Machine learning based computational models with permeability for white matter microstructure imaging

Ioana Hill

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Medical Physics and Bioengineering
Centre for Medical Image Computing
University College London

February 17, 2020

I, Ioana Hill, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Characterising tissue microstructure is of paramount importance for understanding neurological conditions such as Multiple Sclerosis. Therefore, there is a growing interest in imaging tissue microstructure non-invasively. One way to achieve this is by developing tissue models and fitting them to the diffusion-MRI signal. Nevertheless, some microstructure parameters, such as permeability, remain elusive because analytical models that incorporate them are intractable. Machine learning based computational models offer a promising alternative as they bypass the need for analytical expressions. The aim of this thesis is to develop the first machine learning based computational model for white matter microstructure imaging using two promising approaches: random forests and neural networks. To test the feasibility of this new approach, we provide for the first time a direct comparison of machine learning parameter estimates with histology.

In this thesis, we demonstrate the idea by estimating permeability via the intra-axonal exchange time τ_i, a potential imaging biomarker for demyelinating pathologies. We use simulations of the diffusion-MRI signal to construct a mapping between signals and microstructure parameters including τ_i. We show for the first time that clinically viable diffusion-weighted sequences can probe exchange times up to $\approx$1000 ms. Using healthy in-vivo human and mouse data, we show that our model's estimates are within the plausible range for white matter tissue and display well known trends such as the high-low-high intra-axonal volume fraction f across the corpus callosum. Using human and mouse data from demyelinated tissue, we show that our model detects trends in line with the expected MS pathology: a significant decrease in f and τ_i. Moreover, we show that our random forest estimates of f and τ_i correlate very strongly with histological measurements of f and myelin thickness.

This thesis demonstrates that machine learning based computational models are a feasible approach for white matter microstructure imaging. The continually improving SNR in the clinical scanners and the availability of more realistic simulations open up possibilities of using such models as imaging biomarkers for demyelinating diseases such as Multiple Sclerosis.

To my grandfather,
who started this journey with me,
but could not arrive at the other end.
Requiescat in pace.

Acknowledgements

I am thankful to my PhD supervisors, Dr. Ivana Drobnjak, Prof. Olga Ciccarelli and Dr. Marco Palombo for their advice and guidance over the last three years.

I am also very grateful for Ivana's science and life lessons (sometimes indistinguishable from each other!), from which I have had a lot to learn. Thank you for your supervision and for shaping me into an independent researcher. I would also like to thank Marco for his constant support and scientific skills that made a great contribution to this research. This PhD would not have been as enjoyable without our many conversations and coffee breaks. I would also like to thank Dr. Andrada Ianus for her invaluable guidance on diffusion MRI from the very beginning of my PhD. I am also grateful to Prof. Daniel Alexander for all his support at key points throughout this PhD.

There are many colleagues (which I can now call friends) that have made my time at UCL one of the most enjoyable periods of my life and have helped me sail through the more difficult times: Aaron, Alex, Andrada, Irina, Joe, Kiko, Lizzie, Marco, Maria, Maira, Maura, Noemi, Oscar, Paddy, Peter, Raz, Rica, Ryu, Sara, Senda. Thank you, it's truly been a pleasure!

Special thanks go to my family and teachers who have encouraged my academic endeavours from a very early age. I would also like to thank Prof. Peter Tino who introduced me to the world of research and machine learning almost a decade ago and with whom I have thoroughly enjoyed working.

Finally, my deepest gratitude and love goes to my husband and best friend Dom - I would not be here without you.

Journal and Conference Publications

- Hill, I., Palombo, M.[1], Santin, M., Branzoli, F., Philippe, A. C., Wassermann, D., ... Petiet, A., Alexander, D. C., Ciccarelli, O., Drobnjak, I. 'Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination.' Submitted to *NeuroImage* (24/07/2019)

- Nedjati-Gilani, G. L., Schneider, T., Hall, M. G., Cawley, N., Hill, I., Ciccarelli, O., ... Alexander, D. C. 'Machine learning based compartment models with permeability for white matter microstructure imaging.' *NeuroImage, 150* (2017): 119-135

- Hill, I., Palombo, M., Santin, M., Branzoli, F., Philippe, A. C., Wassermann, D., ... Petiet, A., Alexander, D. C., Ciccarelli, O., Drobnjak, I. 'Deep neural network based framework for in-vivo axonal permeability estimation.' 26th Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM). 2018. *One of the five abstracts selected for the high-profile Educational-Scientific Session 'Machine Learning for Magnetic Resonance in Medicine'* - oral presentation

- Palombo, M., Hill, I.[1], Santin, M., Branzoli, F., Philippe, A. C., Wassermann, D., ... Petiet, A., Alexander, D. C., Ciccarelli, O., Drobnjak, I. 'Machine learning based estimation of axonal permeability: validation on cuprizone treated in-vivo mouse model of axonal demyelination.' 26th Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM). 2018 - power pitch (format giving maximum exposure to the most interesting abstracts at the annual meeting)

- Hill, I., Ianus, A., Ciccarelli, O., Drobnjak, I. ' Sensitivity of STEAM diffusion MRI to permeability in white matter tissue: a simulation study'. 25th Annual Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM). 2017

---

[1] shared first author.

Impact Statement

This PhD research provides novel insights into the analysis of diffusion-weighted MRI data in the context of demyelinating neurological diseases such as Multiple Sclerosis. Specifically, this work focuses on the use of machine learning to define new imaging biomarkers of axonal membrane permeability for Multiple Sclerosis.

Our work demonstrates for the first time the clinical potential of machine learning methods for microstructure parameter estimation (Chapters 4 and 5). This research delivers a machine learning based computational framework which could be used in clinical and biomedical research to obtain estimates of informative tissue parameters from diffusion-weighted MRI scans. We demonstrate the clinical applicability of our framework in an example which estimates membrane permeability in the brain white matter. Using healthy volunteers and Multiple Sclerosis patients, we find that our machine learning approach detects changes in the brain tissue in line with expectations from the pathology of Multiple Sclerosis. The preliminary results of this study were published in NeuroImage in February 2017.

We further validate our framework's potential as an imaging biomarker through direct comparison with histology measurements in a preclinical mouse model (Chapters 6, 7 and 8). Our results suggest for the first time, quantitatively and in-vivo, that machine learning based computational frameworks could act as suitable biomarkers for detecting and tracking changes in demyelinating pathologies. Besides Multiple Sclerosis, the applicability of our approach extends to other myelin damaging pathologies such as spinal cord injuries or leukodystrophies. The manuscript presenting these results is currently under review and has been submitted to NeuroImage on 24th July, 2019.

More generally, our framework can easily be extended to estimate a range of other informative microstructure parameters, which could be important for numerous white matter pathologies of the human nervous system. From a clinical perspective, these future developments can have a great impact on the understanding and diagnosis of neurological conditions of the white matter.

In order to maximise the impact of this research and to encourage its adoption by the wider research community, the work in this thesis has also been presented in high-impact educational and scientific sessions at the 2017 and 2018 annual meetings of the International Society for Magnetic Resonance in Medicine.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Studying brain tissue microstructure properties such as the size of axons or the permeability of cell membranes is of paramount importance for diagnosing and monitoring neurological conditions such as Multiple Sclerosis (MS) or Parkinson's disease. By measuring microstructure properties in both healthy and diseased tissue, we can gain insight into the mechanisms underlying pathologies and potentially develop new biomarkers that can help with their diagnosis and treatment. This is particularly important as many neurological conditions are still very poorly understood and lack effective treatment. An example of this is MS, characterised by the formation of lesions in white matter which disrupts the ability of the nervous system to communicate effectively. The main underlying process in MS lesions is demyelination, which is characterised by a breakdown of the myelin sheath surrounding the axons in the tissue. The condition of the myelin sheath has been linked with the permeability associated with the intra-axonal exchange time τ_i [1, 2, 3], which measures the average time a water molecule spends inside the intra-axonal space. The intuitive explanation behind this link is that, as the myelin breaks down, there are less barriers preventing the water molecules from leaving the intra-axonal space, making the axons more permeable. Changes in permeability have also been linked with other pathologies such as Parkinson's disease [4] and cancer [5], leading to a widespread interest in developing permeability-based biomarkers.

Due to its sensitivity to the motion of water molecules within tissue, Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) is potentially able to estimate the intra-axonal exchange time τ_i, together with other important microstructural properties such as cell size or intracellular volume fraction. By fitting mathematical or computational models to the DW-MRI signal, we can estimate the microstructural properties of the underlying tissue in a completely non-invasive and non-ionising way.

So far, measuring the intra-axonal exchange time τ_i has been problematic due to the intractability of the mathematical models which accurately relate τ_i to the DW-MRI signal or, in the case of more simplistic models, due to unrealistic assumptions that are not valid in the brain white matter tissue. Given the challenges involved in deriving accurate analytical models with exchange, there have been alternative approaches which bypass the need for analytical models and use simulations of the DW-MRI signal to learn how this is affected by permeability. One such approach [6] constructs a library of synthetic DW-MRI signals and their associated microstructure parameters including τ_i, and estimates the parameters of a new, unseen signal, by retrieving the closest matching entry in the library using a nearest-neighbour approach. Nevertheless, the nearest-neighbour approach used in [6] does not generalise well to new data as it does not learn the relationship between signals and microstructure parameters, but simply matches it to an existing entry. Machine learning approaches, on the other hand, are capable of learning this relationship and have a very good generalisation capacity. Recently, machine learning algorithms have been shown to produce state-of-the-art results in a variety of medical imaging problems, with random forests and neural networks among the most widely used algorithms in this field [7, 8]. However, their application to microstructure parameter estimation including τ_i has not yet been studied in simulations or in-vivo.

Here, we experimentally investigate the feasibility of machine learning based computational models with permeability for white matter microstructure imaging, using a random forest and a neural network approach. We explore the potential of our computational framework for microstructure parameter estimation, including the intra-axonal exchange time τ_i, both in clinical and preclinical settings, using synthetic and in-vivo human and mouse data.

We acquire in-vivo data from healthy and MS volunteers using a clinically viable sequence in order to demonstrate the clinical potential of our framework as an imaging biomarker for demyelinating pathologies. As this is the first application of machine learning to microstructure parameter estimation, we seek further validation of our results by performing a direct comparison with histology, as the gold standard in microstructure imaging. Since histological samples of brain tissue are extremely invasive and difficult to obtain in the clinical setting, we use a preclinical mouse model of demyelination. This enables us to perform for the first time a direct comparison between machine learning based estimations of microstructure parameters and histology.

1.1 Scope and objectives

The aim of this thesis is to develop and test the feasibility of machine learning based computational models with permeability for white matter microstructure imaging using DW-MRI. Here, we especially focus on the potential of random forest and neural network based models to act as MS imaging biomarkers based on the permeability associated with the intra-axonal exchange time τ_i. The scope of this investigation covers clinical and preclinical scanner settings and substrates including in-vivo human and mouse white matter tissue.

Our specific objectives are to:

1. investigate the sensitivity of DW-MRI sequences to the intra-axonal exchange time τ_i;

2. develop and assess the performance of a random forest based computational model with permeability using simulations and *clinical data* from healthy and MS patients;

3. develop and assess the performance of a neural network based computational model with permeability using simulations and *clinical data* from healthy and MS patients;

4. experimentally investigate the performance of the random forest model in 2 through direct comparison with histology using *preclinical data* from a mouse model of demyelination;

5. experimentally investigate the performance of the neural network model in 3 through direct comparison with histology using *preclinical data* from a mouse model of demyelination;

1.2 Thesis outline and contributions made

This thesis is structured as follows, with contributions made appearing in italic:

- In **Chapter 2**, we give an overview of the MS pathology, which is followed by background theory on diffusion MRI and the models in diffusion MRI, followed by an overview of the machine learning techniques used throughout this thesis.

