized as cascading models akin to various types of brain cells (HUBEL; WIESEL, 1959). By employing algorithms that emulate the functioning of neurons, we endow the network with the capacity to "learn" and solve a wide array of problems, thereby forging a bridge between the human cognition and machine intelligence.

1.3.1 NEURAL NETWORS IN THE BRAIN AND IN THE MACHINE

The human brain is composed by approximately 100 billion nerve cells, creating intricated neural networks responsible for awareness, associations, thoughts, consciousness, and the ability to learn. A neuron is formed by dendrites, a cell body, an axon, and the terminal axon. The cell body of the neuron can store electrical signals, loaded by incoming electrical impulses from other neurons through dendrites. The more electrical signals come in, the higher the voltage, and if this voltage exceeds a certain threshold, the neuron will fire sending a spike over the axon and the synapses (ERTEL, 2017).

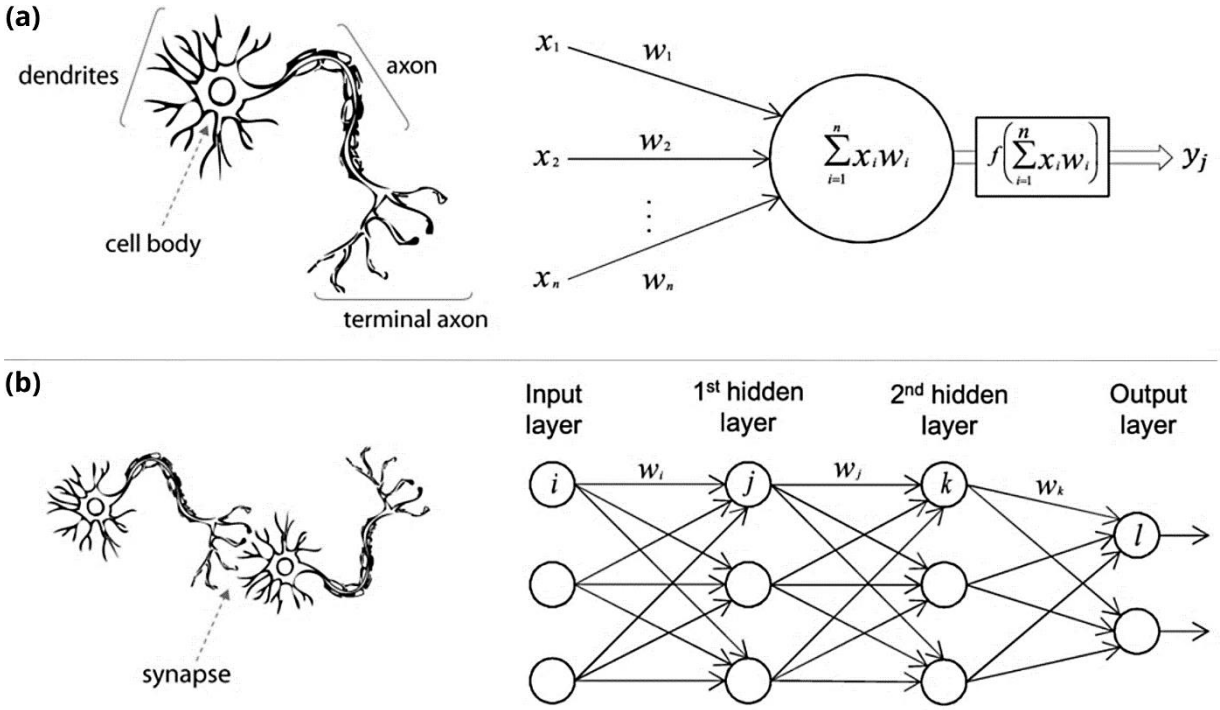
The synapses are responsible for create the connections between neurons. However, there is a small gap among the terminal axons of one neuron and the dendrites of the other one. This small gap is filled with neurotransmitters, that can be ionized when a voltage is applied and then transport the signal through the gap, but this conductivity depends on many factors. Therefore, are not the neurons who are adaptive, but the synapses by changing their conductivity. Intuitively, a synapse is made stronger proportional to the electrical current, than the ones that are used often obtain an increasingly higher weight. And the ones that are not has a decreased conductivity.

Mathematically speaking, this activation potential corresponds to the summation of the weighted output values (ω_{ji}) of all incoming connections (x_i). To see what information will pass to the other neuron (x_j), an activation function (f) is applied on the neighboring neurons as output over the synaptic weights (equation 14).

$$x_j = f\left(\sum_{i=1}^{n} \omega_{ji} x_i\right) \qquad (Eq.\ 14)$$

As we can see in Figure 17 – a, this expression represents one neuron and compose one processing unit in the model. In a neural network there are several hidden layers consisting of various neuron units (Figure 17 – b). Understanding that the adaptive process primarily occurs within the synapses, the learning aspect of the algorithm unfolds through the weighted connections. When a particular feature is deemed crucial for accurate predictions, its corresponding information is assigned a higher weight. For the model learn to cognize which connection is important for a specific task, we need to train or teach them (MENG; HU; ANCEY, 2020).

Figure 17 - (a) Representation of a biological neuron and the mathematical interpretation of them. (b) How this works for more than one neuron in brain and in the machine.



Source: (MENG; HU; ANCEY, 2020).

The learning agent is essentially defined as a function that translates a feature vector into either a discrete class value or a real number. Its primary objective is to discover an efficient data representation in the form of a function that optimally categorizes new data. This function is not predefined or hard-coded; instead, it evolves dynamically throughout the learning phase, adapting to the training data and are also nominated as cost-loss functions. During this learning process, the agent receives the output label, allowing it to fine-tune its mapping from the feature vector to the function value to achieve optimal classification (ERTEL, 2017; CURRIE et al., 2019).

Assessing the quality of this mapping relies on the concept of similarity, which hinges on minimizing the distance between the training samples and the model's predictions within the feature space. To approximate functions based on data points, various mathematical techniques come into play, including polynomial interpolation, spline interpolation, and the method of least squares. Using the least squares, the idea is to look for a function that minimizes the squared error given a pair of training vectors forming by the input $\boldsymbol{q}$ and the target $\boldsymbol{t}$ (equation 15).

$$E(\boldsymbol{\omega}) = \sum_{p=1}^{N} (\boldsymbol{\omega q}^p - \boldsymbol{t}^p)^2 = \sum_{p=1}^{N} \left(\sum_{i=1}^{n} \omega_i {q_i}^p - t^p \right)^2 \qquad (Eq. 15)$$

The gradient, as a vector of all partial derivatives of the error function, points in the direction of the strongest rise of the error function in the space of the weights. And to reach the minimum, we follow the direction of the negative gradient. So, the weights are going to change for each new training example following the equation 16.

$$\Delta\omega_j = -\frac{\eta}{2}\frac{\partial E}{\partial \omega_j} = -\eta \sum_{p=1}^{N} \left(\sum_{i=1}^{n} \omega_i {q_i}^p - t^p \right) {q_j}^p \qquad (Eq. 16)$$

Where η is the learning rate. By replacing the output neuron for applied example training we obtain the delta rule:

$$\Delta\omega_j = -\eta \sum_{p=1}^{N} (t^p - y^p)\, {q_j}^p \qquad (Eq. 17)$$

Thus, for every training example the difference between the target t^p and the actual output of the network y^p is calculated for the given input $\boldsymbol{q}^p$.

For more than two layers of neurons, the delta rule is expanded into the backpropagation algorithm, the most-used neural model for learning, due to their versatility for arbitrary approximation tasks. In this case, it is applied a nonlinear sigmoid function (equation 18) as activation function that allows the network to have more than two layers.

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (Eq. 18)$$

Using the sigmoid function as activation the weights are changed proportionally to the negative gradient of the quadratic error function summed over the output neurons for the training pattern p:

$$E_P(\omega) = \frac{1}{2}\sum_{k\epsilon output}\left(t_k^p - x_k^p\right)^2 \quad , \quad \Delta_p\omega_{ji} = -\eta\frac{\partial E_p}{\partial \omega_{ji}} \qquad (Eq.19)$$

Whitin the equation, the outputs of the neurons of the next deeper layer occur recursively. By multiple applications of the chain rule, we obtain the backpropagation learning rule:

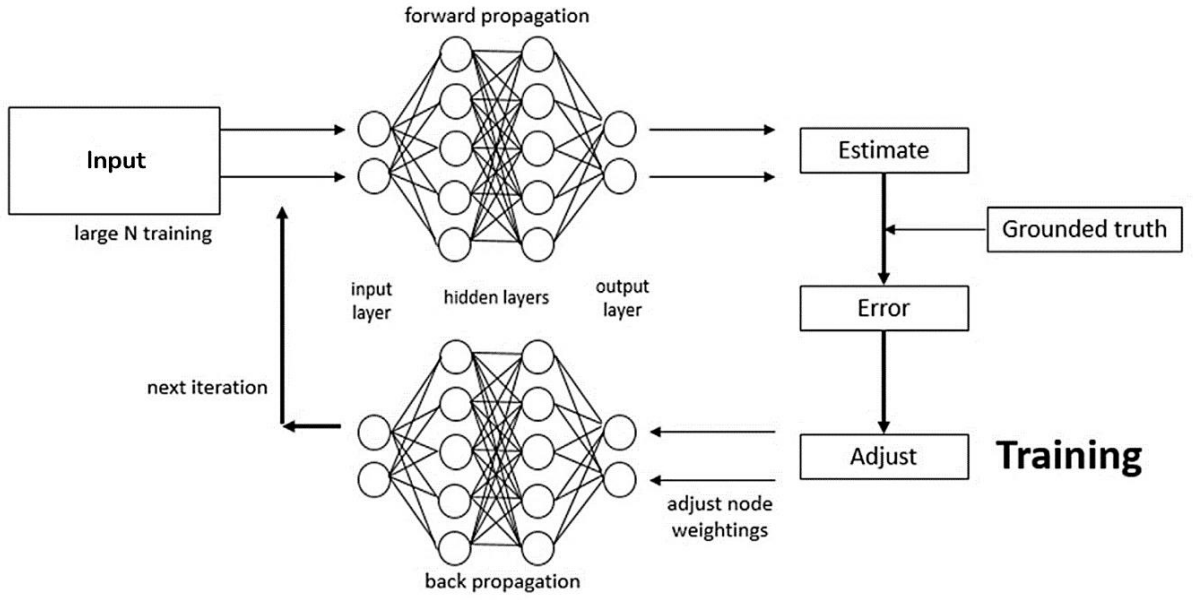
$$\Delta_p\omega_{ji} = \eta\delta_j^p x_i^p,$$

$$\delta_j^p = \begin{cases} x_j^p\left(1-x_j^p\right)\left(t_j^p - x_j^p\right) & if\ j\ is\ the\ output\ neuron \\ x_j^p\left(1-x_j^p\right)\sum_k \delta_k^p\,\omega_{kj} & if\ the\ j\ is\ a\ hidden\ neuron \end{cases} \qquad (Eq.20)$$

The factor $(1 - x_j^p)$ creates the symmetry, for the output neurons, the factor $(t_j^p - x_j^p)$ takes care of the weight change proportional to the error. For the hidden neurons, the value δ_j^p of neuron *j* is calculated recursively from all changes δ_k^p of the neurons of the next higher level.

After calculating the output of the network (forward propagation) for a training example, the approximation error is calculated. This error is the initial point for the backward propagation to alter the weights backward from layer to layer, more details can be found in (VOUTSADAKIS, 2014). The whole process in applied to all training examples and repeated until the weights do not change in a predefined value, or the time limit is reached. A representative scheme of the process in depicted in Figure 18.

Figure 18 - Schematic representation of the learning process in training with backpropagation algorithm.



Source: (CURRIE et al., 2019).

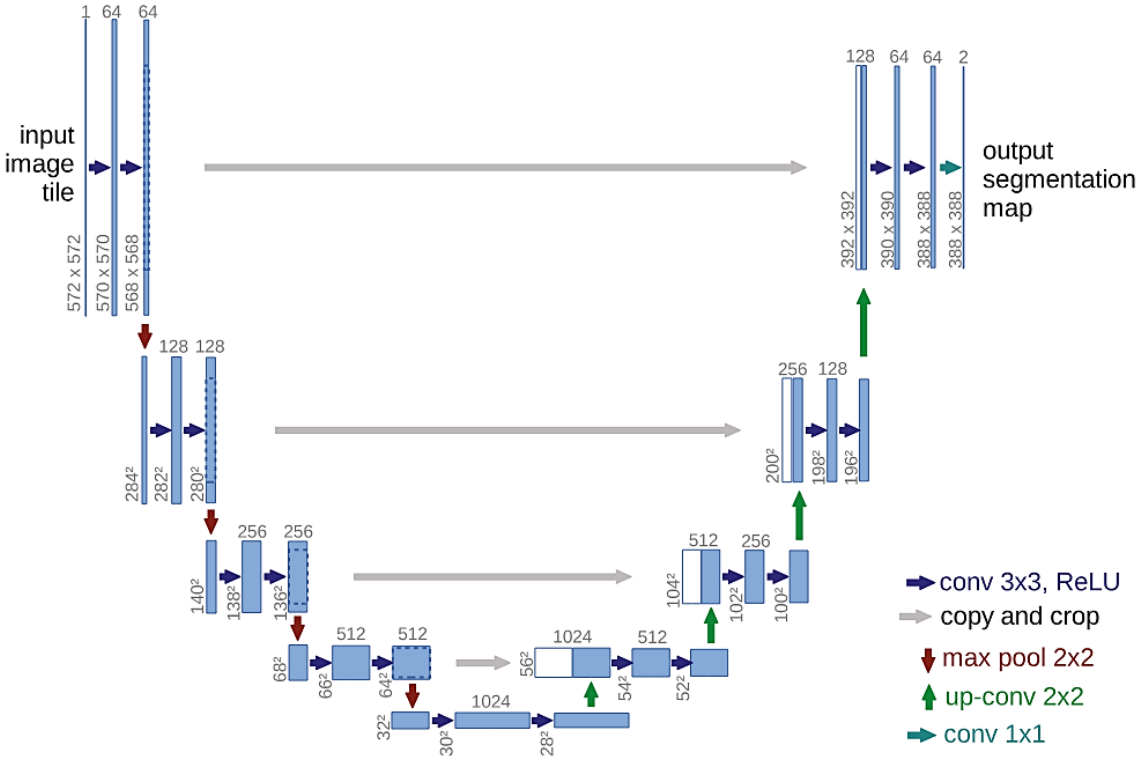
A variety of ANNs architecture have emerged over time, were each one becomes more suitable for specific data types, such as time series or images. The variation in architecture mostly remains in the types of layers, neural units, and connections they use. Examples are convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoder and generative adversarial neural networks (GANs) (JANIESCH; ZSCHECH; HEINRICH, 2021). The CNN is a feedforward neural network (transmit data in one direction) that can extract features from data using several convolutional layers. And its advantages of local connections, weighting sharing and downsampling dimension reduction, make this one of the most representative algorithms in the deep learning field.

The CNN architecture is inspired by the animal visual cortex, where individual neurons' responses to stimuli can be approximated through a convolution operation (HAO; ZHANG; MA, 2016). This operation plays a vital role by reducing the number of free parameters, enhancing model generalization, and making them particularly effective for tasks involving grid-like data such as images.

1.3.2 U-NET AS A CONVOLUTIONAL NEURAL NETWORK

In the field of medical image segmentation, one CNN highlights between all models available. The U-Net was introduced by Ronneberger and his collaborators in 2015 and has become widely adopted (RONNENBERGER; FISCHER; BROX, 2015). This network is characterized by its U-shaped design (Figure 19), which consists of a contracting path (encoder) and an expansive path (decoder).

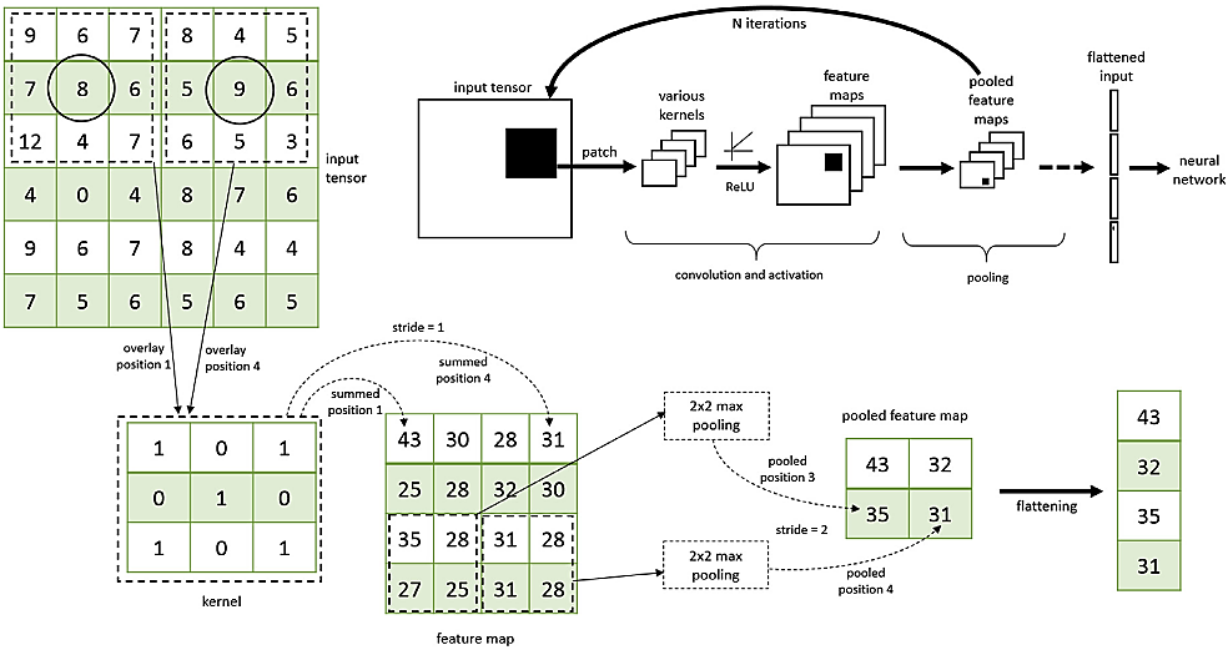
Figure 19 - Typical U-Net network architecture composed by a contracting path and an expansive path.