- In **Chapter 3** *we investigate the sensitivity of DW-MRI sequences to the intra-axonal water exchange time* τ_i. For this, we use Monte Carlo simulations of DW-MRI signals from substrates mimicking white matter tissue and explore a range of clinical and preclinical scanner settings. We show that long Δs ($\geq 300ms$) play a crucial role in the sensitivity of DW-MRI sequences to exchange time effects. *We conclude that, under realistic noise and tissue relaxation conditions, clinically viable diffusion sequences with long* Δs ($\geq 300ms$) *and high b-values* $\geq 3000s/mm^2$ *are sensitive to the effect of exchange times up to 1000ms, a range covering literature estimates for* τ_i *in the human and rat brain tissue [9, 10, 11, 12].*

- In **Chapter 4** *we introduce the first machine learning based computational model with permeability for white matter microstructure imaging, using a random forest regressor. We demonstrate the idea by testing the model's ability to estimate* τ_i*, the intra-axonal volume fraction* f*, and the intrinsic diffusivity* d *in simulations and in clinical scans from two healthy and two MS patients.* We use Monte Carlo simulations to generate synthetic DW-MRI signals for a range of histologically plausible microstructure parameters for white matter. The computational model then learns a mapping between rotationally invariant features of synthetic DW-MRI signals and microstructure parameters. We find that our model has an excellent performance in simulations and a very good in-vivo scan-rescan reproducibility. *We conclude* *the chapter by showing that the random forest estimates in the two healthy subjects are within the plausible range for white matter tissue, and that the estimates in the two MS patients are consistent with expectations from the pathology of MS lesions (increase in* f *and* τ_i*, and a mild increase in* d*), demonstrating the clinical potential of this new computational framework.*

- In **Chapter 5** *we extend our computational model to a neural network approach and test its feasibility in simulations and in both the healthy and MS patients in Chatper 4.* Similar to the random forest model, the neural network has a very good performance in simulations, provides plausible in-vivo estimates for white matter tissue and detects trends in line with the MS pathology, demonstrating the clinical potential of our neural network approach. Additionally, we find that there is a very good agreement between the healthy in-vivo estimates of the two machine learning approaches. *We conclude by showing that, although the neural network provides a marginal improvement over the random forest when estimating* τ_i *in simulations with noise, the random forest parameters have a better scan-rescan reproducibility for* f *and* τ_i*.*

- From Chapter 6 onwards, we experimentally investigate the performance of our machine learning based computational framework through direct comparison with histology using preclinical data from a cuprizone mouse model of demyelination. In **Chapter 6**, *we study the sensitivity to* τ_i *of the preclinical protocol used to acquire the in-vivo mouse data. Additionally, we optimise and test the random forest model with respect to the new protocol and data. We conclude that our imaging protocol has good sensitivity to* $\tau_i \leq 400$ *ms, and that, within this range, the parameters are well estimated by the random forest, despite the presence of noise and the non-optimised protocol.*

- In **Chapter 7** *we compare the in-vivo estimations of the random forest based model with electronmicrosopy (EM) data from the same mouse model.* In addition to this, we investigate the potential confounding effect of dispersion and axonal swelling. We show that we can rule out the confounding effect of dispersion and axonal swelling on the estimated difference in τ_i between the control and the cuprizone groups, suggesting that demyelination is the main process underpinning the estimated decrease in τ_i. *We conclude the chapter by showing that the random forest estimates of* f *and* τ_i *correlate strongly with EM measurements of intra-axonal volume fraction and myelin thickness (*$\rho_{\tau_i} = 0.82$ *and* $\rho_f = 0.98$*), demonstrating the clinical potential of our computational model as an imaging biomarker for demyelinating pathologies such as MS.*

- In **Chapter 8** *we investigate the performance of the neural network based computational model with permeability on the in-vivo cuprizone mouse model of demyelination.* We find that the neural network estimates in the control group are in good agreement with the random forest estimates, and that both methods estimate a decrease in f and τ_i between the control and the cuprizone groups. These findings are in line with the trends estimated in the MS lesions in Chapters 4 and 5, whilst also being consistent with expectations from the pathology of demyelination. *We conclude by showing that the neural network estimates of* f *and* τ_i *correlate well with histological measurements (*$\rho_{\tau_i} = 0.75$ *and* $\rho_f = 0.70$*). However, the p-values of the correlation are above the significance level and above the p-values of the random forest, which show a much stronger correlation with histology.*

- In **Chapter 9** we discuss the findings and the conclusions from the previous chapters, and suggest potential future developments and applications of our machine learning based computational framework in the clinical and biomedical research environments.

Chapter 2

Background

This chapter provides the background knowledge for the work in this thesis, namely microstructure imaging using DW-MRI and machine learning methods. The first section introduces properties of neuronal tissue and then focuses on MS as our pathology of interest. The second section presents the basic principles of DW-MRI, which is the imaging modality of interest for this work. Next, we describe how the diffusion signal is traditionally modelled in order to characterise tissue microstructure. Finally, we outline the main principles of machine learning, with a focus on random forests and neural networks, as the two relevant approaches for this work.

2.1 Multiple Sclerosis

2.1.1 Neurons and neural tissue

Neurons are the main component of nervous tissue and play an elementary role in the human nervous system. The function of neurons is to transmit, process and receive chemical and electrical signals through the nervous system, building a communication network which is essential for the normal functioning of an organism. As shown in Figure 2.1, a neuron has three main components: a *soma* or a body, which contains the cell nucleus, *dendrites*, which are small protuberances that collect information from neighbouring axons, and an *axon*, which is generally a long, wire-like projection which conducts the neuronal signal towards other nerve cells.

In the human central nervous system, most axons are covered by a protective membranous layer which enhances the conduction speed of the neuronal signals. This is called the *myelin sheath* because it is made of myelin, a substance composed

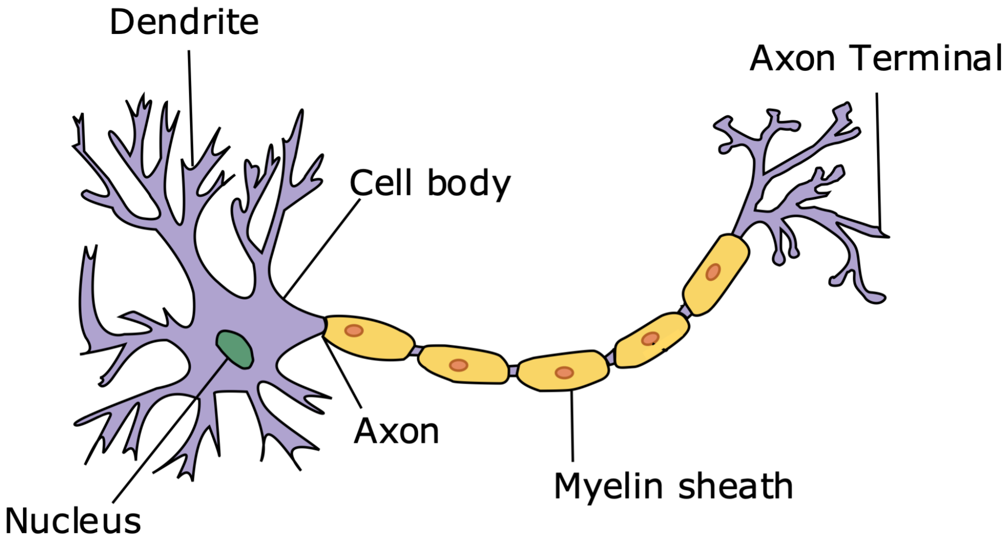

Figure 2.1: Illustration of a neuron, showing the soma (cell body), dendrites and axons. Figure adapted from *https://commons.wikimedia.org/wiki/File:Neuron.svg*

.

primarily of lipids and proteins. The communication speed between neurons is critical to the brain's capacity to perform vital functions, and damages to the myelin sheath can impair this. The myelin sheath is damaged in several neurological conditions, resulting in movement, cognition and sensorial deficiencies. Damage to the myelin sheath can occur as a result of two types of disorders: dysmyelination and demyelination. Dysmyelination is the failure of the organism to produce normal myelin during the myelin formation process [13], and it occurs in conditions such as leukodystrophies [14]. Demyelination, on the other hand, is characterised by damage or loss of previously healthy myelin [13], and it occurs in several demyelinating pathologies, the most well-known of which is MS.

The majority of the myelinated axons are found in *white matter*, which is one of the two main types of neural tissue. White matter is composed of bundles of axons, which connect different regions of *grey matter*, the other main tissue type. While white matter contains the axons of the neurons, grey matter contains the neuronal cell bodies, the dendrites and some shorter axons. An important part of the white matter is the corpus callosum, which is situated in the centre of the brain and is formed of large axonal bundles connecting different grey matter regions in the right and the left hemispheres of the brain. The front (genu) and back (splenium) parts of the corpus callosum connect regions of the grey matter responsible for decision making, auditory and visual tasks [15]. Relative to motion, these are complex tasks that do not need instantaneous transmission. This is reflected directly in their tissue microstructure which is densely packed with small axons and proportionately small myelin sheaths. Conversely, the mid region of the corpus callosum is responsible for motion, which requires instantaneous, but less complex processing. Consequently, the axons in this region are large and covered in proportionately large myelin sheaths in order to help the conductivity of the signals [16]. Therefore, the corpus callosum is described by a high-low-high pattern of axonal density and low-high-low trend of the axonal radius, as shown by histological studies in human [16] and other mammals [17, 18].

2.1.2 Pathology of Multiple Sclerosis

MS is a demyelinating and inflammatory disease which causes the breakdown of the myelin sheaths surrounding the axons in the brain and the spinal cord, forming areas of demyelinated tissue known as *lesions*. The majority of myelinated axons are found in white matter, and thus most MS lesions affect this tissue type. MS causes the loss of the glial cells which produce and maintain the myelin sheath. This results in the thinning or loss of the myelin sheath, and, in later stages of the disease, in axonal loss [19]. Sometimes, in the early stages of the disease, remyelination takes place in the affected axons, however, eventually a scar-like plaque is build up around the damaged axons. In addition to the visible, well-characterised lesions, there are also regions of the tissue called *normal appearing white matter* (NAWM), which usually precede the MS lesions and, despite their normal appearance, are characterised by abnormalities such as axonal injury [20].

The debut of the condition in around 85% of patients manifests through several acute neurological episodes followed by functional recovery [21]. This form of the condition is called *relapsing remitting* MS (RRMS) and is characterised by severe inflammation, demyelination and the breakdown of the blood-brain-barrier [22]. MS usually manifests in the RRMS form for 10-15 years after which the neurological episodes are replaced by irreversible damage and disability during the *secondary progressive* stage of the disease (SPMS). In the remaining 15% of cases, MS manifests from the onset as a progressive disorder and is called *primary progressive* MS (PPMS).

MS is poorly understood and the cause of the disease is not known. Nevertheless, it is believed to be caused either by an immune response or by the failure of the myelin-producing cells due to a combination of genetic and environmentally driven factors [23, 24].

2.1.3 Diagnosis and Treatment

Despite not being fatal, MS reduces life expectancy and drastically affects the quality of life, often resulting in severe disability [19]. The diagnosis of MS is difficult to confirm, especially in the early stages of the disease, because the symptoms are common to those of other diseases [25]. The current diagnosis criteria for MS is based on the revised McDonald criteria, which relies on the detection of MS lesions using magnetic resonance imaging (MRI) as well as other clinical tests to rule out other demyelinating diseases [26, 27]. MRI scans are used to detect in space (DIS) and time (DIT) the developtment of the lesions in the central nervous system of the patient. Demyelinated regions of the brain tissue appear on MRI scans as white dots, as shown in Figure 2.2.

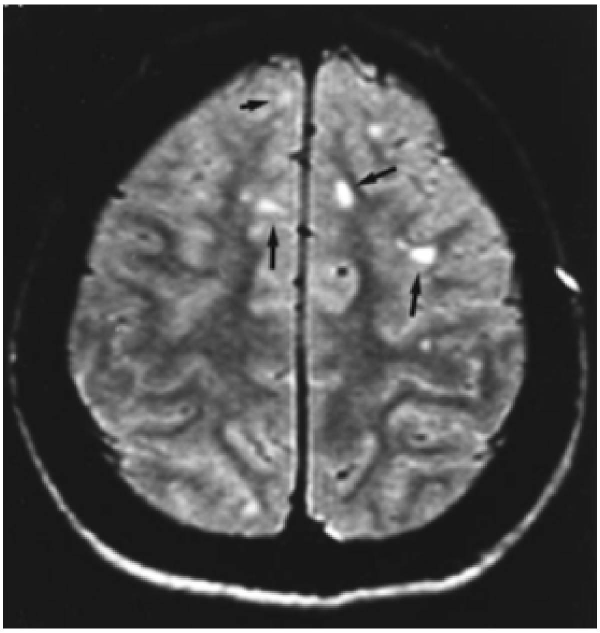

Figure 2.2: An MRI image of the brain showing MS lesions in the frontal and parietal lobe. The lesions are indicated by black arrows and appear as white patches. Image obtained from [28].