Source: (RONNENBERGER; FISCHER; BROX, 2015).

The contracting path takes on a standard CNN architecture, that is composed basically by three main types of layers: convolutional layer, pooling layer, and fully connected layer (Figure 20). The convolutional layer enhances the input features and reduces the noise by performing a scalar product calculation between the weights associated with each neuron and the region of the input volume connected to it. This layer focuses on learnable kernels, where each kernel convolves across the spatial dimensions of the input to generate 2D activation maps (O'SHEA; NASH, 2015).

Figure 20 - CNNs process. The input tensor is passed through a 3x3 kernel, producing several feature maps by the application of an activation function. After, the max pooling is used to reduce the size of the feature map that is finally flattened.



Source: (CURRIE et al., 2019).

The pooling layer subsamples the input data by dividing it into smaller regions and applying certain functions within each region. This operation helps reduce the number of parameters and dimensionality in the representation. By doing so, it effectively reduces the computational complexity of the model, making it more manageable and efficient.

The fully connected layer serves as the link between the information extracted in the previous layers and the final output, which is responsible for producing class scores based on the activations. This layer plays a crucial role in aggregating and processing the learned features to make decisions regarding classification or regression tasks.

More specifically, the contracting path of the U-Net is composed of the repeated application of two 3x3 convolutions together with the rectified linear unit (ReLU) activation function and followed by a 2x2 max pooling operation. These layers progressively reduce the spatial dimensions of the feature maps while increasing the number of feature channels, as represented in Figure 19. Thereby, the encoder concentrates its efforts on acquiring abstract, high-level representations of the input data.

Now, the different part came in the expansive path, involving the unsampling of the feature maps through 2x2 convolution, transposed convolutions, effectively halving the number of features. This result is then concatenated with the corresponding cropped feature map derived

from the contracting part. Following this concatenation, two consecutive 3x3 convolutions are applied, each accompanied by a ReLu activation function. The decoder aims to recover spatial information and fine details lost during the downsampling in the encoder path.

At the final layer, a 1x1 convolution is used to map each component feature to the desired number of classes in the original size of the image. The output of this layer represents the segmented mask, where each pixel is assigned a class label or probability score.

The U-Net architecture is particularly well-suited for medical image analysis tasks, such as organ segmentation, cell detection, and tumor identification, due to its ability to handle small datasets effectively and produce precise segmentation maps (YIN et al., 2022).

1.4 PROBLEM DEFINITION

Tractography is the computational representation of the white matter pathways obtained through diffusion magnetic resonance imaging (BASSER; JONES, 2002; JEURISSEN et al., 2019). To accurately study the brain structural connectivity with tractography, it is crucial to have precise tract segmentation strategies. White matter segmentation enables the identification and characterization of both healthy and abnormal brain microstructures in vivo, providing valuable insights for diagnosis, treatment, and surgical planning (ESSAYED et al., 2017; BHATIA et al., 2018; BASSEL et al., 2020; WENDE; HOFFMANN; MEIXENSBERGER, 2020; YANG et al., 2021). However, the reliability of this type of analysis is highly dependent on the accurate tract delineation, which demands time, anatomical, and imaging processing knowledge. This, in turn, dampens the prospects of reproducibility due to potential variations in delineation methodologies across different researchers. In response, automatic segmentation strategies have been actively investigated to address and enhance these constraints (COVER et al., 2018; ANDICA; KAMAGATA; AOKI, 2023).

Besides virtual dissection remains as the current gold standard, using a combination of inclusion and exclusion of ROIs, automatic approaches are being explored further to decrease time-consuming and variability across operators (NEHER et al., 2015; POULIN et al., 2019; SCHILLING et al., 2021; SINGH et al., 2022). These methods can be categorized in three types: ROI-based; clustering-based and direct segmentation. ROI-based approaches utilize anatomical priors to define ROIs that guide the segmentation of the fibers (ZHANG et al., 2020; WASSERMANN et al., 2010). Clustering-based techniques rely on grouping fibers into coherent clusters by assigning labels to them (GUEVARA et al., 2012; VÁZQUEZ et al., 2020). On the other hand, direct methods aim to obtain the segmentation directly from the volumetric

data, without involving intermediate processes (DYRBA et al., 2013; GONZÁLEZ-VILLÀ et al., 2016). While ROI-based methods are the most used, the advent of artificial intelligence has opened new possibilities for direct segmentation, offering improved reproducibility.

Deep learning methods have already demonstrated significant application in dMRI and great potential in tractography segmentation (POULIN et al., 2019; ZHANG et al., 2022; GHAZI; AARABI; SOLTANIAN-ZADEH, 2023). When it comes to automated bundle segmentation using deep learning, these approaches can be divided into two main groups based on the input data of the network: voxel-based and streamline-based. In voxel-based approaches, the segmentation is predicted using the orientation information derived from the fiber tracts (REISERT et al., 2018; LI et al., 2020). On the other hand, streamline-based approaches obtain bundle segmentation by utilizing pre-defined fiber features (LIN et al., 2019; ZHANG et al., 2020; PERETZKE, et al., 2023).

The U-Net network stands out as the most widely used and renowned architecture in CNNs for segmentation tasks (RONNENBERGER; FISCHER; BROX, 2015). Its unique capability to process entire imaging volumes and produce segmentation maps at the output has contributed to its popularity. When applied to the segmentation of specific tracts, this technique has shown exceptional results, especially when dealing with large tracts and high-quality DWI data (WASSERTHAL; NEHER; MAIER-HEIN, 2018; DONG et al., 2019; LIU et al., 2022, TCHETCHENIAN et al., 2023).

In the study conducted by Wasserthal et al. (2018), stacked 2D U-Net models were employed in a supervised manner to automatic segment tracts, considered the state of the art in direct white matter segmentation. Nevertheless, while the technique excelled in larger tracts, its performance with shorter tracts was less satisfactory. Additionally, the training dataset was sourced from high-quality data, which poses a limitation when attempting to generalize the model to clinical data marked by lower quality, thereby impacting the predictive accuracy of the model.

The success of automatic segmentation techniques has been particularly notable for larger tracts, such as the corticospinal tract and corpus callosum (WASSERTHAL et al., 2017; COVER et al., 2018; PERETZKE et al., 2023). However, shorter tracts, including the fornix and anterior commissure, have posed a challenge for neural networks, and have received less exploration in recent years. These shorter tracts are associated with various neurological diseases; for instance, Alzheimer's disease affects the uncinate fasciculus, hippocampus, and anterior commissure (SHU et al., 2013; MORI; AGGARWAL, 2014; KIUCHI et al., 2009),

while epilepsy is linked to abnormalities in the hippocampus, fornix, and uncinate fasciculus (DIEHL et al., 2008; CONCHA et al., 2010; KALAIVANI; SUNDARESWARAN, 2019). Moreover, the anterior and posterior commissures serve as essential landmarks in neurosurgery (PRAKASH; NOWINSKI, 2006; CHOI et al., 2013; LIU; DAWANT, 2015), further emphasizing the need to automate the segmentation of these shorter tracts for clinical applications.

While deep learning methods have shown impressive performance, there are several factors that complicate the generalizability of these strategies, especially when it comes to hospital acquisitions. Most existing studies in the field of white matter tract segmentation have primarily focused on the Human Connectome Project (HCP) dataset. When these methods have been applied to clinical acquisitions, they often utilize datasets from patients with specific neurological diseases or downgrade the high-quality data to match the clinical acquisition settings (DONG et al., 2019; LIU et al., 2022).

The quality of images obtained during routine hospital procedures tends to differ from high-quality images that constitute the majority of public datasets. This discrepancy poses a significant challenge, primarily due to time constraints imposed by clinical workflows. Moreover, the variability in image acquisition protocols, preprocessing pipelines, and brain anatomy adds further complexity, as these factors can greatly impact the resulting tractography (RHEAULT et al., 2020; SCHILLING, et al., 2021 - (b)). Consequently, addressing these challenges becomes essential to ensure the effectiveness and reliability of deep learning techniques in a clinical setting.

The main objective of this work was to evaluate the capability of the U-Net network in segmenting short tracts using DWI data acquired in different experimental conditions. Five short tracts were studied, including anterior, posterior, and hippocampal commissure, fornix, and uncinated fasciculus. The reference segmentation was obtained using a semi-automatic ROI-based method and applied to a cohort of 175 subjects from HCP, as well as an additional 175 subjects from a local hospital. This approach facilitated the training and assessment of the U-Net neural network using conventional clinical images, seeking for its implementation in the standard workflow of the HCFMRP image processing service.

Considering our previous comments, some key points can be reinforced:

- Although significant advancements have been made in automatic segmentation techniques, it remains crucial to test their effectiveness on clinical images. This is

because there exists a noticeable disparity between the high-quality images found in public datasets and the ones obtained during routine clinical procedures.

- Deep learning approaches have demonstrated successful segmentation results for large tracts. However, short tracts present a greater challenge and require further exploration and development.
- Factors such as the number of diffusion directions, b values, image resolution, and preprocessing pipelines can significantly influence the outcomes of tractography. These variables contribute to variations in the resulting tractography results and warrant careful consideration in the analysis process.

1.4.1 SPECIFIC OBJECTIVES

The objectives of this study were delineated into six distinct questions, all of which are comprehensively addressed within the pages of this dissertation. These questions serve as the guiding framework for the research and subsequent findings, providing a structured and systematic exploration of the research problem.

- To what extent can U-Net effectively segment short tracts characterized by substantial curvature while preserving segmentation accuracy? *Discussion – Paragraph 2*
- Can U-Net consistently deliver accurate predictions when applied to clinical data? *Discussion – Paragraph 4 and 7*
- How does the composition of the training dataset influence the performance of the testing set? *Discussion – Paragraph 5*
- What are the key findings regarding the transfer of a segmentation model trained on clinical data to high-quality data, and what do these findings reveal about domain shift? *Discussion – Paragraph 5 and 6*
- What is the significance of using both lower quality and higher quality data in training the neural network, and how does it impact segmentation performance? *Discussion – Paragraph 7*
- What strategies are recommended to improve the performance of segmentation models, particularly for short tracts, when dealing with diverse datasets in clinical applications? *Discussion – Paragraph 8*

This dissertation is structured to comprehensively address the specific objectives of our research. We initiated with an introductory chapter that establishes foundational concepts, beginning with an exploration of brain anatomy, with a special focus on the intricate white matter structure and its fiber connections. This chapter also delves into the principles of magnetic resonance imaging, emphasizing diffusion-weighted imaging, elucidating the acquisition process and techniques employed to model white matter fibers through tractography. Furthermore, we investigate the convergence of neural networks in both the human brain and computer systems, with particular attention to the U-Net convolutional neural network, which holds pivotal relevance to our study. We then culminate this chapter with a precise problem statement framing our research.

In the subsequent methodology chapter, we delve deeper into the intricacies of our research approach. This includes a comprehensive discussion of the dataset employed, the image pipeline processing techniques utilized, definitions of network parameters, and an elucidation of the training process along with the statistical analysis methodology.

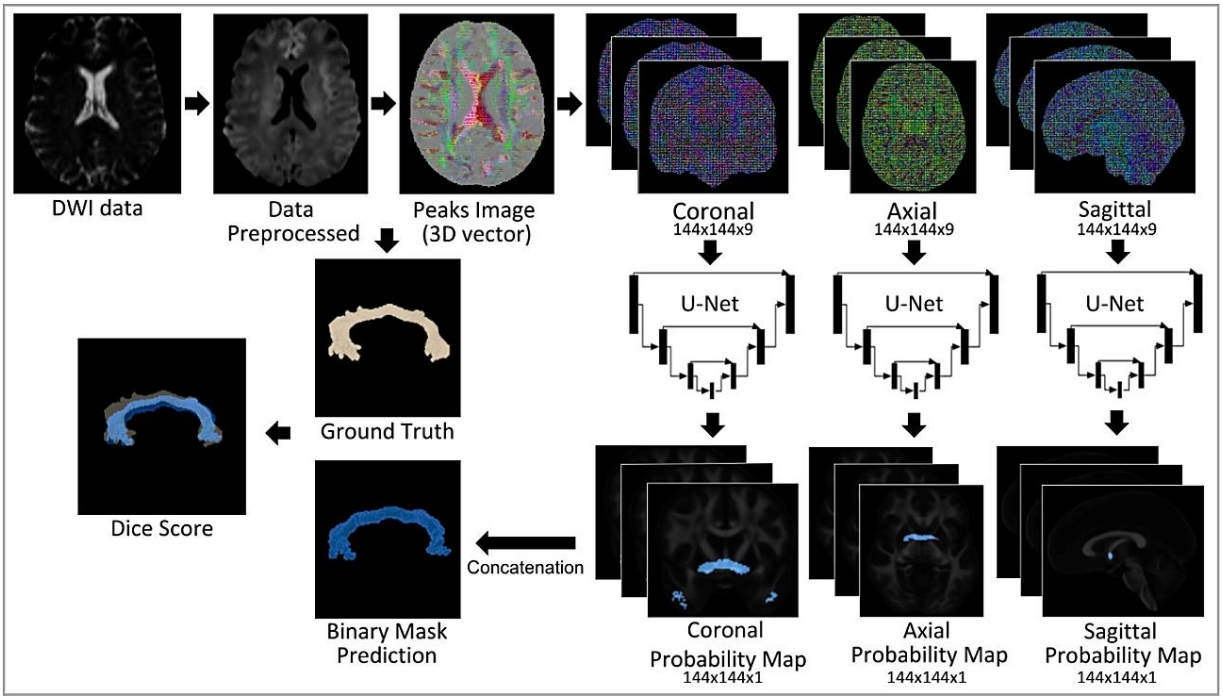
The results chapter serves as a focal point where we showcase the outcomes achieved through the execution of three distinct experiments. This chapter provides an extensive exploration of the U-Net model's performance in the segmentation of short tracts, as well as its potential clinical applications.

Finally, the dissertation concludes with a comprehensive discussion section, wherein we contextualize the findings in relation to the specific objectives outlined in our research. This section facilitates a thorough understanding of the implications of our work and leads to a substantial overall conclusion, summarizing the key contributions and insights gained from our research endeavors.

2 METHODOLOGY

The research involved distinct DWI acquisition methods for 350 subjects from HCP and local hospital datasets: HCP utilized 1.25 mm isotropic resolution with 270 gradients, while the hospital had 2 mm isotropic resolution with 32 gradients. The data preprocessing pipeline included essential motion and distortion corrections. Subsequently, the diffusion signal underwent fitting using the CSD model, enabling the extraction of posterior peaks for input into the neural network. Through a semi-automatic approach, we meticulously generated reference binary masks for seven shorter and four larger white matter tracts. We employed the U-Net network for generating probability maps based on fiber orientation distribution peaks, similar to Wasserthal et al. (2018). To assess the network's prowess in automating short tract segmentation, we conducted three distinct training experiments: exclusive utilization of HCP data, use of local hospital data only, and a fusion of both datasets. Performance evaluation was carried out using Dice score metrics, utilizing unseen data from both HCP and the local hospital dataset. The workflow depicted in Figure 21 offers a visual overview of the process.

Figure 21 - Methodology workflow. First, we preprocess the DWI data and extract the peaks imaging, representing the diffusion directions, using the CSD method. The 3D images are sliced in coronal, axial and sagittal views, and then used as input in the 2D U-Net network. The resulting probability maps obtained from each view are concatenated and merged by taking the mean along the last dimension to generate the final segmentation output. The result prediction is compared with the ground truth using Dice score.



Source: Elaborated by the Author.

2.1 IMAGING DATA AND PREPROCESSING

A total of 370 subjects were selected for the experiments, encompassing participants from HCP, PPMI (PARKINSON PROGRESSION MARKER INITIATIVE, 2011) and the Clinic Hospital of Ribeirao Preto, Brazil. DWIs acquired from the HCP dataset had an isotropic resolution of 1.25 mm. These images were acquired with 270 diffusion gradient directions, consisting of 3 b-values (1000, 2000, and 3000 s/mm²), accompanied by 18 b=0 images with acquisition duration around 60 minutes (Glasser et al., 2013). PPMI dataset is composed by DWIs with an isotropic resolution of 2 mm, with 64 diffusion gradient directions, 1 b-value (1000 s/mm²) and 1 b=0 image, with a total acquisition time of 13 minutes (PARKINSON PROGRESSION MARKER INITIATIVE, 2011). On the other hand, the DWIs from the local hospital dataset (CAAE - 08219712.7.0000.5407) had an isotropic resolution of 2 mm. They were acquired with 32 diffusion gradient directions, including 1 b-value (1000 s/mm²) and 1 b=0 image, with an acquisition time of 6 minutes and 30 seconds for healthy subjects and 10 minutes for epilepsy subjects (PINTO; SANTOS; SALMON, 2021). Table 1 and 2 provides a comprehensive comparison of the principal characteristics between the different datasets.

Table 1 - Dataset specifications for HCP and Local Hospital.

	Public Dataset Human Connectome Project (HCP)	**Local Hospital Dataset Clinical Hospital of Ribeirão Preto (HCRP)**
Subjects	175 (102 women)	175 (91 women)
Age	26 – 30 years	18 – 83 years
Images	DWI and T1w (3T)	DWI and T1w (3T)
	DWI Images	
Metric size	145x174x145 mm	128x128x72 mm
Voxel size	1.25x1.25x1.25 mm	2x2x2 mm
b values	1000, 2000, 3000 s/mm²	1000 s/mm²
Diffusion gradients	270 directions	32 directions
Scan acquisition time	60 min	6 min 30 sec

Source: Elaborated by the Author.

Table 2 - Dataset specifications for healthy subjects acquired in different hospital and epilepsy patients acquired in the local hospital.

	Healthy Subjects Different Hospital	**Epilepsy Patients Local Hospital**
Subjects	10 (5 women)	10 (4 women)
Age	57 – 72 years	26 – 66 years
Images	DWI and T1w (3T)	DWI and T1w (3T)
	DWI Images	
Metric size	116x116x72 mm	128x128x60 mm
Voxel size	2x2x2 mm	2x2x2 mm
b values	1000 s/mm²	1000 s/mm²
Diffusion gradients	64 directions	32 directions
Scan acquisition time	13 min	10 min

Source: Elaborated by the Author.

The HCP images came preprocessed using a minimal pipeline, which involved distortion correction, motion correction, and eddy current correction (GLASSER et al., 2013). For the images obtained from the PPMI dataset and the local hospital, the same pipeline using MRtrix3 (TOURNIER et al., 2019) together with FSL software (SMITH et al., 2004) was applied.

In MRtrix3 package we converted the raw bvecs and bvals files, in NIFTI format, of each subject to MRtrix3 image format (.mif) that combines this three information into one, consisting of a text header, with data stored in binary format within the same file. Data was pre-processed, correcting for noise (VERAART et al., 2016), image distortions (eddy current-induced distortion correction and motion correction, using FSL's tools) (ANDERSSON; SOTIROPOULOS, 2016) and bias correction (B1 field inhomogeneity correction) using ANTS (TUSTISON et al., 2010). This pipeline follows a similar procedure of a previous work in our lab (PINTO; SANTOS; SALMON, 2021).

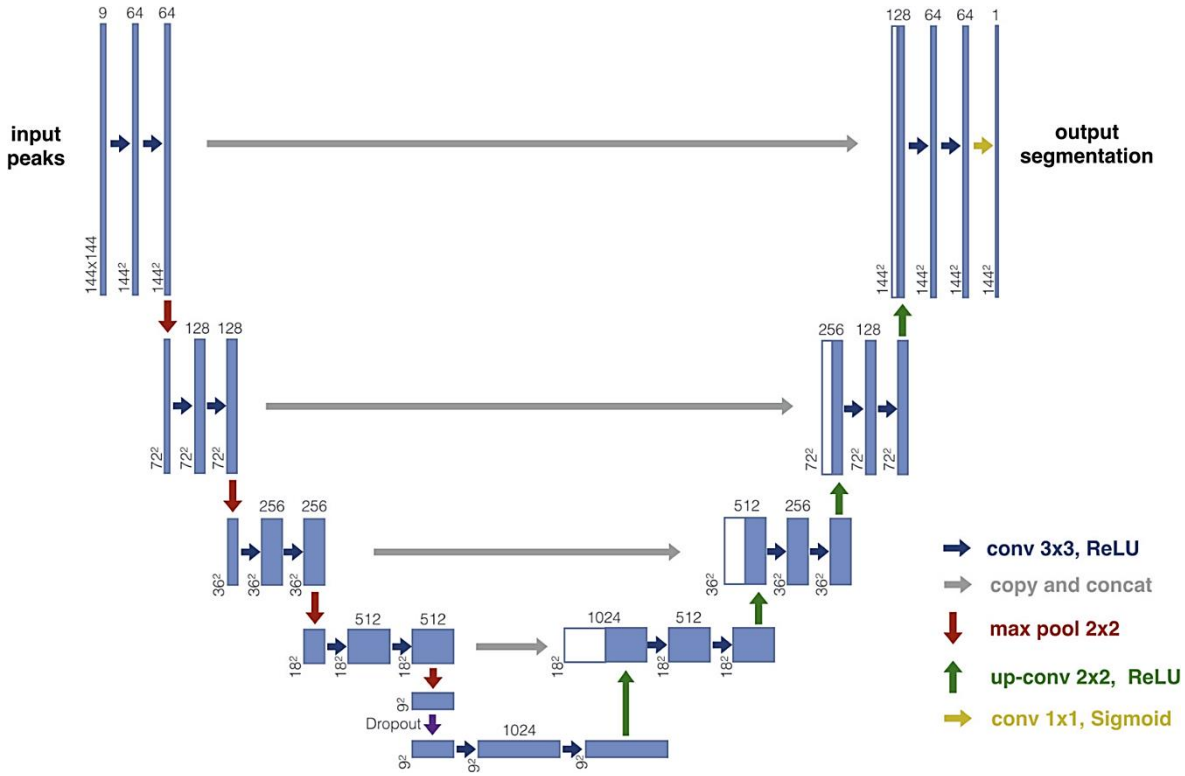
After that, the images were registered to MNI standard space using the MNI152 template (FONOV et al., 2009) using the anatomical images, resulting in a dimension of 182x218x182. Then CSD method (TOURNIER; CALAMANTE; CONNELLY, 2013) was employed to modelling the diffusion signal. And using the peak extraction function available in MRtrix3, the input image (peaks image) was obtained, in which each volume corresponds to the x, y & z component of each peak direction vector in turn. The images were then cropped to a size of 144x144x144 voxels, ensuring utmost care to avoid the removal of any brain tissue during the cropping process.

2.2 CONVOLUTIONAL NEURAL NETWORK

The utilized architecture for this study was the 2D U-Net, originally proposed by Wasserthal et al. (2018) depicted in Figure 22. In this case, the input to the network consisted of a 2D image with dimensions of 144x144 voxels and 9 channels representing the component of each peak direction vector (x, y, z) extracted from the SH coefficients. The initial 3x3 convolution operation generates 64 feature maps while preserving the original image dimensions, followed by the application of the ReLU activation function. Subsequently, the image undergoes downsampling through a max-pooling operation, resulting in a 72x72 image size with 64 channels. Another 3x3 convolution operation is employed, doubling the number of feature maps to 128. Afterward, downsampling reduces the image dimensions to 36x36. This repetition continues until we reach the bottleneck, where we attain a 9x9 image size, enriched with 1024 feature maps. In the expansive path, the opposite takes place. The image size expands

through 2x2 “up-convolutions” while concurrently reducing the number of feature maps. Additionally, the feature maps from the expansion phase are concatenated with their corresponding feature maps from the contracting path. The output is a single-channel map containing the voxel probabilities for a specific tract, with the same spatial resolution as the input.

Figure 22 - U-Net network architecture used for training.



Source: Adapted from (WASSERTHAL; NEHER; MAIER-HEIN, 2018).

In this way, the last layer of the U-Net utilized a sigmoid activation function instead of SoftMax. To keep the size of the output unchanged, SAME padding was used. To convert these probabilities into binary segmentations, a thresholding approach was applied. The threshold value was set to half of the maximum intensity of the prediction.

To overcome the computational and memory requirements of the 3D U-Net, the images were sliced into three different orientations: axial, coronal, and sagittal. Consequently, three separate networks were trained, with one dedicated to each orientation. The resulting probability maps from each network were then concatenated along the last dimension, resulting in an image with dimensions of 144x144x144x3, containing three predictions per voxel per tract. The final segmentation was obtained by taking the mean along the last dimension, as

demonstrated the best performance in literature (WASSERTHAL; NEHER; MAIER-HEIN, 2018).

Although a second training using the output of the first training as input was initially considered, a practical evaluation led to the decision to utilize only one training per view, as the original study suggests that the second neural network is optional (WASSERTHAL; NEHER; MAIER-HEIN, 2018).

2.3 TRAINING

The binary cross-entropy loss was used for training the network. ReLU activation functions were utilized for the intermediate layers, while a sigmoid activation function was applied to the last layer. To optimize the training process, a learning rate of 0.002 was chosen. The Adamax optimizer was utilized for parameter updates. Additionally, dropout regularization with a probability of 0.4 was implemented to prevent overfitting. In order to strike a balance between prediction accuracy and training time, the number of epochs was determined by conducting experiments. In an attempt to reduce the computational cost, the number of epochs and performance of the model were evaluated, with 130 epochs chosen as the optimal value. The validation dataset was utilized to fine-tune the hyperparameters of the network, ensuring improved performance.

We conducted three different types of training experiments. In the first experiment, we exclusively trained the model using 100 subjects for training and 15 for validation of the HCP dataset, which consists of high-quality data. This training involved employing multi-shell multi-tissue CSD and utilizing all gradient directions available in the dataset. To evaluate the performance of the trained model, we conducted tests on two separate sets of unseen subjects. The first test was performed on 60 unseen subjects from the HCP dataset, ensuring that the model was evaluated on data that it had not been trained on. Additionally, we performed a second test on 60 unseen subjects from the local hospital dataset, which presented different characteristics compared to the HCP dataset.

For the second experiment, we focused on the images obtained from the local hospital dataset, with a typical clinical routine acquisition. Equally 100 subjects were used for training and 15 for validation. These images had a b-value of 1000 s/mm², one b=0 image, and 32 diffusion gradient directions. To evaluate the performance of the trained model in this experiment, we used the same test dataset as in the first experiment.

In the third experiment, we employed a training approach that combined images from both the HCP and local hospital datasets. To ensure compatibility, we downgraded the HCP data to match the characteristics of the hospital dataset. This allowed for consistency in terms of data characteristics. Consequently, all training data had a b-value of 1000 s/mm², one b=0 image, and 32 diffusion gradient directions. To ensure diversity and representativeness in the training set, we randomly sampled the images from both datasets. In each experiment, three independent models were trained for each tract, resulting in a total of 99 models.

To evaluate the model's performance in the third experiment, we conducted tests on three separate sets of unseen subjects. The first test involved 60 subjects from the HCP dataset, also downgrading the resolution. The second test was performed on 60 subjects from the local hospital dataset. Lastly, we conducted a test on a mixed dataset consisting of 60 randomly combined subjects from both HCP and local hospital datasets considering 30 subjects from each dataset.

In the third experiment, we aimed to evaluate the generalizability of our approach. We expanded our predictions to include data from ten healthy subjects obtained from a different scanner, openly accessible from PPMI. Additionally, we incorporated data from ten epilepsy patients sourced from the local hospital (see Table 2). These two samples introduced variations in spatial resolutions and gradients, encompassing both healthy and diseased brains.

2.4 STATISTICAL ANALYSIS

To assess the training progress and evaluate the model's performance, we relied on the Dice score, a well-established metric in medical image segmentation (TAHA; HANBURY, 2015). This score offers a quantitative measure of the overlap between the predicted and ground truth segmentations, offering valuable insights into the accuracy of the segmentation results. By utilizing the Dice score, we were able to precisely quantify the model's performance and gauge the effectiveness of our segmentation approach.

The Dice score can be calculated using the following equation:

$$Dice = \frac{2 \times |A \cap B|}{|A| + |B|} \quad (Eq. 21)$$

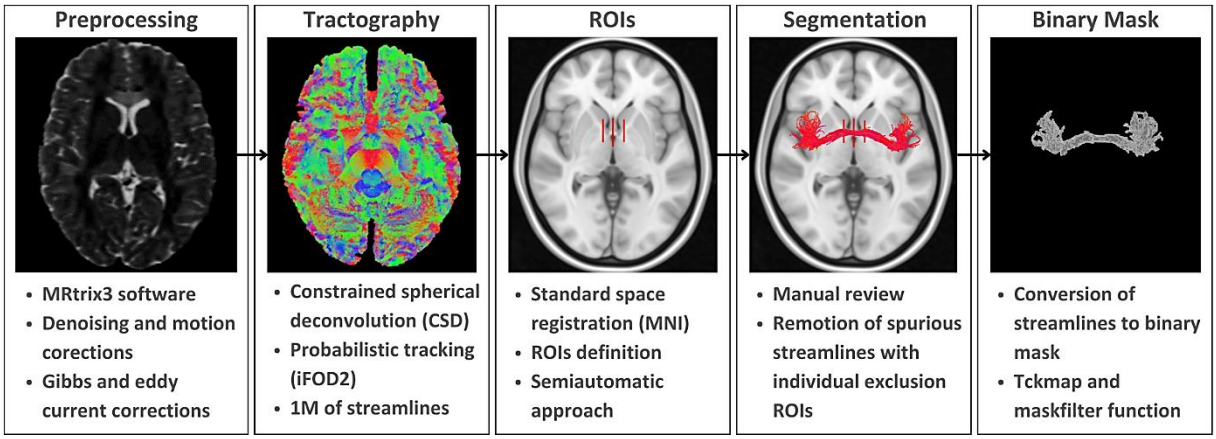
Where $|A \cap B|$ represents the intersection of the predicted and ground truth segmentations, $|A|$ the number of pixels in the predicted segmentation and $|B|$ the number of pixels in the ground truth segmentation. This formula yields a score value between 0 and 1, where 1 indicates

perfect overlap between the predicted and ground truth segmentations, and lower values indicate less accurate segmentation.

2.5 REFERENCE SEGMENTATIONS

To generate the reference binary masks first we preprocessed the DWI images using MRtrix3 (TOURNIER et al., 2019). Fiber orientation distribution was extracted via multi-shell multi-tissue CSD for the HCP dataset and standard CSD for the local hospital dataset. Subsequent steps involved probabilistic tractography using the iFOD2 algorithm, followed by image registration to a standardized space to delineate exclusion and inclusion ROIs for specific tract segmentation. To refine segmentation accuracy, a thorough manual inspection was conducted. The conversion of streamlines into binary masks was achieved through the utilization of two MRtrix3 functions – tckmap and maskfilter –, as visually depicted in Figure 23.

Figure 23 - Example of anterior commissure tract-specific binary mask generation workflow.



Source: Elaborated by the Author.

For each subject, we generated reference binary masks encompassing seven short white matter tracts: UF and FX for both hemispheres, as well as AC, PC, and HC. In addition, we included four larger tracts, corticospinal tract (CST) and inferior fronto-occipital fasciculus (IFO/IFOF) for both the left and right sides, to facilitate comparison with existing literature and validate our work. These reference binary masks served as ground truth labels, enabling thorough evaluation and validation of our segmentation approach.

2.5.1 TRACTOGRAPHY GENERATION

To extract the fODF, a multi-shell multi-tissue CSD method was utilized for the HCP dataset. On the other hand, for the local hospital dataset, which only consisted of one b-value shell, a standard-CSD method was employed. Subsequently, anatomically constrained probabilistic tractography using iFOD2 algorithm (TOURNIER et al., 2010) was performed for the entire brain. This process generated a total of 1 million fibers. For seeding, random placement of seeds within the brain mask was utilized. The streamlines were then cropped at the gray matter-white matter interface to ensure accurate delineation (SMITH et al, 2012).

2.5.2 ROIS DEFINITION

To segment the whole brain tractography into specific tracts, we employed a semi-automatic approach by defining exclusion and inclusion regions of interest. The ROIs were defined in the MNI space to ensure consistency and reproducibility across subject segmentations, following the methods outlined by Stieltjes et al. 2013 and Pinto et al., 2021. Table 3 and Figures 24-30 summarize the criteria used for ROI definition in each tract selection. Inclusion ROIs are highlighted in green, while exclusion ROIs are denoted in red. The segmentation process was made by a technician with over 3 years of experience in anatomical image segmentation.

For tracts where inclusion and exclusion ROIs were not explicitly defined in the literature, such as the posterior commissure and hippocampal commissure, we relied on anatomical descriptions as a reference. This allowed us to establish a basis for accurately segmenting these tracts.