While an MS diagnosis can also be establlished using histology, which allows for the close microscopic examination of tissue samples [29], this is an invasive process, and very frequently tissue samples are difficult or even impossible to obtain from the brain. Unlike histology, MRI is a non-invasive, non-ionising technique which can provide structural information about the tissue [30]. While conventional MRI techniques can be used to detect MS lesions, the sensitivity and specificity is often low and can result in an underestimation of the extent of the disease [28]. Recent MRI techniques aim address this problem by allowing the quantification of relevant structural information about the tissue. Quantitative MRI techniques allow for a better characterisation and diagnosis of pathologies, including MS, by extracting relevant information from normal and diseased tissue [31]. The most promising quantitative MRI techniques for MS are: *quantitative magnetisation transfer imaging* - shown to be able to detect and quantify subtle brain tissue changes that correlate with disability [32], *diffusion tensor imaging* - sensitive to the evolution of tissue damage within MS lesions [33] , *magnetisation sodium MRI* - targeting the pathological deposition of sodium [34], *myelin water imaging* - measuring changes in the myelin water fraction [35], *quantitative susceptibility mapping* - estimating the iron accumulation in lesions [36].

Several therapies aimed at functional recovery following neurological episodes and at the prevention of relapses have been developed as management therapies [19]. Even though there is no cure for MS [37], the current pharmacological management therapies of RRMS have proven to be helpful. On the other hand, there is a lack of therapies that can predict and prevent the transition of RRMS to SPMS or manage PPMS. There are currently several advanced clinical trials for medication addressed at the former two types of MS, however, the need to develop sensitive and specific biomarkers to help with the diagnosis and treatment of MS. In this thesis, we focus on the use of Diffusion-Weighted MRI for developing sensitive and specific biomarkers to aid with the understanding, monitoring and diagnosis of MS.

2.2 Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI)

The microstructure imaging modality that we focus on in this work is DW-MRI. In this section, we begin by briefly describing the main principles of magnetic resonance imaging, and then focus on DW-MRI, for which we explain how the image contrast is generated.

2.2.1 Basic principles of Magnetic Resonance Imaging

MRI is a widely used non-invasive and non-ionising imaging technique which generates high resolution images that allow the probing of internal structures and functions of biological tissue. MRI images are generated using the signal produced by the interaction between certain atomic nuclei when placed under a magnetic field. For this interaction to occur, the nuclei must posses an odd number of protons or neutrons, resulting in a non-zero spin and magnetic moment. Such nuclei are abundantly found biological tissue, especially in the hydrogen atoms of water and fat, making hydrogen the nucleus of interest in most MRI applications.

The interaction of the nuclei with the magnetic field is due to a physical property of the nuclei called *nuclear magnetic moment*, μ, that arises from the rotation of the nuclei around their axis. In the absence of a magnetic field, the direction of μ is random for each nucleus and the cumulative net magnetisation of all the nuclei present in the tissue is zero. Once a strong magnetic field, B_0, is applied, all magnetic moments align either parallel or anti-parallel to it, and precess (i.e. rotate) about the axis of B_0 with a frequency $\omega_0 = \gamma \mathrm{B}_0$, known as the Larmor frequency [38], where γ is a nucleus-dependent ratio known as the gyromagnetic ratio. This results in a non-zero net magnetisation M, equal to the sum of the nuclei's individual magnetic moments μ_i, $\mathrm{M} = \sum_{i=1}^{n} \mu_i$, which is aligned with B_0 and specific to every tissue type.

Since M is very small in comparison to B_0, its signal is non-distinguishable from that of B_0. Therefore, to create a measurable signal, M is tipped in the transverse plane (i.e. perpendicular to B_0) using a radiofrequency (RF) pulse. Once the RF pulse is switched off, the net magnetisation M will start to realign with the external magnetic field in order to return to its equilibrium state. This process is described mathematically through a set of equations known as the Bloch equations [39]. The process through which M realigns with B_0 in the longitudinal plane of B_0 is known as longitudinal or T1 relaxation, while the process through which the net magnetisation decreases in the transversal plane is known as T2 relaxation. Both relaxation times depend on the applied magnetic field as well as on the tissue type. For example, for a 1.5T magnetic field, the T1 and T2 relaxations of white matter are $\approx$880 ms and $\approx$70 ms [40]. By measuring the relaxation times, we can infer details about the tissue domain and produce images that delineate different tissue types.

The signal we measure during the NMR experiment is the transverse magnetisation generated by the RF pulse. Following the RF pulse, M precesses about the axis of B_0 at the Larmor frequency and due to Faraday's law of induction, this oscillating magnetisation induces a measurable oscillating voltage in the receiver coil, which is the NMR signal. In order to localise the NMR signal and extract information about its spatial distribution to construct 3D images, different magnetic field gradients are used. Three different gradients are used to encode spatial information in the NMR signal: G= (G_x, G_y, G_z). They are usually applied in three directions: slice encoding (z direction), frequency encoding (x direction) and phase encoding (y direction) [41], and are generated by coils inside the scanner. Once the gradients are applied and spatial information is encoded in the NMR signal, the final step is to reconstruct the 3D image from the measured raw data. This is done by applying a two-dimensional inverse Fourier transform to each slice of the volume. The results of the Fourier transform have both a real and imaginary component, which can be combined to form both magnitude and phase images.

The sequence, duration and strength of radiofrequency pulses and gradients that result in a set of images with a particular appearance is commonly referred to as an 'pulse sequence' or 'MRI sequence'.

There are several sources of noise that affect MRI images, such as the thermal noise of the RF coils used to measure the signal [42, 43] or the noise coming from the spatial reconstruction process. The overall noise in an MRI image is characterised by a Rician distribution and commonly reported through a metric known as signal-to-noise ratio (SNR), defined as: SNR $= \frac{S_{NF}}{\sigma}$, where S_{NF} is the true, noise-free underlying signal and σ is the standard deviation of the noise. For SNRs>5, the Rician distribution is almost indistinguishable from a Gaussian [44] and thus models that assume Gaussian noise under this SNR condition remain accurate.

Another key advantage of MRI is its versatility compared to computerised tomography (CT) and positron emission tomography (PET), given by the many different types of contrast available. Different contrasts are achieved by using different sequences of RF pulses and magnetic field gradients, known as MRI sequences. The most common types of contrast are proton-density, T1, T2, blood oxygenation level dependent (BOLD) or diffusion-weighted. Each type of contrast reflects different properties of the tissue: the contrast in BOLD images measures changes in blood oxygenation, while DW-MRI measures the diffusion of water molecules in tissue. As the MRI contrast is linked to tissue properties at the microstructural level, one of the ways to obtain information about these properties is by fitting models to the MRI signal. Examples of such properties are the cell size or the fraction of intra- to extracellular space in the underlying tissue. Measuring tissue properties at the microscopic level is particularly important in research as it allows the close monitoring of disease progression by correlating changes in parameters with disease symptoms [45]. This type of imaging is especially relevant for pathologies with unknown cause or cure such as MS.

2.2.2 Diffusion Weighted Imaging

Diffusion theory and contrast

DW-MRI is a quantitative MRI method that measures the diffusion of water molecules within tissue and generates images using the contrast in diffusion [46]. As 70% of white matter is comprised of water [47], DW-MRI is particularly suited to imaging this type of tissue.

Diffusion is the process through which particles move randomly from an area of high concentration to an area of low concentration. In unrestricted mediums with no concentration gradient, the particles continue to move randomly, a pro-

Figure 2.3: Microscopic view of a bundle of axons in the white matter. Image: Prof. George Bartzokis.

cess known as Brownian motion, which was discovered by R. Brown and described mathematically in the 3-dimensional space by Einstein as:

$$< R >= \sqrt{6DT_d} \tag{2.1}$$

where T_d is the diffusion time, D is the diffusion coefficient [48], and R is the average displacement of a molecule from its start position to its end position after T_d has passed.

However, in biological tissue, unrestricted diffusion is rare due to the barriers posed by tissue structures such as cell walls or myelin sheaths, which restrict diffusion along certain directions. Instead, the diffusion pattern in biological tissue is determined by the underlying tissue structure. For example, in the case of the white matter, which is formed of long thin axons that act like tubes, the diffusion takes place mainly along the principal axis of the axons and is restricted in the direction perpendicular to the axons. Figure 2.3 depicts a microscopic view of white matter, where we can observe the long thin axons along which the water diffuses.

The diffusion pattern along one predominant direction, such as along the length of the axons, is called anisotropic. Conversely, when the water molecules encounter no barriers, such as in the cerebrospinal fluid in the brain ventricles, the pattern of diffusion is called isotropic. In the extra-axonal space, the water molecules are hindered by the cell walls, resulting in a movement pattern known as hindered diffusion. Consequently, by measuring diffusion we can infer information about the underlying tissue structure.

Single diffusion encoding (SDE) sequences

In order to obtain MR images sensitive to diffusion, an additional set of gradients, known as diffusion sensitising gradients [49], needs to be applied. A standard imaging sequence for achieving this was developed by Stejskal and Tanner [50] and Carr and Purcell [51], known as 'single diffusion encoding (SDE)' or 'pulsed gradient spin-echo' (PGSE). In this thesis, we will refer to this sequence as SDE, which was recently agreed upon in the diffusion MRI community [52]. This method applies two symmetrical diffusion gradients, with equal duration δ and gradient strength G, usually of rectangular shape and of strength higher than the imaging gradients. The two gradients are applied on either side of a 180° RF pulse, such that the time from the start of the first gradient to the start of the second gradient is Δ, as shown in Figure 2.4. On clinical scanners, the diffusion gradient strength can usually go up to 40-60 mT/m, while on preclinical or Connectom scanners the gradient strength is much higher, up to 300 mT/m.

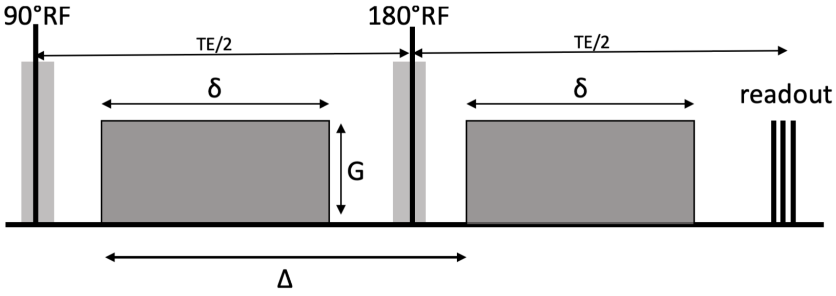

Figure 2.4: Schematic view of a basic single diffusion encoding (SDE) sequence with gradient amplitude G, gradient duration δ and diffusion time Δ.

During the application of the diffusion gradients, the spins in the tissue experience different precession frequencies depending on their position with respect to the gradients. At the end of the diffusion gradient, the spins will have accumulated a net phase ϕ_1, after which they continue to move randomly due to Brownian motion. Next, the 180° RF pulse is applied, inverting the phase of the spins. When the second diffusion gradient is applied, this has the same effect as the first gradient, causing a second location-dependent net phase ϕ_2, but this time in the opposite direction. If the spins had not moved at all in between the two diffusion gradients, the two phase offsets would have cancelled each other and returned to their initial state. But, since the spins experience Brownian motion between the two gradients, ϕ_2 cancels only part of ϕ_1, resulting in a net phase offset $\phi_2 - \phi_1 \neq 0$, which causes an attenuation of the MR signal [53]. The attenuated signal is diffusion-weighted as it characterises the amount of diffusion that takes place within tissue along the direction of the diffusion gradient (i.e. the more the signal drops, the greater the water diffusion in the direction of the gradient). By applying diffusion gradients in different directions equally spaced along a sphere, a complete three dimensional image of water diffusion within the tissue is created [46].

The signal attenuation is controlled by the diffusion weighting or the b-value of the sequence, which is a combination of the sequence parameters, and is defined as:

$$b = \gamma^2 \mathrm{G}^2 \delta^2 (\Delta - \frac{\delta}{3}) \tag{2.2}$$

where γ is the gyromagnetic ratio of the tissue. The higher the b-value, the higher the diffusion weighting. For free diffusing molecules, the b-value is related to the signal attenuation via the equation:

$$\mathrm{S(b, D)} = \mathrm{S}_0 e^{-\mathrm{bD}} \tag{2.3}$$

where S_0 is the signal without diffusion weighting and D is the diffusion coefficient. Since in biological tissue D is affected not just by the diffusivity of the medium, but also by the tissue environment, it is usually referred to as the 'apparent diffusion coefficient' (ADC).