Table 3 - Inclusion and Exclusion criteria of ROIs for each tract selection.

Tract	Inclusion Criteria	Exclusion Criteria
AC	1. In the mid-sagittal plane, positioned below the corpus callosum and above the cerebral peduncle, specifically within the anterior section of the third ventricle. 2. Place identical ROIs in the center of both the left hemisphere and the right hemisphere.	1. Reside within the temporal-frontal division but deviate from the expected trajectory of the AC tract. 2. Extend towards regions beyond the anterior-to-middle section of the brain, including the occipital part.

Continue

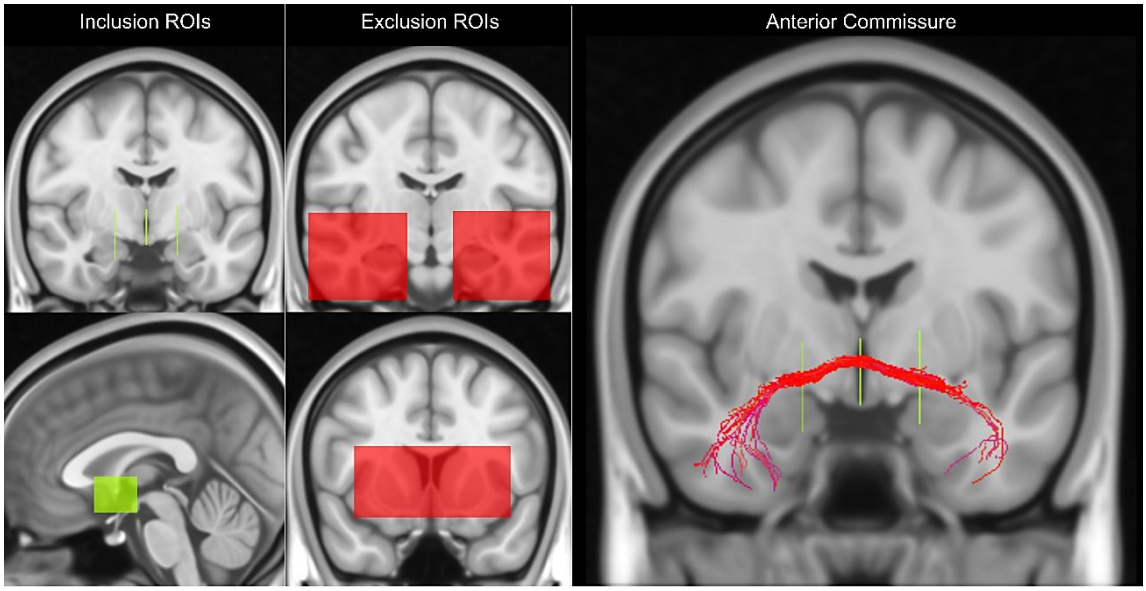
HC	1. Utilize the anatomical landmarks of the end of the corpus callosum and the fornix to precisely define the regions. Place one ROI on the left and another on the right side, ensuring they align within the sagittal plane. 2. In the coronal plane, position the ROIs above the third ventricle to accurately delineate the hippocampal commissure regions.	1. Fibers that are from corpus callosum and fornix. 2. Any fibers extending into the occipital lobe.
PC	1. The posterior commissure is positioned posterior to the anterior commissure. 2. The PC is situated near or slightly anterior to the pineal gland within the third ventricle and can be visualized better in the sagittal plane.	1. Fibers that extend toward regions outside the expected anatomical boundaries of the PC, including the cerebral peduncle, parietal lobe, or cerebellum.
FX	1. In the coronal slice select the body of fornix. 2. In the axial slice, select the region corresponding to the hippocampal fimbria, aligning it with the level of the cerebral peduncles and covering the hippocampus.	1. Cross between the hemispheres of the brain. 2. Fibers from hippocampal commissure.
UF	1. Select the most posterior coronal slice where the temporal and frontal lobes distinctly separate, usually found at the anterior portion of the fornix in the mid-sagittal view. 2. For each hemisphere, designate ROIs within the temporal and frontal lobes.	1. Fibers from corpus callosum. 2. Reside within the temporal-frontal division but are not considered part of the UF tract.

Continue

CST	1. Cerebral peduncle in the midbrain area. 2. Primary motor cortex or precentral gyrus.	1. Traverse between the hemispheres of the brain. 2. Reside near the cerebral peduncle but are not considered part of the CST tract.
IFO	1. In the coronal slice extending into the occipital regions, designate both ROIs for each hemisphere. 2. In the coronal slice where the tract gathers or bunches, identify the ROI located at the posterior edge of the genu of the corpus callosum.	1. Extend towards parietal regions. 2. Belonging to corpus callosum 3. Are not considered part of the IFO tract.

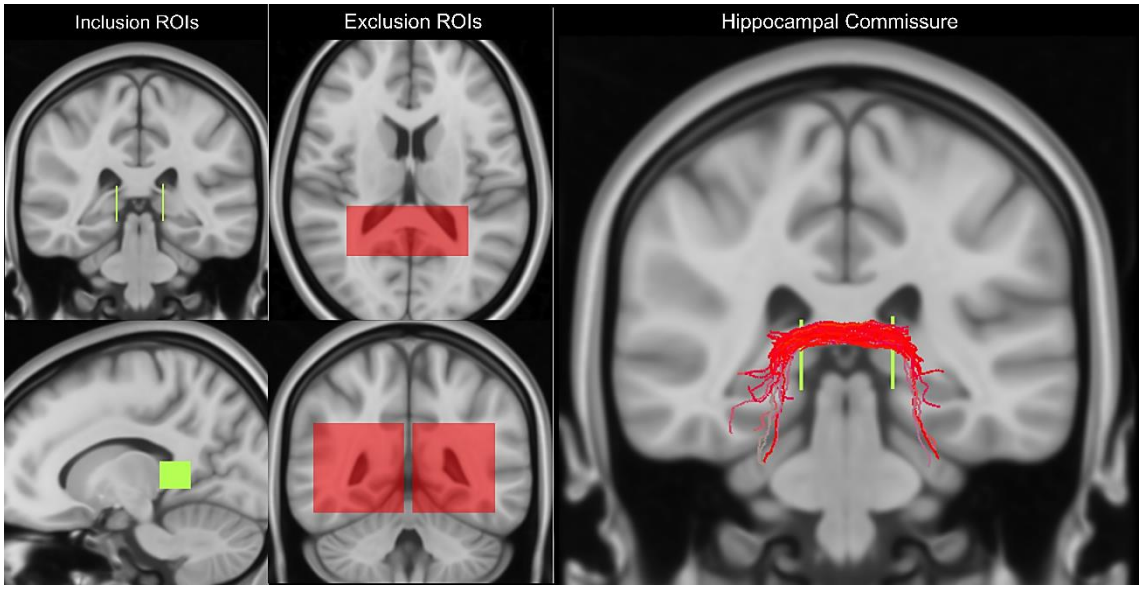
Source: Elaborated by the Author.

Figure 24 - ROIs for Anterior Commissure segmentation.



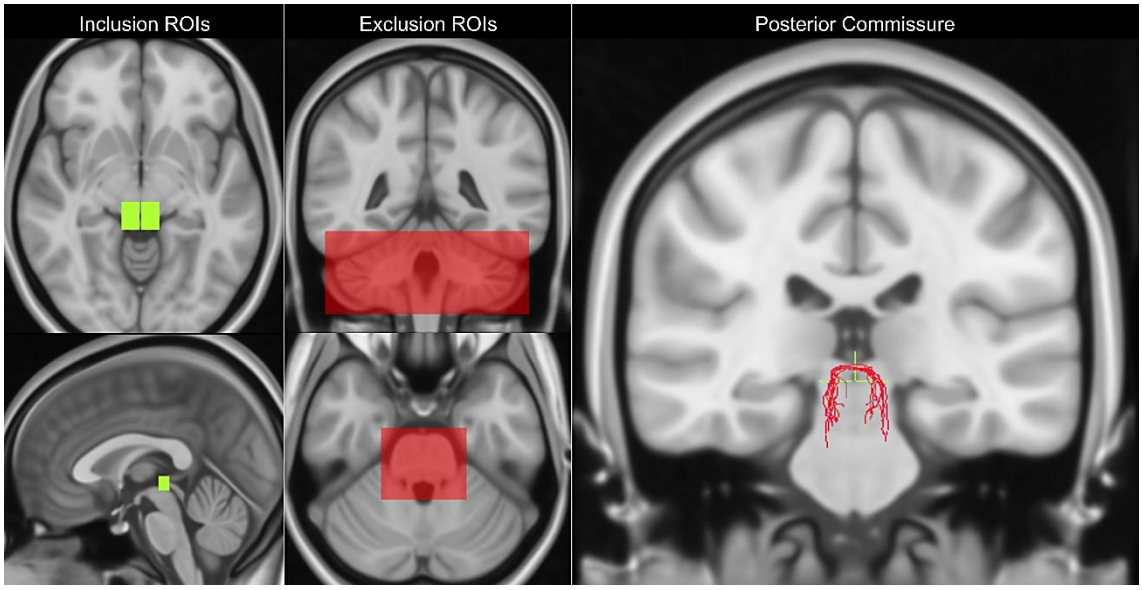
Source: Elaborated by the Author.

Figure 25 - ROIs for Hippocampal Commissure segmentation.



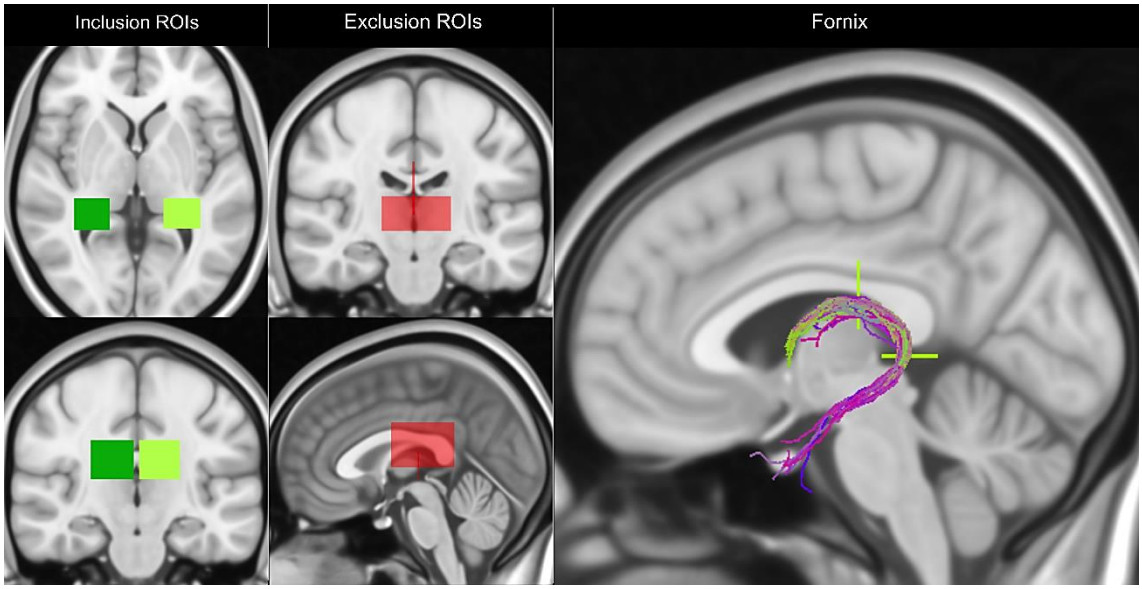
Source: Elaborated by the Author.

Figure 26 - ROIs for Posterior Commissure segmentation.



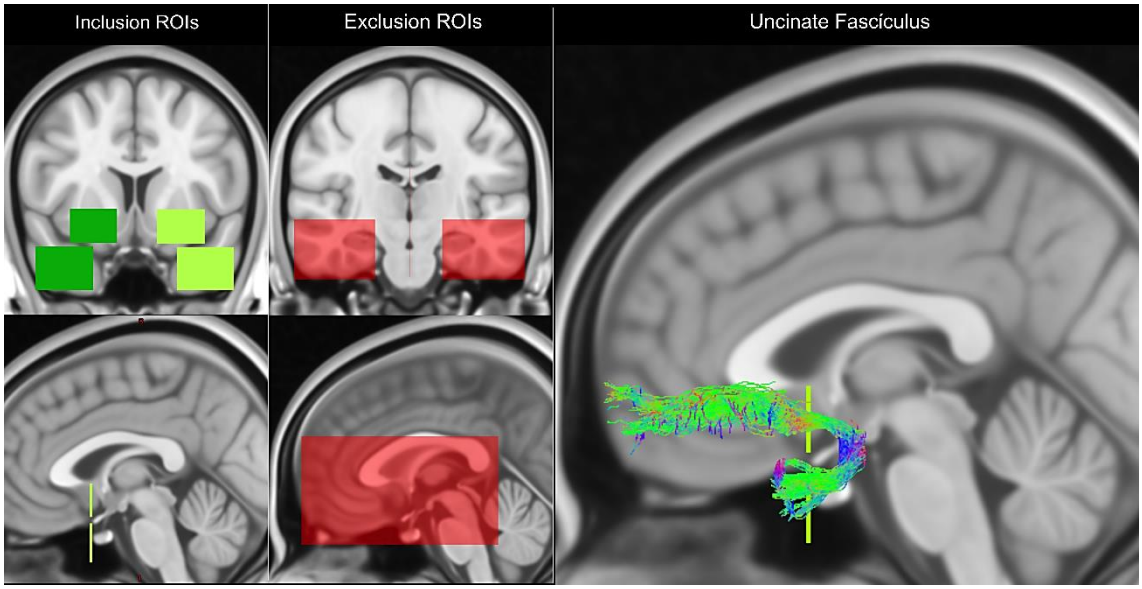
Source: Elaborated by the Author.

Figure 27 - ROIs for Fornix segmentation.



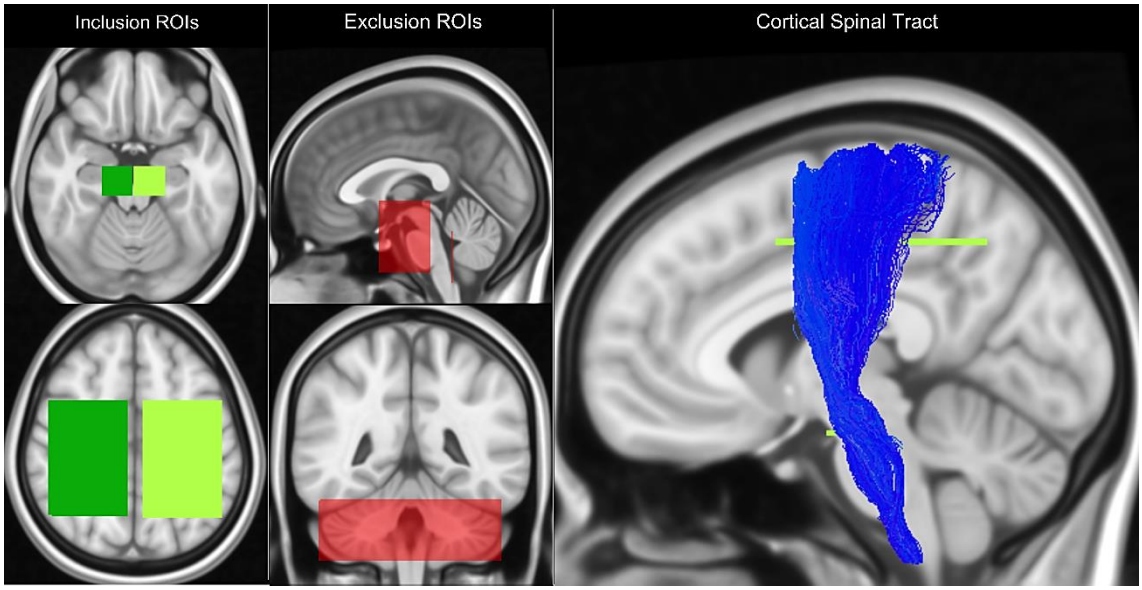
Source: Elaborated by the Author.

Figure 28 - ROIs for Uncinate Fasciculus segmentation.