STEAM sequences

Another DW-MRI sequence is the Stimulated-Echo Acquisition Mode (STEAM). STEAM sequences can be advantageous over the SDE sequence in Figure 6.2.1 by allowing the probing of longer diffusion times [54]. This is achieved by splitting the 180° RF pulse in the SDE sequence into two separate 90° RF pulses, as shown in Figure 2.5. The first 90° RF pulse flips the net magnetisation in the transversal plane, where the signal decays with T2 rate until the next RF pulse. Once the second 90° RF pulse is applied, the net magnetisation is flipped back in the longitudinal plane. During the time between the second and the third 90° RF pulses, known as the mixing time τ_m, the signal decays with T1 rate. Then, the third 90° RF pulse tips the net magnetisation back into the transversal plane, where it again experiences T2 relaxation. In tissue where T1 $\gg$ T2, such as in white matter [40], the signal decays less during the mixing time τ_m, allowing for longer diffusion times Δ to be probed. In contrast, throughout an SDE sequence, the net magnetisation experiences T2 relaxation throughout the whole diffusion time. Consequently, STEAM sequences are commonly used in tissue with short T2, such as muscle or cartilage [55], or in applications where long diffusion times are necessary to ensure sensitivity to certain microstructure parameters, such as axon diameter [56].

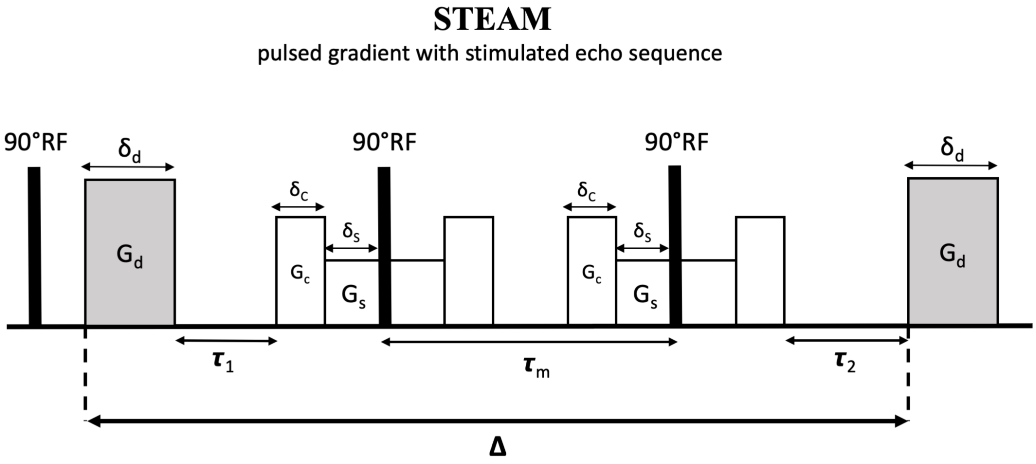

Figure 2.5: Schematic view of Stimulated-Echo Acquisition Mode (STEAM) sequence.

In addition to the two diffusion gradients G, the STEAM sequence includes two crusher gradients G_c, used to terminate the signal decay before the next RF pulse, and two slice-select gradients G_s. The additional gradients have been shown to introduce an unwanted diffusion-weighting bias which has to be accounted for [54]. In contrast, in an SDE sequence, the effect of the additional gradients is usually negligible in practice due to their short duration [54].

The diffusion sensitisation of the MRI signal is the first step towards obtaining tissue microstructure information. The next step is fitting models to the DW-MRI signal measurements in order to estimate specific tissue parameters. The following section presents the most widely used tissue models in DW-MRl.

2.3 Modelling the DW-MRI signal

The versatility of DW-MRI as an imaging technique stems not only from the wide range of contrasts available, but also from the many different ways to model the diffusion signal in order to extract meaningful information about the underlying tissue. Analytical models of the DW-MRI signal are fit to MRI measurements in order to estimate specific tissue parameters from the signal. This section introduces the most commonly used analytical models of the DW-MRI signal and discusses their advantages and limitations. In the first part, we introduce the diffusion tensor imaging (DTI) model [57], as the most widely used DW-MRI model [58]. DTI encodes the cumulative effect of tissue features, but does not fit specific microstructural features. In the second part of this section, we introduce *biophysical* or *multi-compartment models*, which represent the tissue as a collection of key geometrical features called ’compartments’ and fit direct microstructural indices such as axon size or intracellular volume fraction.

2.3.1 Diffusion Tensor Imaging (DTI)

The ADC model in Equation 2.3 assumes free diffusion in an isotropic and homogeneous environment. However, this does not accurately describe diffusion in anisotropic tissue such as white matter or muscle fibres. DTI, on the other hand, uses a 3D Gaussian model to allow for diffusion to vary along different directions [59], providing a more accurate description of free diffusion in anisotropic environments. DTI replaces the scalar D from Equation 2.3 with a 3x3 symmetric matrix

D, known as the diffusion tensor:

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{yx} & D_{xz} \\ D_{xy} & D_{yx} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} \tag{2.4}$$

where $\mathrm{D}_{xx}, \mathrm{D}_{yy}$ and D_{zz} are the diffusivities in the x, y and z directions, and $\mathrm{D}_{xy}, \mathrm{D}_{yz}$ and D_{xz} represent the covariances.

The measured signal S is related to the diffusion tensor by:

$$S = S_0 \exp(-b \hat{\mathbf{G}}^T \mathbf{D} \hat{\mathbf{G}}) \tag{2.5}$$

where S_0 is the signal with no diffusion weighting, b is the diffusion weighting (from Eq. 2.2), and $\hat{\mathbf{G}}$ is the gradient direction.

As **D** is a symmetric matrix, equation 2.4 has only six unknowns and so in order to be able to fit this model to the signal, we need at least 6 measurements for the different gradient directions and an additional 7th in order to measure S_0. However, in practice, the noise in the data means that around 20-30 measures of isotropically distributed gradients are necessary [60]. The estimates can then be obtained using a multitude of either linear or non-linear optimisation techniques [57, 61], although non-linear optimisation has been shown to be less biased [60].

After computing the diffusion tensor **D** from the data, we can estimate the principal directions of diffusion and their corresponding diffusivities by performing an eigendecomposition of **D** to find the eigenvectors $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ and their eigenvalues $(\lambda_1 \geq \lambda_2 \geq \lambda_3)$. The principal eigenvector $\mathbf{e}_1$ estimates the dominant fibre direction in a voxel. The size of the eigenvalues determines the shape of the tensor, as illustrated in Figure 2.6. The tensor on the left depicts anisotropic diffusion and is called *prolate*. In a prolate tensor, one eigenvalue is considerably larger than the other two, indicating anisotropic diffusion, as we expect to see along the main tracts in white matter. The middle tensor is called *oblate* and could, for example, be characteristic of crossing fibres, indicated by the two dominant eigenvalues. The right most tensor resembles a sphere and has approximately equal eigenvalues, indicating isotropic diffusion. This can be representative of either grey matter, where the diffusion is small, but with no dominant directions, or of cerebrospinal fluid, where the diffusivity is large and isotropic.

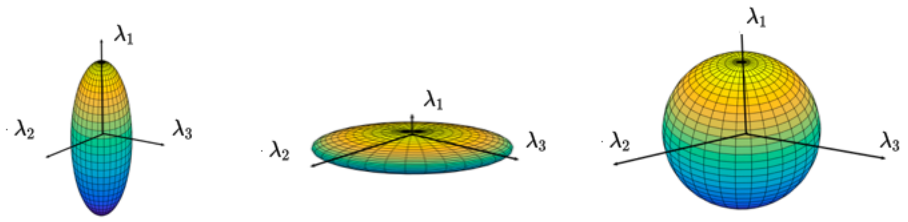

Figure 2.6: Diffusion tensors (from left to right): The prolate tensor characteristic of anisotropic diffusion, the oblate tensor representative of diffusion along a plane (for example due to crossing or bending fibres) and the sphere tensor representing isotropic diffusion.

Another way to describe the diffusion tensor is by deriving rotationally invariant features from the eigendecomposition of D. There are two commonly used rotationally invariant metrics: fractional anisotropy (FA) and mean diffusivity (MD). MD measures the mean amount of diffusion in a voxel and is equal to:

$$\mathrm{MD} = \lambda_1 + \lambda_2 + \lambda_3, \tag{2.6}$$

where λ_i, $i = 1, 2, 3$ are the eigenvalues of $\mathbf{D}$. MD is high in the cerebrospinal fluid, as the water molecules diffuse freely in the absence of barriers, and lower in grey and white matter as diffusion is restricted by axons and glial cells. FA measures the fraction of diffusion anisotropy in a voxel. This takes values between 0 and 1 and is calculated as:

$$FA = \frac{3}{2}\sqrt{\frac{\sum(\lambda_i - MD)^2}{\sum \lambda_i}} \tag{2.7}$$

FA is low in the cerebrospinal fluid and grey matter as there are no dominant directions of diffusion, and takes higher values in white matter, where diffusion happens mainly along the axons.

There are several limitations of this model. Firstly, DTI assumes a trivariate Gaussian model of diffusion, which does not hold in the axons of white matter where the water diffusion is restricted [62]. Secondly, even though DTI offers more specific information than ADC, it remains too simplistic to allow for the estimation of more specific microstructure parameters. For example, in the case of the oblate tensor, the model does not offer sufficient information to determine the main two fibre directions, as it might be necessary in tractography applications. In addition to this, while FA and MD do measure changes in tissue microstructure, it remains impossible to disentangle the contribution of individual parameters such as axon diameter, permeability or fibre dispersion [63]. Detecting changes in MD and FA is not always sufficient as different combinations of tissue parameters could have the same effect on these metrics.

As established in the previous sections, the quantification of tissue parameters is of paramount importance for developing more sensitive and specific imaging biomarkers. In order to achieve this, there is a need for models that directly relate the DW-MRI signal to more specific tissue parameters.

2.3.2 Multi-compartment white matter models without exchange

Biophysical or multi-compartment models that directly relate tissue specific parameters, such as axon radius, to the DW-MRI signal have been proposed to overcome the limitations of the DTI model. Multi-compartment models build a geometrical representation of the tissue as a collection of compartments, each representing a key feature of the underlying tissue microstructure. For example, the axons in the white matter can suitably be represented by an elongated geometrical shape such as a cylinder. The fluid-like aspect of the cerebrospinal fluid can be represented using a sphere, in which water diffuses freely and isotropically, not encountering any barriers. The complexity of the model can be varied by either adding additional compartments to represent more features or by choosing a more complex and accurate geometrical shape for the same features. For example, one can use a zero-radius cylinder to represent an axon or create a more complex model by adding an extra parameter to the model - the axon radius. A trade-off arises here between simpler models, which are easier to fit, and more complex models, which have more parameters and require longer and less clinically-feasible imaging protocols.

A taxonomy of multi-compartment models together with an attempt to rank them according to the Bayesian Information Criterion (BIC) [64] is presented by Panagiotaki et al. [65, 66]. The remaining of this section presents several models commonly used to fit DW-MRI data, ranging in complexity. Essentially, these models are specific cases of the taxonomy framework in [65] and we will adopt this terminology to refer to different compartments: *stick* for a zero radius cylinder, *cylinder* for a non-zero radius cylinder and *gamma-distributed cylinders* to refer to a collection of axons with radii drawn from a gamma distribution, *ball* to refer to an isotropic tensor and, finally, a *zeppelin* to refer to a cylindrically symmetric tensor [65].

One of the first compartment models was introduced by Stanisz et al. [67] and applied to the bovine optic nerve. This model comprises three compartments: prolate tensors representing axons, spheres representing glial cells, and hindered diffusion with exchange for the extracellular space. The model allows for the estimation of microstructure features such as the average axon diameter and length, the glial cell size and the intra and extra cellular diffusivities. Stanisz et al. apply their model to the bovine optical nerve, and the parameter estimates were shown to be in good agreement with histology values [67]. Nevertheless, due to its complexity, the fitting of the model requires not only a great number of measurements, but also a high quality dataset with SNR$>$1000, which is practically impossible to acquire in-vivo in the clinical setting, illustrating the need for simpler, more clinically viable models.

The *ball and stick* model [68] is a much simpler model of the DW-MRI signal than Stanisz et al.'s three pool model. The ball and stick model assumes that there are only two compartments, the intracellular (stick) and the extracellular (ball) compartments, as shown in Figure 2.7. The intracellular compartment represents the axons as zero-radius cylinders and has anisotropic diffusion, while the extracellular compartment has isotropic diffusion, with both compartments having diffusivity D. The measured signal is calculated as a weighted sum of the signals from each compartment, according to the intracellular volume fraction f, computed as the fraction of the signal due to intra-axonal water. The relationship between the parameters and the DW-MRI signal expressed as:

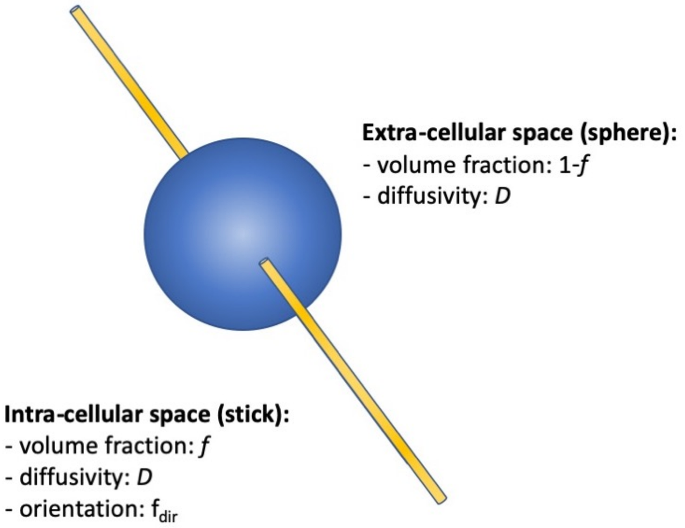

Figure 2.7: Schematic view of the Ball and Stick two-compartment model: the intracellular space is represented as zero-radius cylinders and the extracellular compartment as a sphere with isotropic diffusion.

$$S = S_0(f(\exp(-bD(\hat{f_{dir}} * \hat{G}) + (1-f)\exp(-bD)))) \tag{2.8}$$

where $\hat{f_{dir}}$ is the fibre direction and the rest of the symbols are consistent with the previously defined notation.

While the ball and stick parameters are more descriptive of tissue microstructure than FA or MD, more informative parameters such as axon diameter continue to remain elusive. The next set of models relax some of the ball and stick model assumptions to provide even more specific microstructural features.

The first compartment model that assumes non-zero radius cylinders is the composite hindered and restricted model of diffusion (CHARMED) [69], which fixes the distribution of axon diameters and intracellular diffusivity to typical values for axons in the spinal cord. However, this model still does not allow for the direct estimation of the axonal size. This limitation led to the development of the AxCaliber model [70], which extends the CHARMED model and estimates the axon diameter distribution assuming a known fibre direction. Based on histological observations [69], AxCaliber represents axons as parallel cylinders with non-zero radius following a gamma distribution, as shown in Figure 2.8. Similarly to Stanisz's model, the diffusion in the intracellular compartment is restricted. The

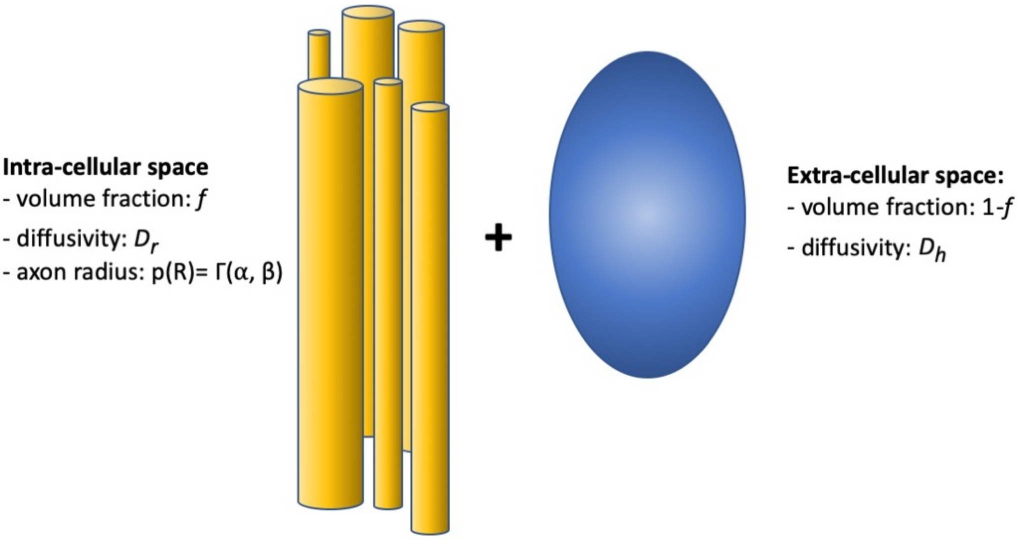

Figure 2.8: Schematic view of the AxCaliber two-compartment model: the intracellular space is represented as parallel, gamma-distributed, non-overlapping cylinders and the extracellular compartment as a zeppelin.

second compartment is a zeppelin [65] aligned with the cylinders and represents the hindered diffusion in the extracellular space. While ex-vivo [70] and in-vivo studies [71] show that there is a good agreement between the model estimates and histology, there are also limitations which prevent the model from being widely used in clinical in-vivo human studies. Firstly, the data acquisition is very long and uses high gradient strengths not available in the clinical setting. Secondly, AxCaliber assumes that the fibre orientation is known, which is an unrealistic assumption in clinical practice.

Alexander [72] aims to address the limitations of AxCaliber by developing an optimisation framework (Active Imaging) which produces clinically feasible and orientationally invariant imaging protocols. Active Imaging is designed to minimise the variance of axon diameter estimates, and was introduced together with the ActiveAx model, which represents axons as randomly packed, single-radius cylinders. The extracellular compartment of ActiveAx is modelled as a zeppelin diffusion tensor, with the principal eigenvector parallel to the cylinders, as shown in Figure 2.9. ActiveAx permits the estimation of the intracellular volume fraction, the fibre orientation and the axonal diameter index α, a parameter that has been shown to correlate with the mean volume-weighted axon diameter. *ActiveAx* was shown to provide good ex-vivo monkey brain and in-vivo human brain estimates [73]. Although the

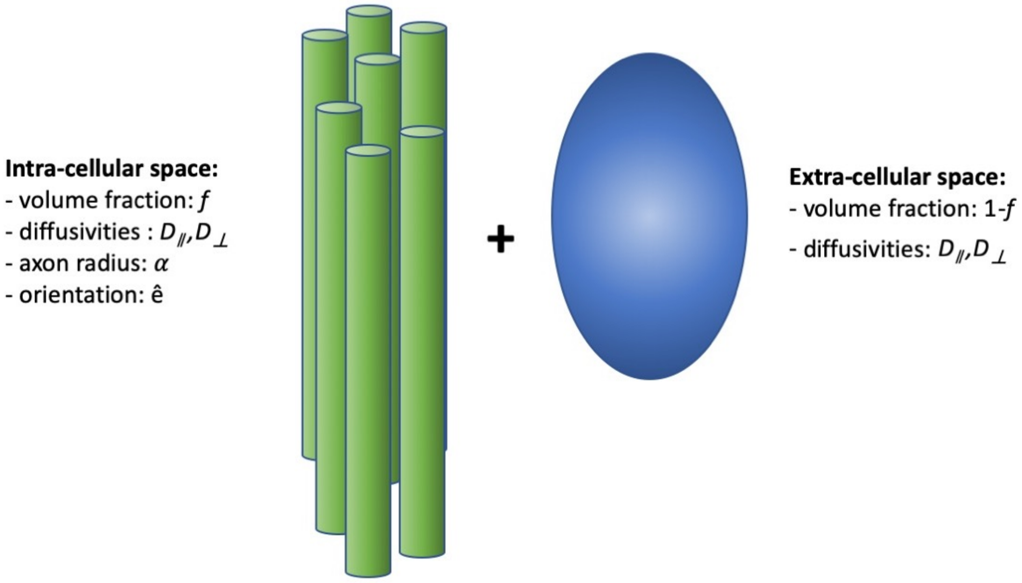

Figure 2.9: Schematic view of the ActiveAx two-compartment model: the intracellular space is represented as parallel, non-overlapping, single radius cylinders and the extracellular compartment as a zeppelin.

model assumes single-radius axons via the α parameter, *ActiveAx* provides a first window into more informative tissue microstructure parameters.

The models introduced so far represent axons as straight cylinders, without accounting for bending or fanning, as it is often the case in some parts of the white matter [74]. While the assumption of straight parallel cylinders holds in regions of the white matter such as the corpus callosum [16] or the spinal cord [75], this assumption does not hold in other regions of the white matter, in which axonal tracts can fan out and disperse [76]. One of the first models that accounts for fibre dispersion was developed by Jespersen et al [77]. The models calculates the signal from the intracellular space by integrating over all possible orientations. Nevertheless, despite accounting for fibre dispersion, Jespersen's model leaves out other important microstructure parameters such as the axon radius. Another model which accounts for fibre dispersion is the 'the neurite orientation dispersion and density imaging' model (NODDI) [78], which allows for the estimation of the fibre density and orientation dispersion. NODDI comprises three compartments: a Watson distribution of sticks for the intra-neurite signal, a zeppelin for the signal outside the neurites and a ball for the cerebrospinal fluid. This model is commonly employed by researchers in in-vivo studies [79, 80] due to the straightforward and clinically feasible imaging protocol. NODDI has so far been used to characterise healthy [79, 81] and diseased tissue [80, 82] proving the clinical feasibility of compartment models. Kaden et al [83] replace the symmetrical Watson distribution of NODDI with an asymmetric directional distribution known as Bingham, which allows for an anisotropical distribution along the main direction of dispersion. This is used in order to build a model that better characterises fanning tracts such as the corticospinal tract.

2.3.3 White matter compartment models with exchange

The multi-compartment models discussed above account for various tissue features, however, they do not take into account the effect of membrane permeability. The permeability of the cell membranes influences the diffusion of the water particles and, it is, therefore, expected to have an effect on the DW-MRI signal. The cell membrane permeability is usually measured via the intra-axonal water exchange time τ_i [6, 84, 85, 86], defined as the average time a water molecule spends inside the intra-axonal space represented by the white matter axonal fibres [87]. τ_i is also known as 'the intracellular water lifetime' or 'the residence time', and, throughout this thesis, we will refer to τ_i as the intra-axonal exchange time. τ_i is inversely related to the permeability k of the cell membrane and can be calculated with:

$$\tau_i = \frac{V}{Ak},$$

where V is the cell volume and A is the cell surface area [87, 6, 84, 88].

The first part of this section describes the main analytical models that account for water exchange between the cell membranes. Nevertheless, deriving accurate analytical models with exchange is inherently difficult due to the intractability of the mathematical expressions that incorporate permeability. To address this limitation, there have been alternative numerical approaches which bypass the need for mathematical expressions, which we outline in the second part of this section.

Analytical Models

Analytical models that relax the assumption of restricted diffusion and account for exchange, usually incorporate permeability either in a simplistic way [89, 90] or under assumptions not valid in white matter [85, 91]. An example of this is the simplified permeability model introduced by Cod and Callaghan [89], which assumes that once the water particles leave the intracellular space they cannot re-enter the same compartment. A more recent approach by Grebenkov et al. [90] aims to estimate permeability using high gradients and narrow pulses. Nevertheless, their current results are only valid for compartments significantly larger than those present in white matter.

One of the most well known approaches that explicitly incorporates τ_i is the Kärger model [85]. The model assumes two compartments or 'pools' and accounts for exchange between them by coupling their signals via the τ_i parameter. Similar to the compartment models in the previous section, the total DW-MRI signal is computed as the sum of the signals coming from the two compartments, coupled through the exchange time τ_i. Due to its compatibility with data acquired using clinically feasible protocols (SDE and STEAM), the model is very widely used in practice [67, 92, 6]. Despite its popularity, the assumptions of the Kärger model do not hold in white matter tissue, limiting its capacity to accurately describe exchange times in the brain. The Kärger model assumes that the water pools in the two compartments are well mixed and it does an account for restricted diffusion, an assumption which does not hold in white matter. Fieremans et al. [84] show that while the Kärger model may provide a reasonable description of the DW-MRI signal in the long time limit (i.e. when τ_i is much longer than the time it takes to diffuse across the cell) and when the cells are close to impermeable, the model fails in the case of fast exchange and results in overestimated exchange times and underestimated permeabilities. In addition to this, Nilsson et al. [6] use Monte Carlo simulations with restriction and exchange to show that the Kärger model yields biased estimates of the intra-axonal volume fraction f for exchange times < 350 ms.

Pfeuffer at al [93] and Dietrich et al [94] address one of the limitations of the Kärger model by modelling restricted diffusion in the intracellular space. However, the lack of an analytical model with exchange for the extracellular space has prevented significant improvements to the Kärger model for over 20 years.