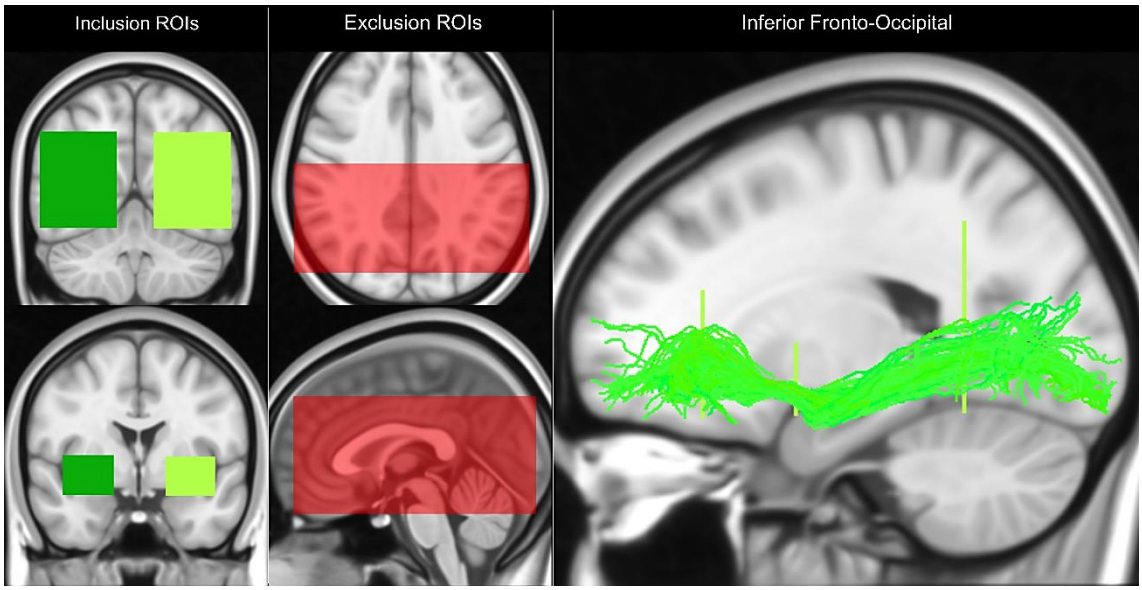
Source: Elaborated by the Author.

Figure 29 - ROIs for Corticospinal Tract segmentation.



Source: Elaborated by the Author.

Figure 30 - ROIs for Inferior Fronto-Occipital segmentation.

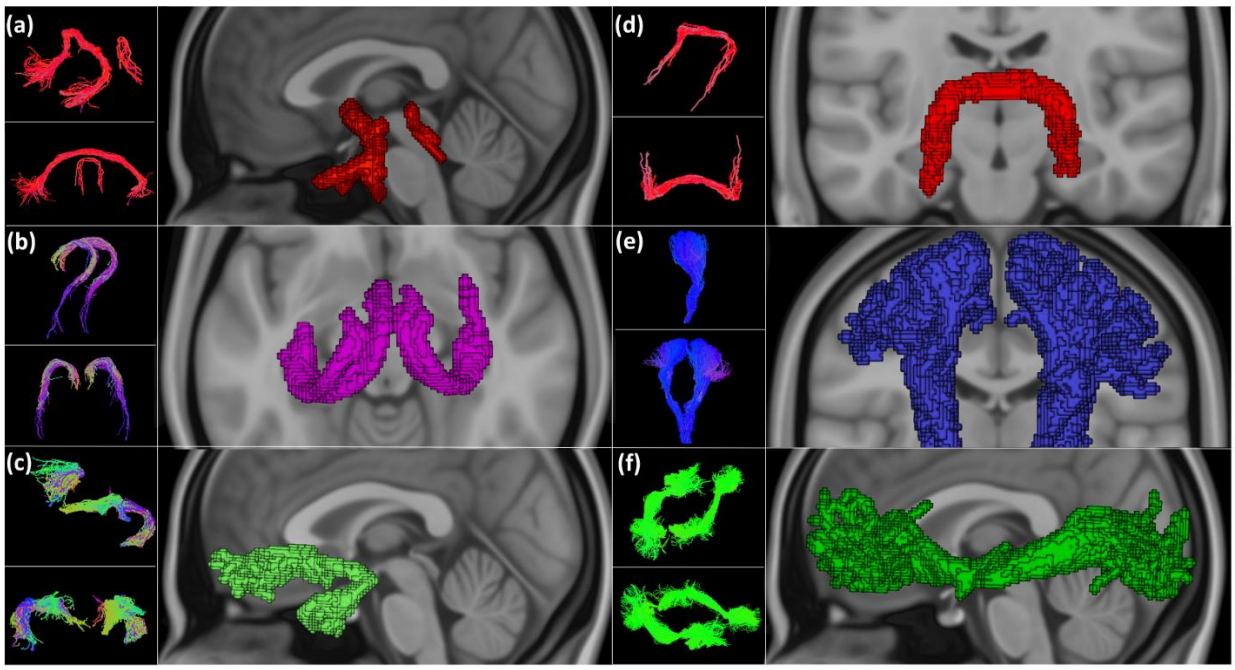


Source: Elaborated by the Author.

2.5.3 MASK GENERATION

To generate binary tract masks from the final sets of streamlines, we utilized the tckmap function in conjunction with the maskfilter function available in MRtrix3 (Figure 31). This combination allowed us to achieve a smoother and more refined result. These binary masks now serve as reference segmentations and are used for training and testing of the proposed segmentation approach. Utilizing these masks, we ensure a reliable and standardized basis for evaluating the performance of the segmentation method.

Figure 31 - Specific tracts visualization: anterior and posterior commissure (a), fornix (b), uncinate fasciculus (c), hippocampal commissure (d), corticospinal tract (e), and inferior fronto-occipital fasciculus (f).



Source: Elaborated by the Author.

3 RESULTS

By exploring three different training scenarios, we aimed to evaluate the performance and generalization capabilities of our segmentation approach under varying data conditions, including both high-quality and routine clinical acquisition settings. For that, we separate the results into qualitative and quantitative to demonstrate better the impacts of the different conditions in public and local hospital datasets. Moreover, we extended our investigation by subjecting the final experiment to a separate dataset originating from an alternative hospital routine and epilepsy patients, to see the transferability of the training networks to different scanners and non-healthy brains.

3.1 MANUAL QUALITY CONTROL AND CLEAN-UP

Following the semi-automatic approach, we conducted a thorough manual inspection of the individual results. Any remaining spurious streamlines were carefully addressed by implementing new exclusion ROIs (Figure 32). Given the need for precise and accurate placement, the exclusion ROIs had to be manually drawn for each individual subject rather than employing the same ROIs across all subjects. The percentages of subjects that needed manual additional exclusion ROIs are exhibit in Table 4.

Table 4 - Percentages of subjects with additional exclusion ROIs to remove spurious streamlines in manual quality control.

Tract	*Public Dataset*	*Local Hospital Dataset*
AC	31,3%	56,3%
HC	57,4%	73,4%
PC	51,9%	81,1%
FX L	75,4%	80,6%
FX R	73,1%	75,7%
UF L	42,2%	70,3%
UF R	36,3%	51,7%
CST L	68,0%	77,1%
CST R	56,6%	73,1%
IFO L	26,3%	29,1%

Continue

Tract	*Public Dataset*	*Local Hospital Dataset*
IFO R	22,3%	25,7%

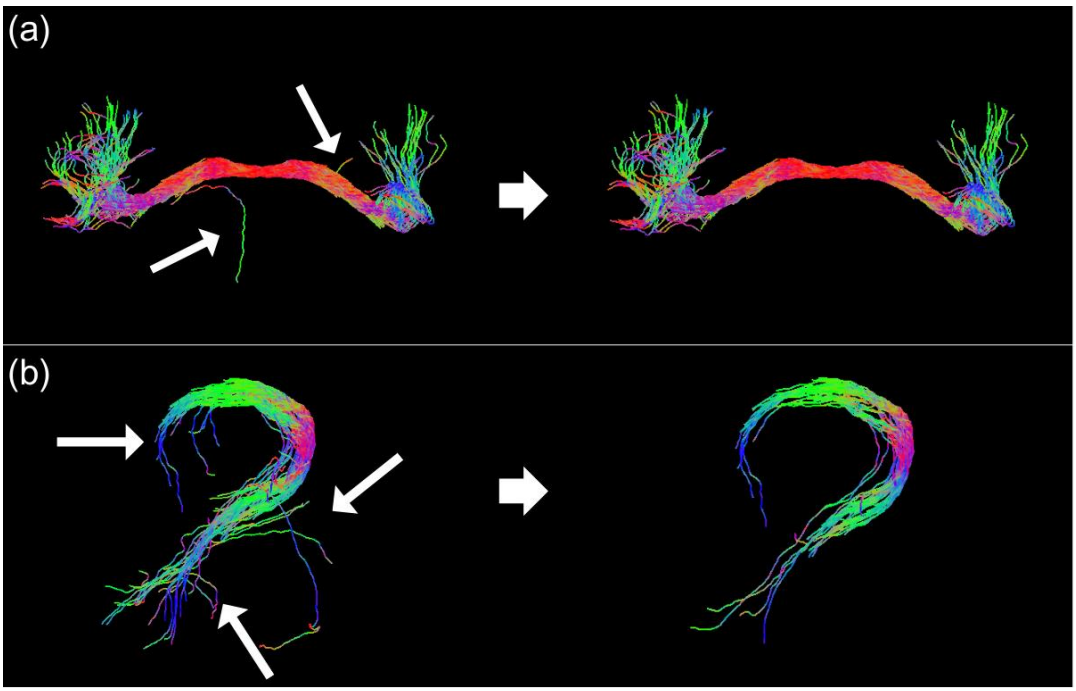
Source: Elaborated by the Author.

The higher percentages of additional corrections in all tracts using the images of the local hospital dataset detach the impact of quality in tractography segmentation. This outcome is partially a consequence of the registration process, which introduces deformations in the images to align them within a common spatial framework. Images with lower quality have more problems with this approach, bringing more necessity of manual quality controls in segmentation.

Nevertheless, certain tracts like the CST, PC, FX, and HC present unique difficulties due to their specific anatomical locations and structures. The CST, being a lengthy tract, presents complexities, especially in accurately segmenting its lateral projections near the cortex. The proximity of the PC to the CST path further complicates matters as these fibers, along with those extending towards the cerebellum, tend to persist even when using exclusion ROIs to sever these connections. The fornix, characterized by a high curvature level and its central position within the brain, presents challenges in separating its fibers from those belonging to neighboring tracts. HC faces a similar dilemma, as it passes through the corpus callosum and resides at the crura of the fornix, making it particularly challenging to isolate it.

In this way, the manual checkup ensures the effective removal of any undesired streamlines and reinforced the necessity of additional steps besides the automatic approach to improving the overall quality of the segmentation results.

Figure 32 - Example of the necessity of manual checkup for an adequate tract segmentation. (a) Anterior Commissure. (b) Left Fornix.

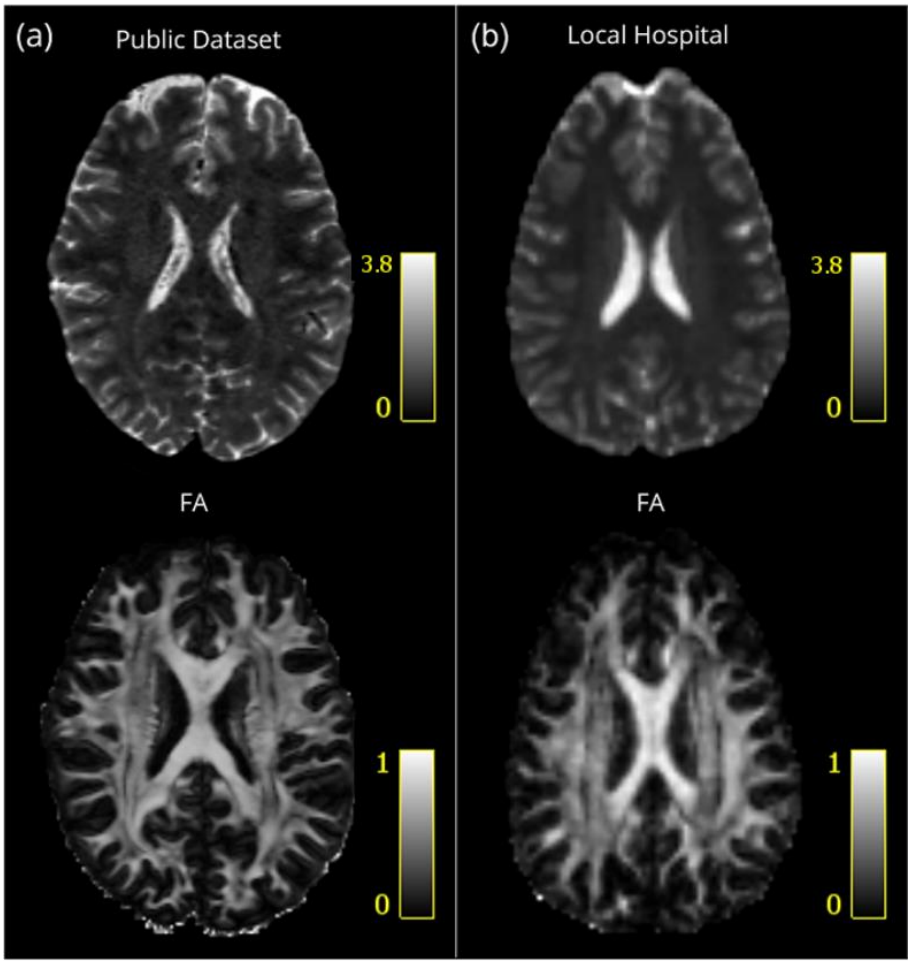


Source: Elaborated by the Author.

3.2 QUALITATIVE RESULTS

In our qualitative assessment, we carefully selected a representative subject from both the test set of the public dataset and the local hospital dataset, ensuring their Dice scores were closest to the mean certifying the chosen subjects served as representative samples for their respective datasets. Although, before delving into this analysis, we provide an illustrative depiction of the effect of different experimental conditions (see Figure 33). This visual aid aims to facilitate a deeper understanding of how varying conditions can significantly affect the quality of information within the images.

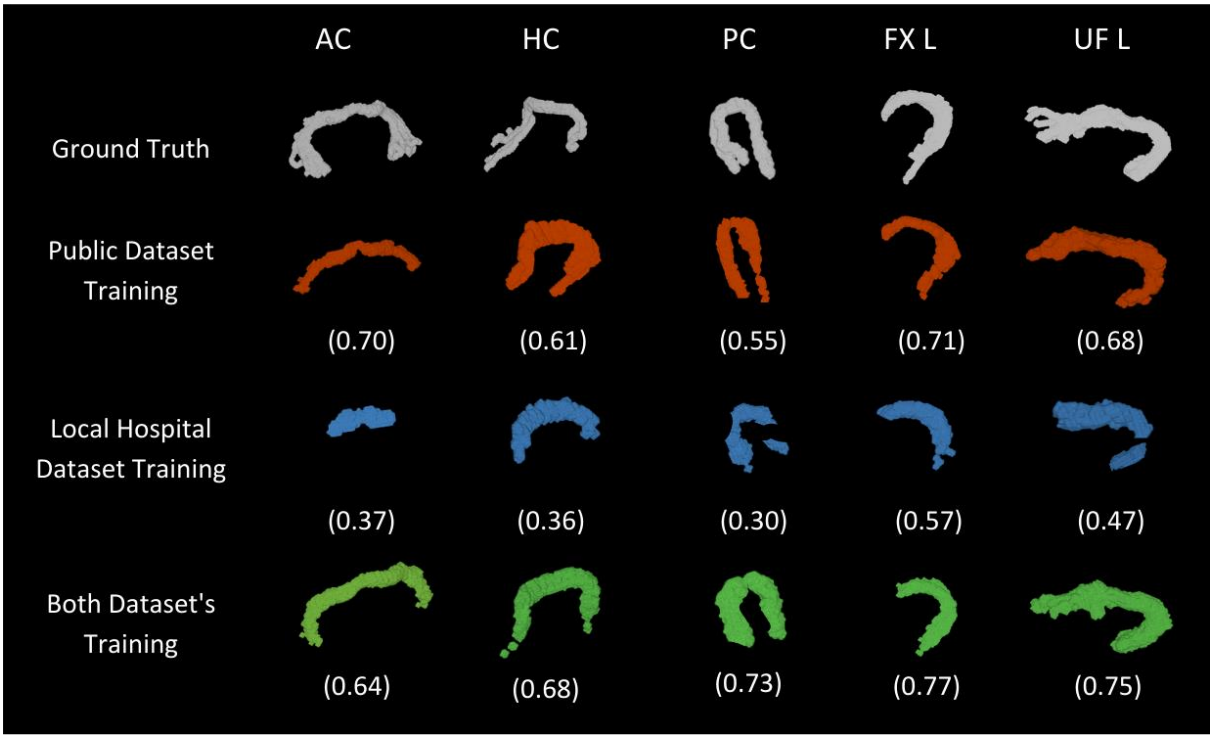
Figure 33 - Comparison between two images acquired in two experimental conditions (above) and arising FA maps (bellow).



Source: Elaborated by the Author.

As depicted in Figure 34, we presented the 5 short tracts explored, displaying the left-side representation of those existing in both hemispheres, and the outcomes from the first experiment exhibited a commendable level of satisfaction. However, discernible challenges arose when the network was trained on a different dataset, as evident in the second experiment. The most visually compelling tract reconstructions emerged from the training involving both datasets, demonstrating enhanced fidelity when compared with ground truth.