Recently, the Apparent Exchange Rate model (AXR) [91, 95] was introduced as an alternative to estimating τ_i via the Kärger model. AXR combines the intra-axonal volume fraction f and τ_i under a single apparent exchange rate parameter $AXR = \frac{1}{\tau_i f}$ [1]. Despite preliminary findings demonstrating AXR's feasibility for estimating τ_i in in-vivo human data [95], there are several limitations of the model which could preclude its use as an in-vivo clinical tool. Firstly, AXR requires imaging sequences with double diffusion encoding, which are less widely available than the conventional SDE required by the Kärger model. Secondly, AXR preserves some assumptions not valid in white matter: it assumes that the water is split between a 'fast' and a 'slow' pool and that the diffusion in each compartment is time-independent [96]. Thirdly, as the volume fraction f is conflated with the exchange time τ_i, it is difficult to disentangle the contribution of each parameter. Finally, the model allows for only two compartments, unless further parameter conflation occurs [97].

The analytical models illustrated so far encode the relationship between tissue microstructure parameters and the DW-MRI signal through an analytical expression. However, for some parameters such as the exchange time or the extracellular space, there is a lack of tractable mathematical expressions that accurately express their relationship to the DW-MRI signal. Because of this, the accurate estimation of such parameters has so far remained elusive and alternative approaches have been developed, which we present next.

Computational Models

Given the inherent difficulties involved in deriving accurate analytical models of exchange with permeable axonal membranes, there have been alternative models which bypass the need for analytical expressions altogether and use simulations to learn how permeability affects the DW-MRI signal. Rather than deriving an analytical expression to encode the relationship between τ_i and the DW-MRI signal, these models adopt a computational approach and use numerical simulations to study this relationship.

A great majority of studies simulating the DW-MRI signals use Monte Carlo Markov Chain (MCMC) algorithms [6, 84, 98], which are a class of methods used for sampling from probability distributions using Markov chains. Markov Chains are stochastic models in which the probability of each event depends only on the previous event. MCMC chains are used extensively for sampling distributions with an unknown form as they bypass the need for analytical expressions. MCMC algorithms are very flexible and can be used to simulate the diffusion motion by placing a sufficiently large amount of spins (or walkers) in a geometrical representation of the tissue and updating their motion step by step over a period of time. At the end of each step update, the phase of each spin is computed according to its position and the sequence settings, and the final signal attenuation is calculated as the sum of the phases of all spins. For example, Hall et al. [98] study the convergence of the diffusion signal in highly complex tissue substrates and include effects that analytical models cannot capture. The main drawback of MCMC simulations is their computational complexity, which renders them unsuitable as real-time techniques. Nevertheless, this limitation can be mitigated by adopting a database approach, in which databases of MCMC simulations are generated a priori.

Nilsson et al. [6] build such a computational approach by generating a library of Monte Carlo simulations of DW-MRI signals and their corresponding ground truth microstructure parameters (the cell diameter, the intra-axonal exchange time, the diffusion coefficient and the intra-axonal volume fraction). Then, given a previously unseen signal (i.e combinations of tissue parameter values not explored in library), they adopt a nearest-neighbour approach to retrieve the closest matching signal in the database by finding the entry that minimises the squared difference to the unseen signal. When comparing the performance of their computational model with that of the analytical Kärger model, they show that their approach improves the accuracy of the estimated parameters, except for the exchange time [6]. Nevertheless, their model has several drawbacks. The main limitation is that the nearest-neighbour algorithm they use is not capable of generalisation (i.e. it cannot infer the parameter estimates for a new, unseen signal). Their approach does not learn the mapping between signals and parameters, but simply creates a dictionary and retrieves the closest match, which in reality could be very far away from their simulated database. Another limitation stems from using the raw signals to perform the matching. This is potentially inefficient, as it requires new libraries to be generated for each acquisition protocol as well as for each fibre orientation.

While the approach of Nilsson et al. simply builds a mapping between synthetic DW-MRI signals and microstructure features, machine learning approaches can potentially learn this relationship and provide better generalisation, especially for parameters which only weakly influence the DW-MRI signal [95]. Machine learning algorithms have been shown to excel in learning to generalise from high-dimensional data [7], offering a promising new approach for microstructure parameter estimation. The next section introduces such machine learning approaches, with a focus on random forests and neural networks as the algorithms of interest for this work.

2.4 Machine Learning

Machine learning is a field of artificial intelligence in which algorithms and statistical models are trained to learn and infer patterns from datasets without using explicit instructions. A high-level definition of machine learning would describe it as 'the field of study that gives computers the ability to learn without being explicitly programmed' (Arthur Samuel, 1959). During the learning stage, a machine learning algorithm uncovers hidden insights or patterns in the training data, which are not always easily detectable by humans. Using the data-driven insights uncovered during training, the machine learning algorithm builds a generalisable model that can predict to new, previously unseen data.

Through bypassing the need for explicit programming as well as being able to generalise to previously unseen data, machine learning is particularly suited and excels in problems for which there are no available solutions using traditional approaches. Machine learning is also suited for problems for which traditional approaches do work, but use a long set of rules or require considerable hand-tuning [99]. Even though the term 'machine learning' was drafted in the late 1950s, recent developments culminating with the rise of deep learning [7] have brought machine learning at the forefront of technological advances: automatic object identification [100], speech recognition [101, 102], product recommendations, autonomous driving, sentiment analysis. Machine learning algorithms have also exceeded human-level performance in playing games such as Atari or AlphaGo, pushing the boundaries of artificial intelligence. In addition to this, machine learning is currently applied in many areas of medicine and medical imaging: from predicting the effect of DNA mutations on diseases [103] to automatic X-Ray based diagnosis [104] and tumor classification [105]. While the progress in the medical field has been slower compared to others fields such as image recognition, the recent advances in computing power facilitate the application of machine learning to real-time diagnosis tasks and enhanced analysis of medical images [106].

Machine learning approaches use a large set of data $\{x_1, x_2, ...x_N\}$, named a *training set*, to optimise the parameters of an adaptive model with respect to a particular task. The model's output is a function $y(x)$, whose parameters depend on the training data and are optimised during the training stage, also known as the learning phase [107]. The function optimisation is done with respect to a task-dependent objective or cost function, by employing numerical optimisation techniques such as gradient descent. The goal of the training process is to increase the generalisation of the algorithm, defined as the ability to correctly perform the task for which the model is optimised given previously unseen inputs. Usually, the generalisation capacity of an algorithm is assessed on an unseen set of data called the *test set*.

Prior to training, there is usually a need to preprocess the dataset given the high-dimensionality and variability of real-life data. During preprocessing, all x_i instances of the training set are usually represented in a common space, designed so that the task at hand is easier to solve and less computationally expensive [107]. For example, given an image recognition task, the pre-processing stage could consist of scaling and translating the images in the training set into a single picture size. If the training set was formed of MRI signals, the pre-processing stage might involve extracting rotationally invariant features of the signal or normalising them.

Machine learning algorithms can be divided into three broad categories, based on whether labels for the dataset are available. By labels, we refer to the correct task outcome associated with each training example. The first type or learning is *supervised learning*, in which the training set labels are made available during training and used to inform the algorithm on the type of output desired by the user. The second type of learning is *unsupervised learning*, in which the training takes place without any labels being provided. Clustering problems belong to this category, such as discovering distinct healthy and control groups within a population. More recently, a third type of learning, known as *semi-supervised*, has emerged. This falls in between the first two types as it makes use of both labelled and unlabelled data. This follows from recent findings that unlabelled data, which are usually easy to produce, when used in conjunction with labelled data, which are usually more time-consuming to generate, can lead to an improvement in performance.

Machine learning approaches can also be categorised into *classification* and *regression* algorithms according to the type of task they are designed for. Classification algorithms are used to assign inputs to a finite number of categories [107]. An example of this is a digit recognition algorithm which assigns images a unique label from 0 to 9. Regression algorithms are used to assign inputs to a continuous variable or to a collection of continuous variables. Following from this categorisation, the task of estimating microstucture parameters falls into the category of regression algorithms.

In this thesis, we adopt a supervised regression approach and the specific machine learning algorithms that we implement are random forests and neural networks, as two of the most widely used regression algorithms in medical imaging. In the remainder of this chapter, we present in more detail the specific random forest and neural network algorithms that we implement. For a more in-depth discussion about the two techniques, we advise the reader to consult [8] for random forest models, and [107] for neural networks as well as for a more general introduction to the field of machine learning.

2.4.1 Random Forests

A random forest [108] is an ensemble of decision trees, in which each tree is trained randomly and independently of each other [109]. Ensemble methods are one of the most powerful machine learning approaches, well known for their robustness to noise and good generalisation [110]. An ensemble model, also known as a *strong learner*, is formed of a collection of less powerful algorithms, known as *weak learners*. In our case, a random forest acts as a strong learner, while the individual decision trees act as weak learners. Even though decision trees were introduced in the 1980s, their popularity has resurfaced in the recent years with the rise of ensemble methods, which have become one of the most widely used algorithms in the medical imaging field [111, 112, 113, 114, 115]. Depending on the task, random forests can either be regressors or classifiers. In this thesis, we use only random forests and decision trees for regression, and the discussion below is focused accordingly.

Decision Trees

A decision tree (DT) is formed of a set nodes and edges grouped hierarchically as a directed graph with no loops. The nodes of a DT can either be internal or terminal (also known as 'leaf' nodes). The internal nodes split the incoming data and pass it on to the nodes on the lower levels according to a split function which looks for the best partition of the data. Terminal nodes store a predictor function which outputs the result of the regression task. The DTs in a random forest are usually binary trees, in which each internal node has exactly two outcoming edges [110]. Figure 2.10 shows an example of a binary decision tree, which we exclusively use in this work.

TRAINING. During the learning phase, the training data is passed through the DT, starting at the root node towards the terminal nodes. At each node, the DT searches for a partition of the incoming data such that having separate partitions on either side of the node improves the estimation. If such a partition exists, the node is split and two child nodes are added on the level below. This procedure is repeated for every child node until splitting the data into smaller partitions does not improve the estimation anymore. If no better partition is found, the node becomes a terminal node. Therefore, the structure of a DT is determined by the complexity of the training dataset and is set during training.

Conceptually, the training process is equivalent to hierarchically splitting com-

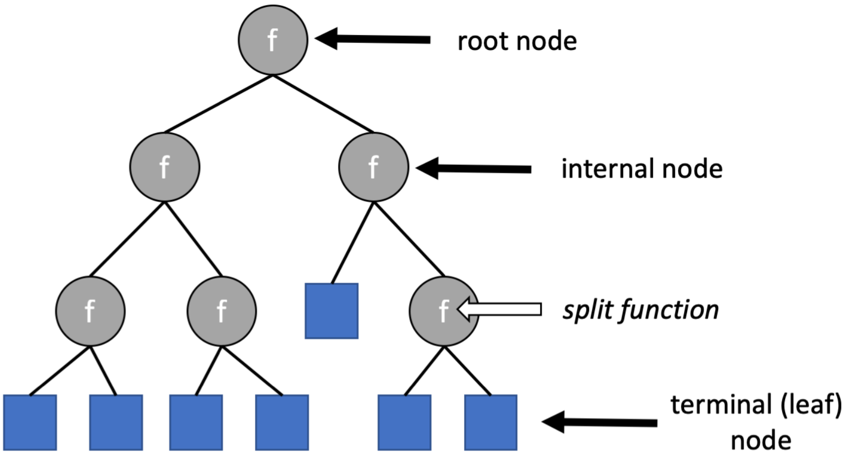

Figure 2.10: Schematic view of a decision tree.

plex tasks into simpler ones [110]. Mathematically, the training process is guided by two functions: the split function of each node and the optimisation or cost function. The split function of each node stores the best partition at that node. The split function parameters are the features and thresholds according to which the data is split. The cost function is used to determine the best split at each node. The optimisation function searches for the feature-threshold pairs (f_i, t_{f_i}) that produce the best split.