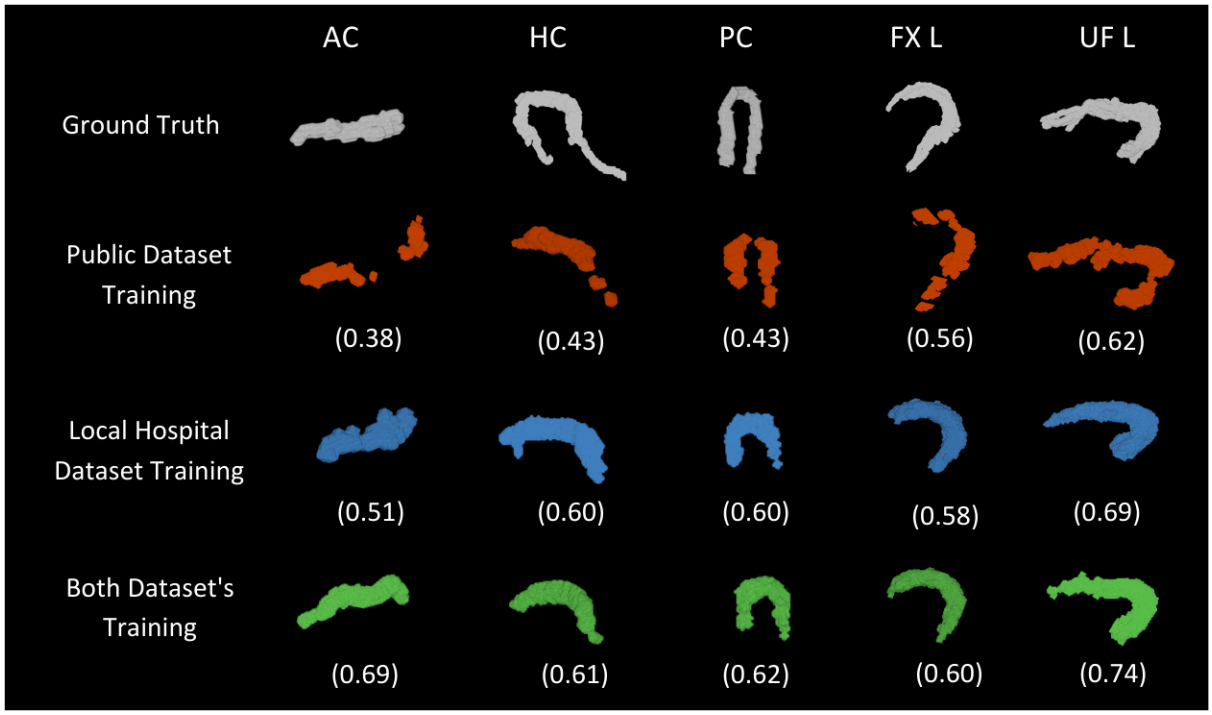
Figure 34 - Qualitative results of one random subject from the public dataset for the first experiment (public dataset training), second experiment (local hospital dataset training) and third experiment (training with both datasets) (Dice Score).



Source: Elaborated by the Author.

The same aspects can be visualized for one subject with data acquired at the local hospital (Figure 35). Here, the tract reconstruction faced difficulties when exclusively trained on the public dataset, with the third experiment yielding the most favorable outcomes. Notably, short tracts with higher curvature levels exhibited greater difficulty in learning and maintaining continuity in predictions, as exemplified by the posterior commissure, fornix, and uncinate fasciculus. These tracts exhibited gaps in both boundary and curvature sections, underlining the intricacies of their representation.

Figure 35 - Qualitative results of one random subject from the local hospital dataset for the first experiment (public dataset training), second experiment (local hospital dataset training) and third experiment (training with both datasets) (Dice Score).



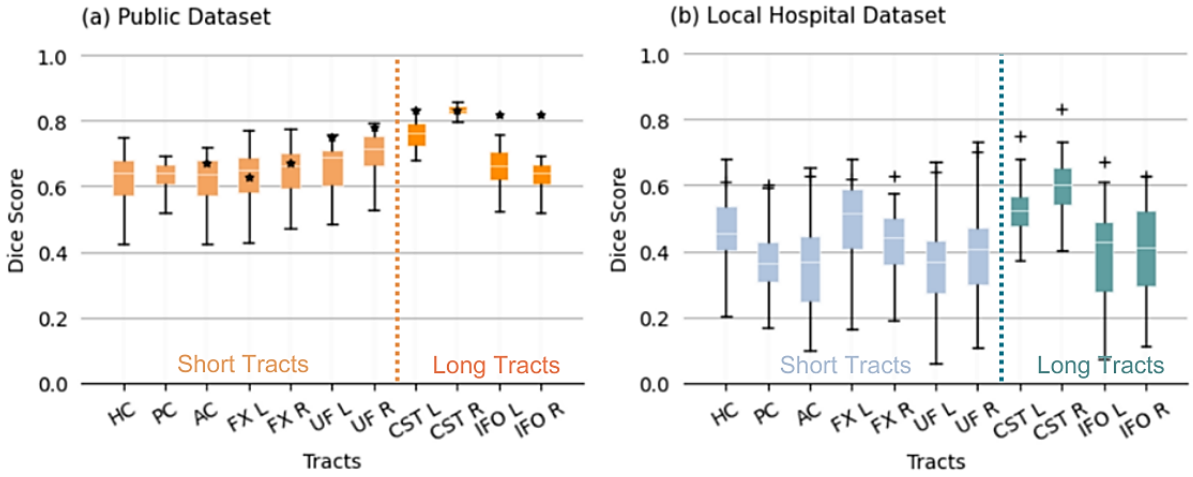
Source: Elaborated by the Author.

3.3 QUANTITATIVE RESULTS

3.3.1 1ST EXPERIMENT

The results of the first experiment are presented in Figure 36. For the public dataset, the short tracts achieved Dice scores ranging from 0.54 to 0.70, while the large tracts attained scores between 0.63 and 0.83. As expected, the long tracts exhibited higher prediction scores, as they are relatively easier to construct and do not have high levels of curvature. Despite CST being considered more challenging due to its lateral projections, the network successfully learned a significant portion of the tract, achieving impressive predictions of 0.75 for the left hemisphere and 0.83 for the right hemisphere. Among the short tracts, UF demonstrated the highest prediction scores, with 0.64 for the left hemisphere and 0.70 for the right hemisphere. This was followed by FX and AC, both achieving a score of 0.62. PC obtained the lowest scores, likely due to its shorter and thinner nature, which makes segmentation and consistent reference generation more challenging, particularly considering the anatomical variability across subjects.

Figure 36 - Dice score results for the training using only HCP data and testing in unseen subjects from the HCP dataset (a) and local hospital dataset (b) for each tract. (*) Results from literature (WASSERTHAL; NEHER; MAIER-HEIN, 2018). (+) Mean Dice score Training with Public Dataset.



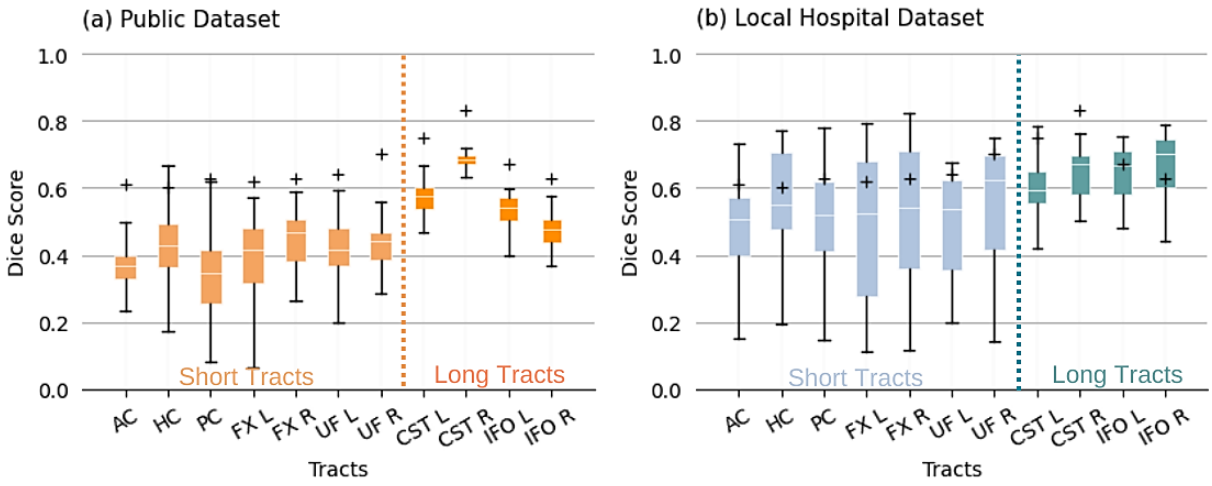
Source: Elaborated by the Author.

When the network trained solely on HCP data was tested on the hospital acquisitions, a noticeable decline in performance was observed. While the long tracts still achieved scores above 0.61, the shorter tracts obtained lower Dice scores, ranging from 0.36 for AC, HC, and FX L. Additionally, there was a larger standard deviation in the predictions across the subjects. This drop in performance can be attributed to several factors. The hospital acquisitions typically have lower signal-to-noise ratio, more distortions, and artifacts that impact the image quality. These factors, in turn, affect the accuracy of the orientation information used by the network to make precise tract segmentation predictions. The reduced amount of information and higher levels of distortions in the preprocessing stage can contribute to the decreased performance observed when applying the network trained on HCP data to the hospital acquisitions.

3.3.2 2ND EXPERIMENT

The results of the second experiment demonstrated an average improvement of 32% in the predictions for the local hospital dataset compared to the first experiment (Figure 37). Specifically, AC, PC, left FX, and left UF achieved Dice scores of 0.50, while right FX obtained a score of 0.53. Additionally, HC and right UF achieved a score of 0.56. These scores indicate the challenges faced by the network in learning and accurately segmenting these short tracts, even when the training is based on the same type of data. Despite the overall improvement observed in the second experiment, the lower Dice scores for these short tracts highlight the difficulty in capturing their specific characteristics and accurately predicting their boundaries.

Figure 37 - Dice score results for the training using only local hospital data and testing in unseen subjects from HCP dataset (a) and local hospital dataset (b) for each tract. (+) Mean Dice score Training with Public Dataset.



Source: Elaborated by the Author.

When applying the trained network to the HCP dataset, a decrease in performance was observed. The long tracts experienced a relatively smaller decrease of 20%, while the short tracts exhibited a more significant decrease of 37%. These findings highlight the dependency of the training process on the quality of the input data. Although the HCP dataset is known for its higher-quality data, the network trained with data of lower quality from the local hospital dataset could not achieve high scores when applied to the high-quality HCP data.

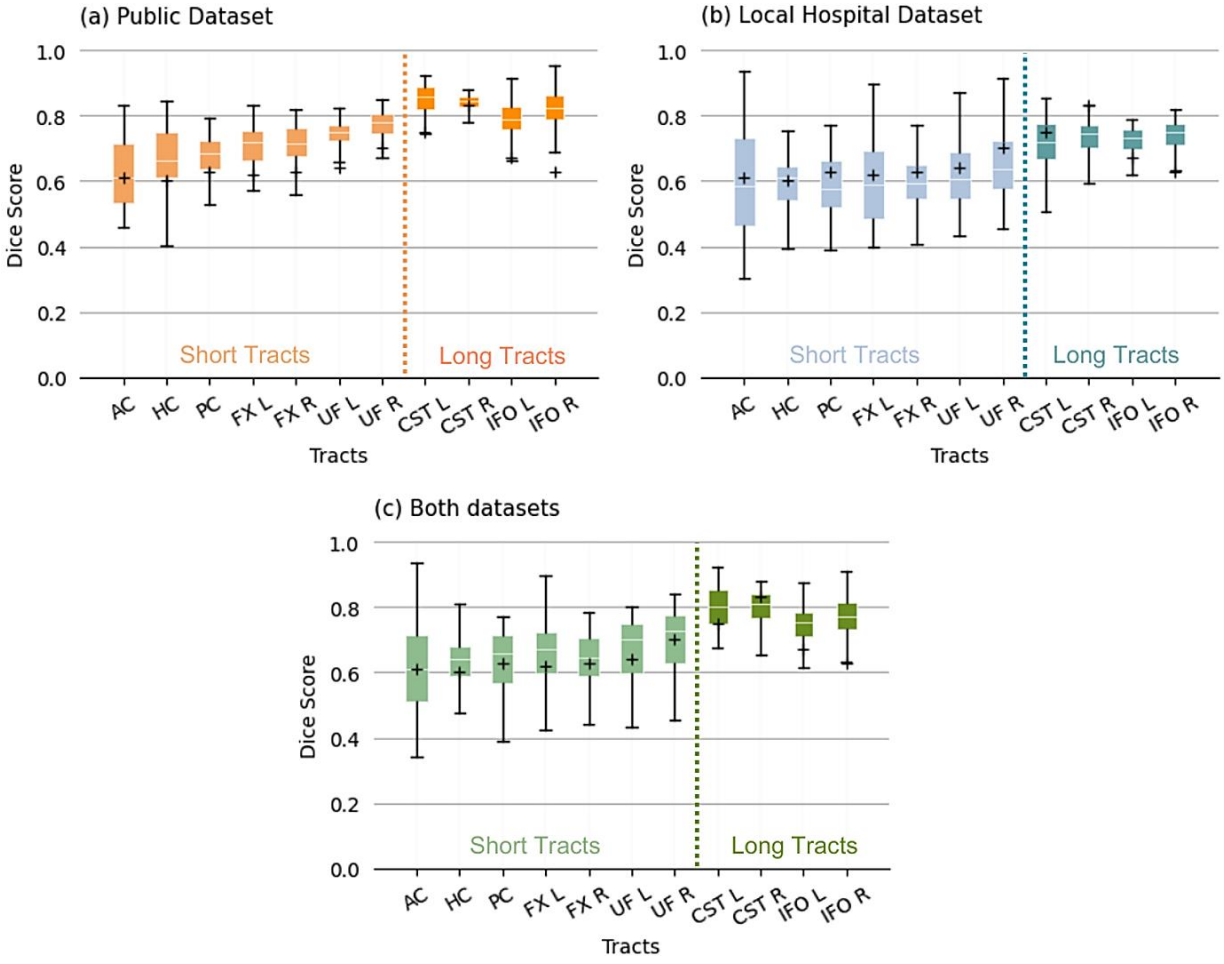
The performance of the trained network heavily relies on the information present in the training data. The network learns patterns and features from the input data to make predictions. Therefore, if the training data lacks certain information or exhibits specific characteristics, it becomes difficult for the network to generalize its predictions to unseen data. In the case of shorter and highly curved tracts, the intricacies of their structures pose additional challenges for accurate segmentation. The network may struggle to capture the fine details and variations in these tracts, leading to decreased performance and difficulty in generalizing predictions beyond the training data.

Overall, these results emphasize the need for robust training approaches that can handle variations in data quality and effectively address the complexities associated with segmenting shorter and highly curved tracts. This was what led us to carry out the third experiment.

3.3.3 3RD EXPERIMENT

By combining the datasets from the HCP and the local hospital, and standardizing the resolution, we aimed to leverage the strengths of both datasets and create a training environment that better encapsulates the characteristics of the tracts in question. This approach allowed us to mitigate some of the limitations encountered in the previous experiments and enhance the network's ability to generalize predictions across different datasets. As we can see in Figure 38, the scores obtained were higher compared with the first and second experiments for both datasets.

Figure 38 - Dice score results for the training using both datasets and testing in unseen subjects from HCP dataset (a), local hospital dataset (b), and both datasets randomized (c) for each tract. (+) Mean Dice score Training with Public Dataset.



Source: Elaborated by the Author.

We observed an average enhancement of 10% in performance for the public dataset when contrasted with outcomes achieved through training solely on the same public dataset. Among the short tracts, particularly substantial improvements were evidenced for the FX, 0.70 for both

hemispheres and UF, 0.72 for the left and 0.75 for the right hemisphere. On the other hand, AC (0.62) and the right CST (0.84) exhibited relatively modest enhancements, yet these increments remained satisfactorily consistent with the values obtained during the first experiment.