Here, we use the Classification and Regression Tree (CART) algorithm cost function, introduced by Breiman [116]. The CART cost function is defined as:

$$J(f_i, t_{f_i}) = \frac{m_{\text{left}}}{m} MSE_{\text{left}} + \frac{m_{\text{right}}}{m} MSE_{\text{right}} \tag{2.9}$$

where $m_{\text{left/right}}$ is the number of training instances in the left/right subset and 'MSE' stands for the 'mean-squared-error' [99]. The smaller the cost $J(f_i, t_{f_i})$, the better the split. In a regression tree, each node outputs a continuous value equal to the average of the labels of all training instances associated to a node, and is calculated as:

$$\hat{y}_{\text{node}} = \frac{1}{m_{\text{node}}} \sum_{i \in \text{node}} y^{(i)} \tag{2.10}$$

where m_{node} is the number of training instances associated to the node and $y^{(i)}$ is the label of training example i. The MSE in Equation 2.9 for each node is calculated as the sum of squared differences between the output of the node $\hat{y}_{\text{node}}$ (Eq. 2.10) and the true labels of the training instances associated with the node:

$$MSE_{\text{node}} = \sum_{i \in \text{node}} (\hat{y}_{\text{node}} - y^{(i)})^2. \tag{2.11}$$

The CART optimisation algorithm is part of a class of algorithms called 'greedy', a heuristic which aims to find the global optimum by making locally optimal choices at each stage (i.e. we greedily search for a split that optimises the cost at the each node without checking whether it also leads to the lowest cost at other levels) [99]. Even though greedy algorithms are not guaranteed to converge to the global optimum, they have been shown to generally converge to a sufficiently good approximation of the global optimum. Employing a problem solving heuristic such as a greedy algorithm is needed because finding the optimal tree is an NP-Complete problem, intractable even for very small training sets [99].

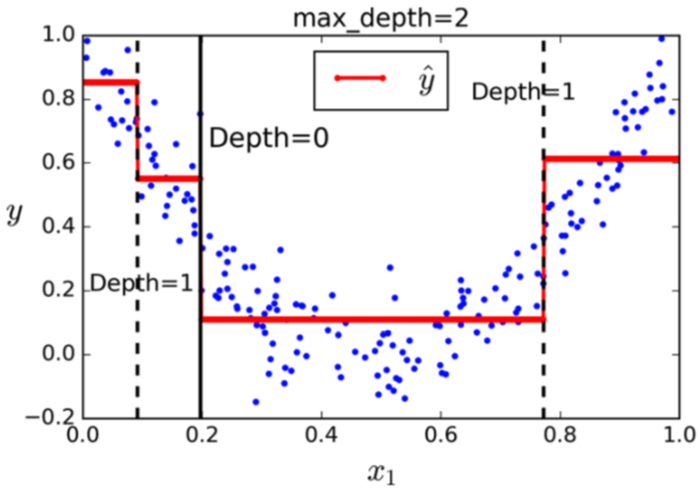

Figure 2.11: Estimations of a regression decision tree. The blue points represent the training examples (y) and the red line represents the estimated values ($\hat{y}$). In a DT, the estimations are calculated as the average of the training examples associated with each node. The image is obtained from [99].

There are several criteria according to which training can be stopped. One criteria is to stop training once the CART algorithm cannot find a split that minimises the MSE. This option usually leads to overfitting as the DT can be grown to fit the training data perfectly. To prevent this, several other approaches have been developed. One example is to stop training when the tree depth is grown beyond a certain value (denoted by the 'max_depth' parameter in Figure 2.11), but there are additional ways to achieve this and a list of other options can be found in the scikit-learn online documentation *http://scikit-learn.org/stable/documentation.html*.

After training is completed, a DT estimates a new value by passing the new instance through the tree nodes until a terminal node is reached. The estimated value is computed as the average of all the training instances assigned to that node. This process is illustrated in Figure 2.11, where the estimated values (the red line) in each node are the average of the training instances associated with that node.

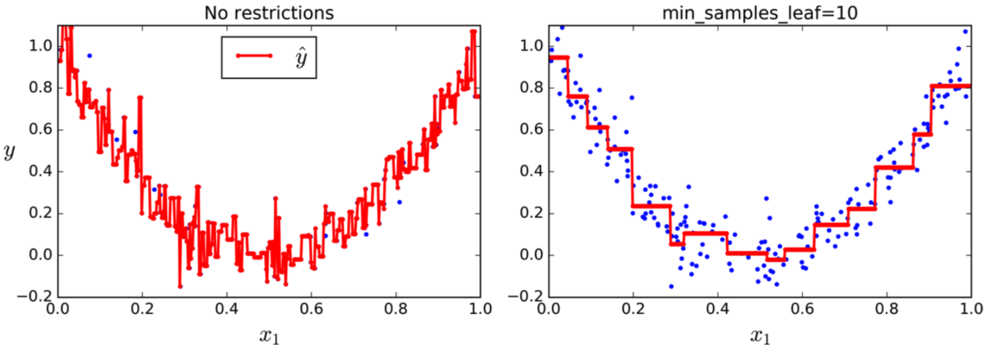

Figure 2.12: The figure on the left represents the estimations of the regression DT ($\hat{y}$) without any constraints on any of its parameters. It can be noticed that the DT fits perfectly every example, leading to overfitting and reducing its capability to generalise well. The figure on the right displays the same DT, but with a restriction on its *min_samples_leaf* parameter, which has the effect of reducing overfitting. The image is obtained from [99].

OVERFITTING. DTs make very few assumptions about the data, unlike other simpler models such as linear regression. On one hand, this constitutes an advantage as DTs can sometimes fit idiosyncrasies in the data better. On the other hand, if the training process is left unconstrained, DTs are able to fit the data perfectly, as illustrated in the left graph of Figure 2.12. This leads to the DT failing to capture the general underlying trend. This phenomenon is known as *overfitting* and refers to ML algorithms that learn to fit the training data perfectly, including the inherent noise. This results in a reduced generalisation ability, which means that the algorithm fails to accurately predict new data. Reducing the overfitting of an algorithm is called *regularisation* and, in the case of DT, is done by constraining a few parameters. Such parameters can be the minimum number of samples before a split is performed, the minimum samples a leaf must have (*min_samples_leaf* in Figure 2.12), the maximum number of leaf nodes or the maximum features evaluated when splitting each node. The effect of including some of these parameters can be seen in the right plot of Figure 2.12, where the *mean_samples_leaf* parameter has been set to 10.

The overfitting problem of single DT models led to the development of random forests, which aim to address this limitation and are outlined in the next section.

Combining DTs into a random forest

As mentioned above, a random forest is a collection of independently trained DTs. In a regression task, random forests make estimations by averaging the answer of all individual DTs. Random forests have been shown to improve the generalisation performance and robustness to noise of single DTs by introducing randomness into the training process [108]. Randomness is typically introduced in two ways, via bagging and feature selection.

BAGGING. The first method to introduce randomness in a random forest is called bagging [110]. The idea behind bagging is that each DT is trained on a different subset of the training data, chosen through random sampling with replacement. Therefore, each sample in the training set can be used more than once during the training of an individual tree as long as the size of the training set size remains constant. An alternative sampling method is *pasting*, when the sampling is performed without replacement.

FEATURE SELECTION. The second method of introducing randomness is by using only a random subset of features to search for the best partition at each node [8]. If the same features were used by every node, all decision trees would be identical. Additionally, if all features were used to search for the best partition, the performance of the random forest would depend heavily on how well the test data is described by the training data. For example, if the test data has one feature very different from the train data, the random forest can ignore this feature in the DTs where it is not selected, and will instead use the other more similar features for the estimation. This aspect is particularly relevant for this work, as we train our random forests using simulated data but estimate using real data, between which we expect to see differences.

The two randomness methods ensure that there is a low correlation between the DTs of a random forest, which has, in turn, been shown to improve the generalisation capacity of a random forest [108]. In addition to this, the performance of a random forest is also affected by other properties, the most important of which are the forest size and the tree depth. As with any machine learning approach, these parameters need to be optimised in order to improve the generalisation of the RF.

FOREST SIZE. The forest size parameter specifies the number of trees in the random forest. Increasing the number of decision trees has been shown to improve the generalisation performance of a random forest [108]. The forest size influences the smoothness of the decision boundary and of the class posteriors [110], which is shown in Figure 2.13. The larger the number of trees, the smoother the regression interpolation. The green line denotes the conditional mean of the predictions and the redness of the points indicates the value of the posterior distribution (the higher the confidence, the darker the colour). It can be noticed that as we increase the number of trees from 1 to 400, the decision boundary becomes smoother. Figure 2.13 also reveals that the confidence is higher for data points closer to the training points.

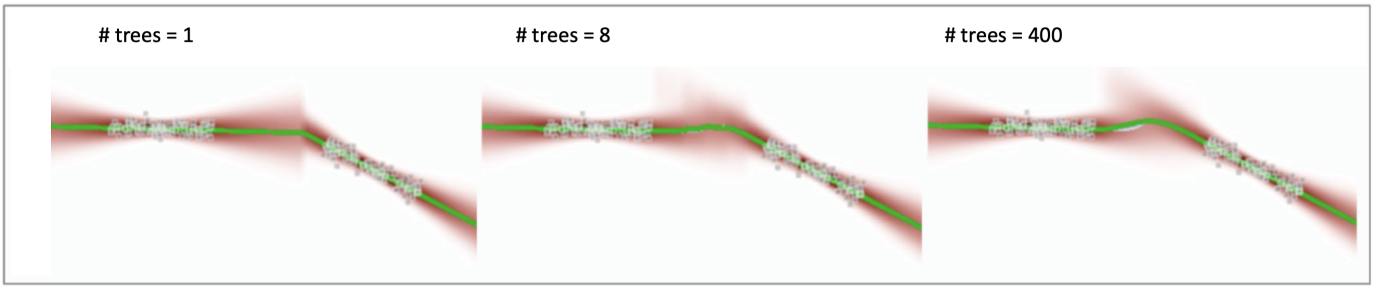

Figure 2.13: The effect of the *forest size* parameter on the class posteriors (red points) and mean curves (green line), which are smoother as the number of trees in the forest increases. Image adapted from [110].

TREE DEPTH. The tree depth parameter specifies the maximum number of levels that each DT in a random forest can have. Usually this parameter is set according to the complexity of the training data. A large tree depth might lead to overfitting if the data is not complex enough, while a small tree depth usually results in underfitting, as the tree is not large enough to catch the complexity of the data.

FEATURE IMPORTANCES. One of the main advantages of random forests is their interpretability, which is particularly relevant in fields such as medical imaging. One way to interpret the results of a random forest is by looking at the *feature importances*. The importance of a feature is usually computed by looking at how often the feature is used as a split criteria. Another indication of a feature's importance is the level of a DT at which this feature is used. As important features are likely to appear on the first levels of the DT, the higher the level where a feature is used as a split criteria, the more important the feature. The interpretability of random forest is in direct contrast with other machine learning approaches such as neural networks, which act as black-box models.

To conclude with, random forests are one of the most widely used powerful machine learning algorithms. We choose random forests over other approaches due to their robustness to noise, excellent generalisation ability, easy parameter tuning and interpretability.

2.4.2 Neural Networks

Neural network models are at the forefront of machine learning research after recent advances in deep learning have dramatically improved the state-of-the-art in a multitude of fields such as object and speech recognition, autonomous driving, drug discovery, genomics or medical imaging [7]. Conceptually, neural networks aim to build an intelligent model by replicating the architecture of the brain's neuronal networks. Despite their recent increase in popularity, neural networks have been around for a long time. They were first introduced in 1943, when McCulloch and Pitt [117] developed the first mathematical model of an artificial neuron, called the McCulloch and Pitt neuron. However, their progress was limited by a lack of computational power and a need for large datasets for training. The current resurgence of interest in neural networks came with the rise of deep learning [7] which proposes the use of neural networks with multiple processing layers (hence the name 'deep'). This was made possible by the exponential increase in computer power which allows the training of deep networks in a feasible time and by the availability of large sets of data [99].

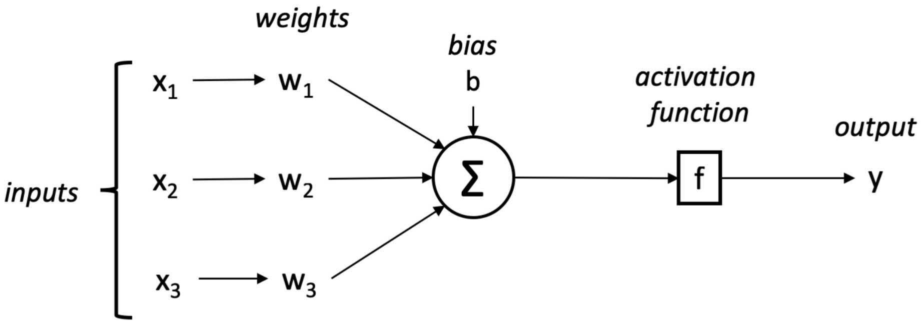

Figure 2.14: Schematic view of an artificial neuron. The inputs x_i are multiplied with their weights w_i, summed and a bias term b is added. Then, the result is transformed by a usually non-linear activation function and passed on either as a network output or as an input to the neurons on the next layer.