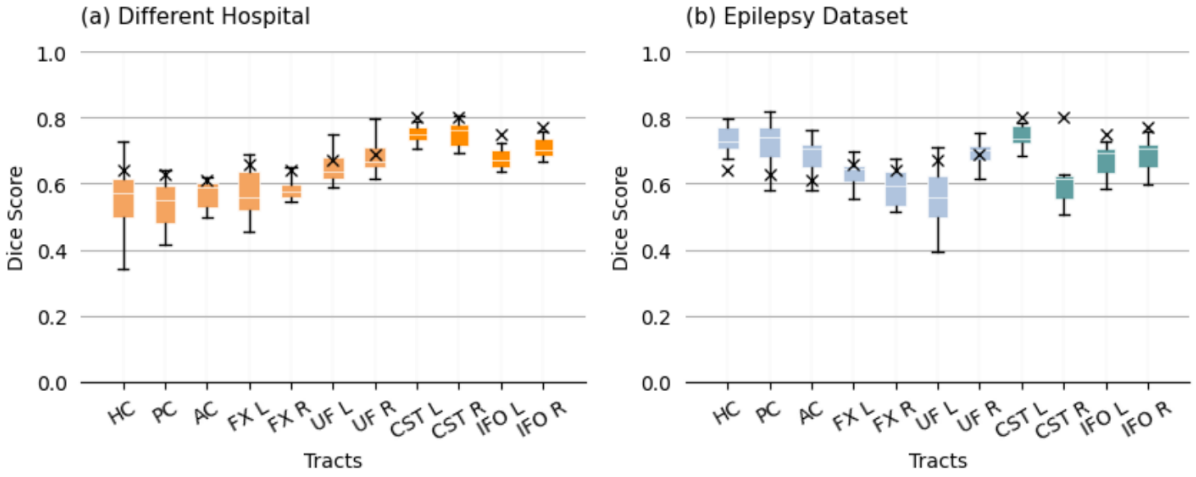
The local hospital dataset exhibited a substantial enhancement of 15% when compared to outcomes derived solely from training with the local hospital dataset. Notably, these results closely paralleled those of the initial experiment involving the public dataset, as depicted in Figure 38 – b. AC (0.60), left FX (0.59), and left UF (0.61) showcased notable performance gains within the local hospital dataset compared to the outcomes of the second experiment with mean dice scores of 0.48, 0.49, and 0.49 respectively. Although the achieved results weren't identical to the public dataset's, their proximity underscores the efficacy of our approach in different acquisition settings.

In the scenario where predictions were made using a combination of both datasets, the outcomes consistently surpassed those achieved solely with training on the public dataset from the first experiment, except for the right CST (Figure 38 – c). This outcome underscores the remarkable generalization capabilities of the neural network, notably pronounced for the short tracts. For instance, left FX achieved a high score of 0.66, HC scored 0.64, and left UF reached an impressive 0.67, exemplifying the network's robust performance across all three experiments.

3.4 TRANSFERABILITY BETWEEN SCANNERS AND ACQUISITION SETTINGS

The outcomes garnered from the dataset acquired on a distinct scanner were notably promising, exhibiting a slightly lower performance compared to those attained with the randomized dataset in the third experiment, similar to the behavior of the local hospital dataset result in the third experiment, as demonstrated in Figure 39 – a. Nonetheless, these results effectively underscore the network's generalization prowess, demonstrating its aptitude for generating predictions across varying acquisition settings and diverse scanners.

Figure 39 - Results for the dataset acquired on different scanner – healthy subjects (a) and in local hospital – epilepsy patients (b) using the network of the third experiment. (x) Mean Dice score Training with Both Datasets – Randomized.

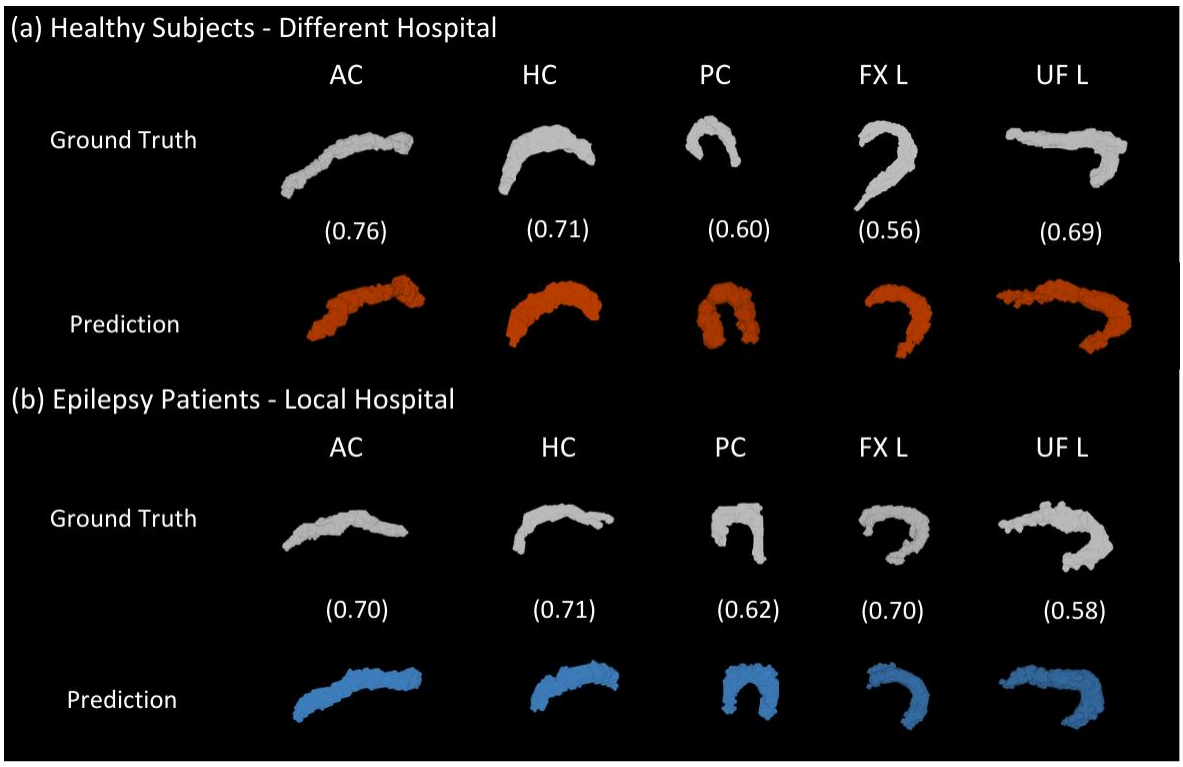


Source: Elaborated by the Author.

In the case of epilepsy patients, our achievements surpassed the results obtained with the randomized dataset in the third experiment, as depicted in Figure 39 – b. Despite the exclusive training on healthy subjects, the inherent variability within the training data proved sufficient for the network to yield commendable predictions even in cases of brain diseases. This underscores the model's adaptability and robustness in accommodating variations and complexities across diverse clinical scenarios.

The qualitative analysis displays the network's adeptness in reconstructing short tracts with remarkable fidelity across both datasets (Figure 40). Notably, the network exhibited proficiency in ensuring seamless continuity, devoid of gaps, and minimal disruptions in the tracts' boundaries or curvature. Thus, validating its precision in representing the structural details for datasets acquired in different conditions and with healthy and non-healthy brains.

Figure 40 - Qualitative results for the testing in data acquired in different scanners and a non-healthy brain for one random subject of each dataset (Dice Score).



Source: Elaborated by the Author.

3.5 RUNTIME

Table 5 presents an overview of the time allocation associated with various stages of the process, including preprocessing, tractography generation, standard space registrations, and method predictions. One of the numerous advantages of automation is its capacity to yield time efficiencies, which is why we have delineated the time expenditure for each of these processes.

The preprocessing, tractography generation, and registration procedures, conducted using MRtrix3 and FSL software, were executed on a machine equipped with a 1.80GHz Intel Core i7 processor with 8GB RAM. Conversely, the training and predictions were performed on a machine equipped with 2 Intel(R) Xeon(R) Gold 6130 CPU, 2.10GHz (16 cores x 2 threads per core) and 512GB RAM. The first and second experiment taken a mean time of 8 hours to train in one view, totalizing 24 hours to finish the total training for one specific tract. In contrast, the third experiment required a cumulative training time of 72 hours across all views. This was anticipated, as the number of training subjects had doubled, resulting in the extended duration.

Table 5 - Time consumption related with the preprocessing images, tractography generation, registration to standard space and predictions with the model for one subject.

Process	***Public Dataset***	***Local Hospital Dataset***
Preprocessing*	69 min	66 min
Tractography	52 min	18 min
MNI Registration	10 min	11 min
Prediction with the proposed method	1,6 min	1,6 min

* Public dataset preprocessing refers to bias correction and multi-tissue multi-shell CSD modelling. Local hospital dataset preprocessing is referent to distortion, motion, Gibbs, eddy current corrections and CSD modelling.

Source: Elaborated by the Author.

As illustrated in Table 5, the most time-consuming phases were the data preprocessing (particularly peak extraction using CSD), accounting for approximately 84% of the total processing time. It's noteworthy that the segmentation prediction, executed after the training phase, is highly efficient, taking less than 2 minutes to generate predictions on new data. This timing includes the necessary preparations for testing data, such as cropping, slicing, tensor formatting, concatenating results from each viewpoint, and saving the final predictions.

4 DISCUSSION

The automatic segmentation of white matter tracts is of great interest in the clinical setting. While deep learning methods have shown remarkable results in recent years, several steps are still necessary before this tool can be integrated into routine hospital practice. These steps include validation on more clinical-like datasets and the improvement of short tracts segmentation.

When comparing our results with the literature, we obtained similar outcomes for long tracts such as the CST, and short tracts like AC, FX, and right UF when training and testing on public datasets (WASSERTHAL; NEHER; MAIER-HEIN, 2018). However, left and right IFO showed lower prediction scores (17%), which may be attributed to differences in factors such as data augmentation, number of epochs, reference segmentation generation methods, and number of the testing dataset. We also reported the segmentation of the HC and PC using deep learning, achieving dice scores comparable to other short tracts. This achievement underscores the effectiveness of U-Net in successfully segmenting short tracts characterized by significant curvature while maintaining a reasonable degree of segmentation accuracy.

The predominant focus in white matter tract segmentation research has been on the HCP dataset with the clinical testing involving small datasets from patients with neurological diseases or data downgrading, potentially constraining the reliability and generalizability of the conclusions drawn (DONG et al., 2019; LIU et al., 2022). In our study, we aimed to bolster the dependability of our results and offer a more comprehensive insight into the capabilities and constraints of our proposed method by enlarging the sample size, encompassing a wider range of ages, and incorporating images from clinical acquisitions.

Upon applying our segmentation model to clinical data, an anticipated reduction in performance was observed, attributed to the inherent lower quality of the data and the broader age range encompassed within clinical datasets. As Figure 33 illustrates, the distinction between data acquired outside the clinical routine and in a hospital daily setting is quite evident. For instance, comparing the SNR between the high-quality public dataset (47.54) and the clinical acquisition (37.60) reveals a substantial 21% decrease, highlighting the considerable quality gap between these data types.

Furthermore, the clinical dataset encompasses a wider age range (18-83 years), unlike the narrower span of the HCP dataset (18-36 years). Considering the well-documented variations in healthy aging across various brain tracts, the expanded age diversity within the

clinical dataset introduces a broader spectrum of variability (PINTO; SANTOS; SALMON, 2021). This variability is likely a contributing factor to the observed higher standard deviation and potentially elucidates the decrease in performance. The neural network, lacking exposure to such variability during training, might encounter challenges in effectively accommodating the diverse age-related differences present in the clinical dataset.

With this in mind, we made an interesting observation when we reversed this scenario and trained the model using clinical data, and then made predictions on high-quality data. Surprisingly, we obtained scores lower than the local hospital data and gaps in the visual segmentation reconstruction for the unseen public data. This finding emphasizes the limitations of the supervised learning approach, which relies on the combination of training, validation, and testing data type more than the availability and quality of the training data.

When the model encounters data from scanners and acquisition sequences that are different from the training data, a domain shift occurs, leading to a reduction in performance. The decrease in performance when applying the model to high-quality data suggests that the model has learned specific features and patterns that are unique to the clinical dataset. These features may not be as relevant or applicable to high-quality data, leading to a decrease in dice scores. Highlight the challenges of transferring the segmentation model trained on one dataset to another dataset with different acquisition characteristics.

By combining both lower quality and higher quality data, we can achieve better generalization for the neural network predictions, as evidenced by the resulting performance which consistently exceeded that achieved when training with only one type of dataset. This trend was particularly pronounced in the context of short tracts, which obtained the highest scores across all experiments. This heightened performance can be attributed to the broader variability encompassed during training, allowing the network to better adapt to diverse data scenarios. The method's adaptability is remarkably showcased in its application to images obtained from diverse scanners and epilepsy patients.

The results derived from a dataset acquired using a different scanner exhibited significant promise, displaying a marginally lower performance in comparison to the outcomes of the third experiment. This minor reduction in performance can be traced back to variations in diffusion gradients. Precisely, the training data comprised only 32 directions, whereas the dataset in question featured 64 directions. This difference in directional information during both training and evaluation could explain the observed performance variation.

In the context of epilepsy patients, our achievements surpassed the results obtained in the third experiment, conducted with a random sample. This particular dataset, mirroring the acquisition traits of our training data by originating from the same hospital and spanning an age range of 26 to 66 years, produced scores well within the optimal performance range of the neural network. This validation emphasizes the model's accuracy in depicting the structural complexities inherent in datasets gathered under diverse conditions, capturing both healthy and non-healthy brain characteristics.

This leads us to the conclusion that in order to make the application more viable in a clinical setting, the path forward involves training the model using MRI datasets that closely mimic real clinical scenarios, and whenever possible, incorporate authentic clinical data.

In closing, our study focused on the application of deep learning for automatic short tracts segmentation, with a particular emphasis on clinical relevance. U-Net emerges as a great model for tract segmentation, especially when paired with a well-balanced approach to training, validation, and testing. Our findings indicated that the key to clinical applicability of deep learning segmentation lies in training the model on MRI datasets that closely resemble real clinical scenarios and incorporating authentic clinical data whenever possible. This research contributes to the ongoing efforts to enhance the reliability and applicability of deep learning-based white matter tract segmentation in clinical settings.

4.1 LIMITATIONS

The segmentation of short white matter tracts presents unique challenges that are evident in our results, even when we matched both datasets. Besides the improved results, the dice scores still stay between 0.60 and 0.75. One of the major limitations in tractography is the lack of a definitive ground truth, which makes the generation of reference data for training and evaluation challenging.

Short tracts pose additional challenges due to their complex and highly curved nature. For example, the PC and FX exhibit significant levels of curvature, making it difficult to consistently identify and segment these tracts across subjects. Similarly, the lateral part of the AC can vary in its detailed anatomical definition, further contributing to the difficulty in accurate segmentation.

Another factor to consider is the registration process, which in the present work was used bring the images into a standard space for a faster segmentation. However, this process can introduce image deformations that may impact the generation of final tract masks. These

deformations can particularly affect the accuracy of segmenting short tracts, as their smaller size and intricate anatomical features are more susceptible to distortion. Impacting the resulting tract used as label.

Lastly, it's important to underscore that the quantity of training examples has a direct and profound impact on the performance of AI systems. Increased data volume consistently yields superior generalization, reduced bias, enhanced model complexity, heightened stability, and greater robustness.

4.2 DATA AVAILABILITY

In recent years, the availability of code has greatly benefited the research field. However, many existing pipelines require multiple steps and advanced coding knowledge to understand and utilize effectively. Additionally, when dealing with diffusion magnetic resonance imaging and deep learning, training models can be computationally demanding, requiring substantial memory and GPU resources. This can result in lengthy runtimes, sometimes spanning weeks or months.

To address these challenges, we aimed to optimize and develop an approach that offers a favorable cost-benefit ratio with shorter runtimes and an easier setup. Our focus was on creating an understandable and didactic Python code that can be utilized by researchers who are new to this area or do not possess advanced programming skills. We are pleased to share that our code is publicly available on GitHub at

https://github.com/inbrainlab/Short-Tracts-Automatic-Segmentation.

It can be easily employed to train new tracts or expand the dataset using Google Collaboratory. Furthermore, we have made the pre-trained weights accessible, allowing for transfer learning or making predictions on new data.

5 CONCLUSION

Our study highlights both the potential and challenges associated with the automatic segmentation of white matter tracts for clinical applications. While deep learning methods show considerable promise, there are crucial steps to be taken before this technology can seamlessly integrate into routine medical practice.

Many prior studies in WM tract segmentation have primarily focused on limited datasets, potentially hindering the generalizability of their findings. In contrast, our study sought to enhance the reliability of results by expanding the dataset, incorporating clinical acquisition images, and accommodating a wider range of age groups.

When applying our segmentation model to clinical data, we anticipated a decrease in performance due to issues related to data quality and the broader age ranges present in clinical datasets. This observation underscores the inherent challenges of supervised learning and the necessity to address domain shifts when transferring models across datasets with varying acquisition characteristics.

Through our findings, we observed the most significant score enhancements when augmenting the dataset with examples from various sources, all subjected to a standardization process to closely match the characteristics of the clinical dataset. As a result, we achieved elevated scores in the evaluation of short tracts, ranging between 0.61 and 0.69. Furthermore, in the transferability assessment, the network demonstrated elevated performance even when confronted with non-healthy brain data, experiencing only a minimal decline attributed to disparities in diffusion directions across different hospitals.

In conclusion, automatic WM tract segmentation using AI holds immense promise for clinical applications and has the potential to significantly enhance reproducibility in the context of tractography segmentation. Nevertheless, continuous research efforts are imperative to tackle the complexities associated with short tracts, bolster the model's generalization capabilities, and fine-tune the techniques to achieve heightened accuracy and reliability across various clinical settings.

6 REFERENCES

ALEXANDER, A. L., et al. Diffusion Tensor Imaging of the Brain. Neurotherapeutics: the journal of the American Society for Experimental **NeuroTherapeutics**, v. 4, p. 316–329, 2007.

ANAND, K. S.; DHIKAV, V. Hippocampus in health and disease: An overview. **Ann. Indian Acad. Neurol.**, v. 15, p. 239–246, 2012.

ANDERSSON, J. L. R.; SOTIROPOULOS, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. **NeuroImage**, v. 125, p. 1063–1078, 2016.

ANDICA, C.; KAMAGATA, K.; AOKI, S. Automated three-dimensional major white matter bundle segmentation using diffusion magnetic resonance imaging. **Anat. Sci. Int**., v. 98, p. 33–336, 2023.

BASSER, P. J.; MATTIELLO, J.; LEBIHAN, D. MR diffusion tensor spectroscopy and imaging. **Biophysical Journal**, v. 66, n. 1, p. 259–267, 1994.

BASSER, P. J.; JONES, D. K. Diffusion-tensor MRI: theory, experimental design and data analysis – a technical review. **Nmr Biomed.**, v. 15, p. 456–467, 2002.

BASSEL, J. et al. Diffusion Tensor Imaging Abnormalities in the Uncinate Fasciculus and Inferior Longitudinal Fasciculus in Phelan-McDermid Syndrome. **Pediatr Neurol**, v. 106, p. 24–31, 2020.

BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. Neuroscience: exploring the brain. Fourth edition. **Jones & Bartlett Learning**, February 17, 2015.

BEHRENS, T. E. J., et al. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? **Neuroimage**, v. 34, p. 144–155, 2007.

BHATIA, K., et al. Diffusion Tensor Imaging Investigation of Uncinate Fasciculus Anatomy in Healthy Controls: Description of a Subgenual Stem. **Neuropsychobiology**, v. 75, p. 132–140, 2018.

BULLOCK, D. N., et al. A taxonomy of the brain's white matter: twenty-one major tracts for the 21st century. **Cereb. Cortex**, v. 32, p. 4524–4548, 2022.

CONCHA L., et al. In Vivo Diffusion Tensor Imaging and Histopathology of the Fimbria-Fornix in Temporal Lobe Epilepsy. **Journal of Neuroscience**, v. 30, p. 996–1002, 2010.

COVER, G. S., et al. Computational methods for corpus callosum segmentation on MRI: A systematic literature review. **Comput. Methods Programs Biomed.**, v. 154, p. 25–35, 2018.

CHOI, S.-H., et al. Anterior Commissure - Posterior Commissure Revisited. **Korean J Radiol**, v. 14, p. 653–661, 2013.

CURRIE, G., et al. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. **J. Méd. Imaging Radiat. Sci.**, v. 50, p. 477–487, 2019.

ÇAVDAR, S., et al. The Complex Structure of the Anterior White Commissure of the Human Brain: Fiber Dissection and Tractography Study. **World Neurosurg.**, v. 147, p. 111–117, 2021.

DIEHL, B. et al. Abnormalities in diffusion tensor imaging of the uncinate fasciculus relate to reduced memory in temporal lobe epilepsy. **Epilepsia**, v. 49, p. 1409–1418, 2008.

DONG, X., et al. Multimodality white matter tract segmentation using CNN. **ACM TURC-AIS**, n. 145, p. 1-8, 2019.

DYRBA, M., et al. Robust Automated Detection of Microstructural White Matter Degeneration in Alzheimer's Disease Using Machine Learning Classification of Multicenter DTI Data. **Plos One**, v. 8, id. e64925, 2013.

ESSAYED, W. I. et al. White matter tractography for neurosurgical planning: A topography-based review of the current state of the art. **Neuroimage Clin**, v. 15, p. 659–672, 2017.

ERTEL, W. Introduction to Artificial Intelligence. 2nd Edition. **Springer**, 2017.

FENLON, L. R., et al. The evolution, formation and connectivity of the anterior commissure. **Semin Cell Dev Biol**, v. 118, p. 50–59, 2021.

FONOV, V. et al. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. **NeuroImage**, vol. 47, Suppl 1, p. S102, 2009.

GARYFALLIDIS, E., et al. QuickBundles, a method for tractography simplification. **Front. Neurosci.**, v. 6, 175, 2012.

GARYFALLIDIS, E., et al. Recognition of white matter bundles using local and global streamline-based registration and clustering. **NeuroImage**, v. 170, p. 283–295, 2018.

GEVA, S.; CORREIA, M.; WARBURTON, E. A. Diffusion tensor imaging in the study of language and aphasia. **Aphasiology**, v. 25, p. 543-558, 2011.

GHAZI, N.; AARABI, M. H.; SOLTANIAN-ZADEH, H. Deep Learning Methods for Identification of White Matter Fiber Tracts: Review of State-of-the-Art and Future Prospective. **Neuroinformatics**, p. 1–32, 2023.

GLASSER, et al. The minimal preprocessing pipelines for the Human Connectome Project. **Neuroimage**, v. 80, p. 105–124, 2013.

GONZÁLEZ-VILLÀ, S. et al. A review on brain structures segmentation in magnetic resonance imaging. **Artif Intell Med**, v. 73, p. 45–69, 2016.

GUEVARA, P. et al. Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. **Neuroimage**, v. 61, p. 1083–1099, 2012.

HAACKE, E. M., et al. Magnetic Resonance Imaging: Physical Principles and Sequence Design. 1st Edition. **Wiley**, June 10, 1999.

HAO, X.; ZHANG, G.; MA, S. Deep Learning. **Int. J. Semantic Comput.,** v. 10, p. 417–439, 2016.

HEINE, J. et al. Transdiagnostic hippocampal damage patterns in neuro-immunological disorders. **NeuroImage: Clin.**, v. 28, 102515, 2020.

HUBEL D. H.; WIESEL T. N. Receptive fields of single neurones in the cat's striate cortex. **J Physiol.** v. 148, p. 574–91, 1959.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electron. Mark**., v. 31, p. 685–695, 2021.

JEURISSEN, B., et al. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. **Neuroimage.**, v. 103, p. 411–426, 2014.

JEURISSEN, B., et al. Diffusion MRI fiber tractography of the brain. **Nmr Biomed.**, v. 32, id. e3785, 2019.

KALAIVANI, P.; SUNDARESWARAN, N. Hippocampal Abnormalities in Adults with Unilateral Temporal Lobe Epilepsy: A Diffusion Tensor Imaging Study. **International Journal of Science and Research**, v. 8, p. 1564–1572, 2019.

KIUCHI, K. et al. Abnormalities of the uncinate fasciculus and posterior cingulate fasciculus in mild cognitive impairment and early Alzheimer's disease: A diffusion tensor tractography study. **Brain Res**, v. 1287, p. 184–191, 2009.

KROGH, A. What are artificial neural networks? **Nature Biotechnology**, v. 26, n. 2, p. 195–197, 2008.

KUZNIECKY, R., et al. Quantitative MRI in temporal lobe epilepsy: evidence for fornix atrophy. **Neurology Aug**, v. 53 (3) 496, 1999.

LI, B. et al. Neuro4Neuro: A neural network approach for neural tract segmentation using large-scale population-based diffusion imaging. **NeuroImage**, v. 218, id.116993, 2020.

LIN, Z. et al. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. **Med Phys,** v. 46, p. 3101–3116, 2019.

LIU, W. et al. Volumetric Segmentation of White Matter Tracts with Label Embedding. **Neuroimage**, v. 250, n 118934, 2022.

LIU, Y; DAWANT, B. M. Automatic Localization of the Anterior Commissure, Posterior Commissure, and Midsagittal Plane in MRI Scans using Regression Forests. **Ieee J Biomed Health**, v. 19, p. 1362–1374, 2015.

MANDAL, P.K. Magnetic Resonance Spectroscopy (MRS) and Its Application in Alzheimer's Disease. **Wiley InterScience**, Concepts in Magnetic Resonance Part A, v. 30A(1), p. 40–64, 2007.

MARIEB, E. N.; HOEHN, K. N. Human Anatomy & Physiology. 9th Edition. **Benjamin-Cummings Publishing Company**, January 16, 2012.

MENG, Z.; HU, Y.; ANCEY, C. Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. **Water**, v. 12, n. 600, 2020.

MORI, S. et al. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. **Annals of neurology**, v. 45, n. 2, p. 265–9, 1999.

MORI, S; AGGARWAL, M. In vivo magnetic resonance imaging of the human limbic white matter. **Front Aging Neurosci**, v. 6, n. 321, 2014.

MOON, W.-J., et al. Atrophy Measurement of the Anterior Commissure and Substantia Innominata with 3T High-Resolution MR Imaging: Does the Measurement Differ for Patients with Frontotemporal Lobar Degeneration and Alzheimer Disease and for Healthy Subjects? **Am. J. Neuroradiol.**, v. 29, p. 1308–1313, 2008.

NEHER, P. F., et al. A Machine Learning Based Approachto Fiber Tractography Using Classifier Voting. **Lect Notes Comput Sc**, v. 9349, p. 45–52, 2015.

ONGSULEE, P. Artificial intelligence, machine learning and deep learning. **15th Int. Conf. ICT Knowl. Eng.** (ICTKE), p. 1–6, 2017.

O'SHEA, K.; NASH, R. An Introduction to Convolutional Neural Networks**. ArXiv**, 2015. Doi:10.48550/arxiv.1511.08458.

OZDEMIR, N. G. The Anatomy of the Posterior Commissure. **Turk. Neurosurg.** v. 25, p. 837–43, 2015.

PARKINSON PROGRESSION MARKER INITIATIVE. The Parkinson Progression Marker Initiative (PPMI). **Prog Neurobiol.**, v. 95(4), p. 629-35, 2011.

PERETZKE, R. et al. atTRACTive: Semi-automatic white matter tract segmentation using active learning. **ArXiv,** 2023. Doi:10.48550/arxiv.2305.18905.

PINTO, M. S.; SANTOS, A. C. dos; SALMON, C. E. G. Age-related assessment of diffusion parameters in specific brain tracts correlated with cortical thinning. **Neurol Sci**, v. 42, p. 1799–1809, 2021.

POULIN, P., et al. Tractography and machine learning: Current state and open challenges. **Magn Reson Imaging**, v. 64, p. 37–48, 2019.

PRAKASH, K. N. B.; NOWINSKI, W. L. Morphologic Relationship Among the Corpus Callosum, Fornix, Anterior Commissure, and Posterior Commissure MRI-Based Variability Study. **Acad Radiol**, v. 13, p. 24–35, 2006.

RONNENBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. **Lect Notes Comput Sc**, v. 9351, p. 234–241, 2015.

REISERT, M., et al. HAMLET: Hierarchical Harmonic Filters for Learning Tracts from Diffusion MRI. **Arxiv**, 2018. Doi:10.48550/arxiv.1807.01068.

RHEAULT, F. et al. Tractostorm: The what, why, and how of tractography dissection reproducibility. **Hum. Brain Mapp.**, v. 41, p. 1859–1874, 2020.

SALADIN, K. Anatomy & Physiology: The Unity of Form and Function. 6th Edition. **New York: McGraw Hill**, 2012.

SEIBERLICH, N. et al. Advances in Magnetic Resonance Technology and Applications. Volume 1. **Elsevier Inc**, 2020.

SCHILLING, K. G. et al. Tractography dissection variability: What happens when 42 groups dissect 14 white matter bundles on the same dataset? **Neuroimage**, v. 243, n. 118502, 2021 – (a).

SCHILLING, K. G. et al. Fiber tractography bundle segmentation depends on scanner effects, vendor effects, acquisition resolution, diffusion sampling scheme, diffusion sensitization, and bundle segmentation workflow. **Biorxiv**, 2021. Doi:10.1101/2021.03.17.435872 – (b).

SHU, X., et al. Voxel-Based Diffusion Tensor Imaging of an APP/PS1 Mouse Model of Alzheimer's Disease. **Mol Neurobiol**, v. 48, p. 78–83, 2013.

SINGH, N. M. et al. How Machine Learning is Powering Neuroimaging to Improve Brain Health. **Neuroinformatics**, v. 20, p. 943–964, 2022.

SMITH, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. **NeuroImage**, v. 23, p. 208–219, 2004.

SMITH, R. E., et al. Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. **Neuroimage**, v. 62, p. 1924–1938, 2012.

SPENCER, S. S., et al. Human Hippocampal Seizure Spread Studied by Depth and Subdural Recording: The Hippocampal Commissure. **Epilepsia,** v. 28, p. 479–489, 1987.

STEJSKAL, E.O.; TANNER, J.E. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient. **The Journal of Chemical Physics**, v. 42, p. 288-292, 1965.

TAHA, A. A.; HANBURY, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. **BMC Med Imaging**, v. 15, n. 29, 2015.

TCHETCHENIAN, A. et al. A comparison of manual and automated neural architecture search for white matter tract segmentation. **Sci. Rep.**, v. 13, n. 1617, 2023.

TOFTS, P. S.; DOWELL, N. G.; CERCIGNANI, M. Quantitative MRI of the Brain: Principles of Physical Measurement. 2nd edition. **CRC Press**, January 26, 2018.

THOMAS, A. G.; KOUMELLIS, P.; DINEEN, R. A. The Fornix in Health and Disease: An Imaging Review. **Radiographics**, v. 31, p. 1107–1121, 2011.

TOURNIER, J. D. et al. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. **Neuroimage**, v.23, n. 3, p. 1176–1185, 2004.

TOURNIER, J. D.; CALAMANTE, F.; CONNELLY, A. Improved probabilistic streamlines tractography by 2nd order integration over fiber orientation distributions. In: **Proceedings of the 19th International Society for Magnetic Resonance in Medicine**, p 1670, 2010.

TOURNIER, J. D.; MORI, S.; LEEMANS, A. Diffusion Tensor Imaging and Beyond. **Magn Reason Med**, v. 65(6), p. 1532–1556, 2012.

TOURNIER, J. D., et al. MRtrix3: a fast, flexible and open software framework for medical image processing and visualization. **Neuroimage**, v. 202, n.116137, 2019.

TUSTISON, N. J. et al. N4ITK: Improved N3 Bias Correction**. IEEE Transactions on Medical Imaging**, v. 29, n. 6, p. 1310–1320, 2010.

VAN ESSEN, D. C., et al. The WU-minn human connectome project: an overview. **Neuroimage**, v. 80, p. 62–79, 2013.

VÁZQUEZ, A., et al. FFClust: Fast fiber clustering for large tractography datasets for a detailed study of brain connectivity. **Neuroimage**, v. 220, n.117070, 2020.

VERAART, J. et al. Denoising of diffusion MRI using random matrix theory. **NeuroImage**, v. 142, p. 394–406, 2016.

VOUTSADAKIS, G. Introduction to Artificial Intelligence. Lectures, February 2014. Available in: http://www.voutsadakis.com/TEACH/LECTURES/INTROAI/Chapter7.pdf. Accessed in 15 sept. 2023.

WAKANA, S. et al. Fiber tract-based atlas of human white matter anatomy. **Radiology**, v. 230, n. 1, p. 77–87, 2004.

WANG, F., et al. Microsurgical and tractographic anatomical study of insular and transsylvian transinsular approach. **Neurol. Sci.** v. 32, p. 865–874, 2011.

WASSERMANN, D., et al. Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. **Neuroimage**, v. 51, p. 228–24, 2010.

WASSERMANN, D., et al. The white matter query language: a novel approach for describing human white matter anatomy. **Brain Struct. Funct.**, v. 221(9), p. 4705–4721, 2016.

WASSERTHAL, J., et al. H. Direct White Matter Bundle Segmentation using Stacked U-Nets. **Arxiv**, 2017. Doi:10.48550/arxiv.1703.02036.

WASSERTHAL, J.; NEHER, P.; MAIER-HEIN, K. H. TractSeg - Fast and accurate white matter tract segmentation. **Neuroimage**, v. 183, p. 239–253, 2018.

WENDE, T.; HOFFMANN, K.-T.; MEIXENSBERGER, J. Tractography in Neurosurgery: A Systematic Review of Current Applications. **J Neurological Surg Part Central European Neurosurg**, v. 81, p. 442–455, 2020.

WYCOCO, V., et al. White matter anatomy: what the radiologist needs to know. **Neuroimaging Clin N Am.**, v. 23, p. 197-216, 2013.

YANG, J. Y.-M., et al. Diffusion MRI tractography for neurosurgery: the basics, current state, technical reliability and challenges. **Phys Medicine Biology**, v. 66, n. 15TR01, 2021.

YEATMAN J.D., et al. Tract profiles of white matter properties: automating fibertract quantification. **PLoS ONE**, v. 7:E49790, 2011.

YENDIKI A., et al. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. **Front Neuroinform.**, v. 5:23, 2011.

YIN et al., X. X. U-Net-Based Medical Image Segmentation. **Journal of Healthcare Engineering**, ID 4189781, 2022.

ZHAN, L. et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. **Front Aging Neurosci**., v. 7, n. 48, 2015.

ZHANG, F. et al. Deep white matter analysis (DeepWMA): Fast and consistent tractography segmentation. **Med Image Anal**, v. 65, n. 101761, 2020.

ZHANG, F. et al. Quantitative mapping of the brain's structural connectivity using diffusion MRI tractography: A review. **Neuroimage**, v. 249, n. 118870, 2022.