ARCHITECTURE. The general idea behind artificial neural networks is to group artificial neurons in multiple connected layers stacked on top of each other. The elementary unit of a neural network is the artificial neuron, a mathematical function conceived as a model of biological neurons. Similarly to the synapse mechanism in the brain, an artificial neuron receives inputs signals from other neurons, processes them and then outputs the result to other neurons. The processing is usually done by calculating a weighted sum of the input signals, adding a bias term and then applying a non-linear activation function. Figure 2.14 illustrates this process. Mathematically, this can be written as:

$$y = f(\sum_{i}^{n} w_i x_i + b),$$

where x_i is the input, w_i is the associated weight, b is the bias, f the activation function and y the final output of the neuron. The role of the activation function f is to combine and transform the inputs in a non-linear way. Two very commonly used activation functions in regression tasks are the hyperbolic tangent function or 'tanh', $f(i) = \frac{e^{2i}-1}{e^{2i}+1}$, and the rectified linear unit or 'ReLu', $f(i) = \max(0,i)$. The weights and bias terms of each neuron are learnt during training so that they best represent the data.

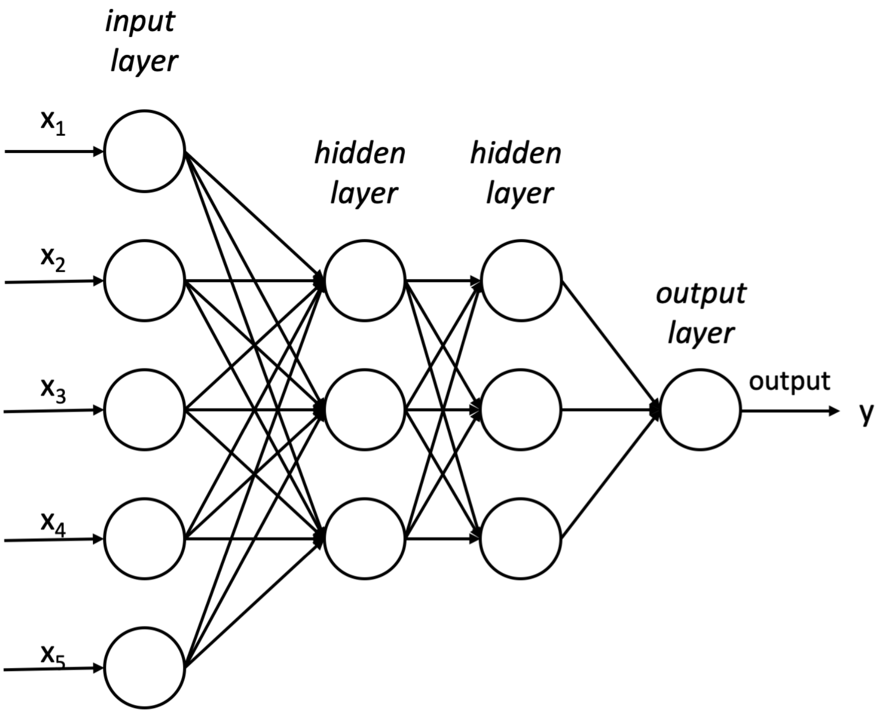

Figure 2.15: Schematic diagram of a multilayer perceptron. The feedforward neural network is made up of an input layer, two hidden layers and one output layer.

In a feedforward neural network, the data is transmitted in one direction. The first layer of a network, called the input layer, receives the network input, processes it and then passes it on to subsequent layers until the last layer is reached. The last layer of the network outputs the result of the regression or classification task and it is called the output layer. The layers in between the input and the output layer are called hidden layers. A feedforward artificial neural network with at least three layers of neurons (input, hidden and output layers) is called a multilayer perceptron Figure 2.15.

While there are motivations for high-level design decision, usually the architecture of a neural network is set by trial and error. This is because the impact of design parameters such as the number of neurons and layers remains poorly understood and, instead, needs to be evaluated empirically on a case by case base

[118]. The lack of understanding of the effect of different parameters makes neural networks act like black-box models, with low interpretability.

Although the basic building blocks of a neural network are simple non-linear mathematical functions, the combined effect of multiple layers and neurons results in a complex system capable of uncovering hidden structures and storing abstract representations of the data [119].

TRAINING. Training a neural network refers to the process of adjusting the weights and biases of each neuron so that the network best represents the data. In supervised learning, this is done by comparing the outputs of the network with the true labels via a loss function J, and minimising the difference between the two using an optimisation algorithm. The optimisation algorithm in a neural network is an automatic differentiation technique based on the 'chain rule' of derivation and it is called backpropagation [120].

During training, the neural network takes a training example, passes it through the hidden layers and outputs an estimation, which is then compared to the true label via the loss function J (usually the mean squared error for a regression task). Next, the error computed by the loss function is propagated backwards, from the output layer to the input layer, to all neurons. For each neuron, the weight and bias parameters are changed in proportion to the neuron's contribution to the overall error. The update of the weight and bias of each neuron is calculated by computing the gradient of the loss function with respect to the neuron.

The parameter update is usually computed via gradient descent, an iterative optimisation approach to find the minimum of a function, here the loss function J. In gradient descent, the parameters of a function are optimised through iteratively making small steps in the direction that decreases the loss. The steps are proportional to the negative gradient of the loss function:

$$p_{n+1} = p_n - \gamma \nabla J,$$

where p is the parameter being updated, γ is the the learning rate (or the step size) and ∇J is the gradient of the loss function. Minimising the loss function is equivalent with ensuring that the network estimates are as close as possible to the true values and, thus, the network learns how to best represent the training data.

There are several variations of the gradient descent algorithm, a comprehensive review of which is presented in [121]. Our experiments were ran using 'mini-batch stochastic gradient descent' and 'Adaptive Moment Estimation' (Adam) [122], both modified versions of the gradient descent algorithm, widely used in the neural network community. In the mini-batch stochastic gradient descent, the parameters are updated using a batch of randomly selected training instances rather than the entire training set. A batch refers to the subset of instances in the training set. Once all batches of the training set have been used to update the network parameters, a training epoch is completed. Unlike conventional gradient descent, Adam adapts the learning rate during training using exponentially decaying estimates of the mean and variance of the gradients. This has been shown to improve the robustness of neural networks in some datasets [122], leading to Adam being widely used in many state-of-the-art neural networks.

Once training is completed, the performance of a neural network is assessed on previously unseen data. This is important as it measures not only the ability of the network to learn the training data (just by memorising it), but also to generalise the learned trends to new data. Due to their extremely high number of parameters, neural networks have been shown to be able to perfectly memorise the training data, including the noise [123], making them extremely prone to overfitting.

REGULARISATION. Regularisation techniques aim to address the overfitting problem of a machine learning model and improve its generalisation performance. This is usually achieved by constraining the solution space or by introducing randomness, often at the cost of increasing the training loss [124]. However, regularisation techniques are known to be less effective in the case of neural networks, most likely due to the their very high number of parameters [123]. Nevertheless, regularisation is usually still applied as it can result in marginal improvements in performance [123]. For an in-depth analysis of regularisation techniques the reader is advised to consult [124]. For the remainder of this sections will focus on the ones relevant for this work.

Early stopping is the simplest regularisation technique and it refers to stopping the training process early. This is usually determined by looking at the model performance on a validation set and interrupting training when the performance no longer improves. Nevertheless, these are just indicative guidelines and, in practice, is it often difficult to determine the best point to stop the training process [123].

Other popular regularisation techniques are the L1- or L2-norm regularisation. As in the case of linear models, this is applied by adding the regularisation term to the loss function of the model. L1-norm regularisation encourages model sparsity, while the L2-norm aims to keep the neural network's weights as small as possible.

Stochastic gradient descent can also be seen as a form of regularisation, as it introduces randomness into the training process by performing updates in a direction that is determined by a random subset of training instances. This has sometimes been shown to improve generalisation, especially when the batch size is small [125].

Finally, the most popular regularisation technique in the deep learning field is *dropout* [126]. Despite the simple idea behind it, dropout has been shown to reduce overfitting in the most complex neural networks. Output works by 'dropping out' or 'turning off' each neuron with a probability p at each training step. Once training is completed, the neural network uses all neurons for the estimation stage. Essentially, this results in a different neural network being trained at each step as different neurons are 'dropped out' in each training step. While the resulting neural networks share parameters between them, they are, nevertheless, different [99] and dropout can be seen as having an ensemble effect. At a neuron level, dropout has the effect of making individual neurons be less reliant on the other neighbouring neurons since they can be switched off at any time, resulting in more robust neural networks [99].

In summary, neural networks are powerful black-box algorithms that have achieved state-of-the-art performance in a variety of domains. We choose neural networks over other approaches due to their promising performance and ability to discover hidden patterns in high-dimensional, complex datasets. We implement a multilayer perceptron as a starting point for more complex neural network approaches that are undergoing continuous development.

2.4.3 Machine learning in DW-MRI microstructure imaging

The application of machine learning methods to DW-MRI is a recent avenue, with most research studies published over the last five years. Current machine learning applications to DW-MRI data cover different stages of the imaging pipeline: image reconstruction [127], super-resolution [128, 129], tractography [130] or parcellation [131].

Prior to the work presented in this thesis, machine learning had not been used for the estimation of microstructure parameters. Nilsson et al. [6] attempt to estimate microstructure parameters including τ_i using a dictionary approach, but without explicitly learning the relationship between DW-MRI signals and microstructure features. The computational framework presented in this thesis is the first machine learning approach towards microstructure parameter estimation, as published in [132].

Since then, a similar approach has been adapted in [133] to map axon diameter from DW-MRI using a random forest trained on matching histological data. Very recently, Rensonnet et al [134] propose a fingerprinting method to estimate microstructure properties (axonal density and radius indices) using Monte Carlo simulations of DW-MRI signals (fingerprints) and demonstrate the potential of the method on an in-vivo rat model of spinal cord injury. Their approach estimates microstructure parameters as a combination of fingerprints by solving the inverse sparse optimisation problem. Another fingerprinting approach is proposed in [135] to classify DW-MRI voxels in cancer. The authors show that their method can distinguish between young and old tumours and detect whether a tumour has been treated with chemotherapy using a mouse model. An additional application of machine learning for microstructure parameter estimation was proposed by Reisert et al. [136] based on a Bayesian estimator to disentangle the microscopic cell properties of the human brain from the effects of the mesoscopic structure.

Another emerging application enabled by the use of machine learning methods and DW-MRI data is Image Quality Transfer [128]. Alexander et al. propose to use machine learning algorithms to learn and transfer detailed structural information from images in high-quality dataset to enhance lower quality datasets and demonstrate the benefits of their method for both connectivity mapping and microstructure imaging (with NODDI and SMT parameter maps). The concept is initially illustrated using a random forest regressor [128], and further developed using a convolutional neural network which enables uncertainty quantification [129].

To the best of our knowledge, there are no studies that attempt the estimation of the exchange time parameter using machine learning methods.

2.4.4 Summary and motivation for the thesis

In this chapter, we introduce microstructure imaging using DW-MRI and machine learning, with a focus on the permeability measured via the intra-axonal exchange time τ_i. First, we show that permeability based biomarkers are of great interest for the understanding, monitoring and diagnosis of otherwise poorly understood demyelinating pathologies such as MS. Next, we discuss the suitability of DW-MRI for estimating τ_i and other specific tissue parameters due to its sensitivity to the diffusion of water molecules in biological tissue. In addition to this, we present an overview of the most widely used modelling techniques for analysing DW-MRI data. Our review covers tissue models both with and without exchange, and discusses their possible pitfalls. We show that progress in estimating the exchange time has been limited due to the intractability of the mathematical expressions that incorporate this parameter. This suggests that there is a need for a different approach that bypasses analytical expressions. Computational models that learn the relationship between tissue microstructure parameters and the DW-MRI signal directly from simulations constitute a promising alternative.

Prior to this work, there has only been one previous attempt to estimate permeability using a computational model and simulations of the DW-MRI signal [6]. The work, however, uses a dictionary approach that is not able to generalise to new, unseen signals. The approach in [6] uses a basic nearest-neighbour algorithm which simply retrieves the closest matching signal in the simulated library and does not attempt to learn the relationship between the DW-MRI signals and microstructure parameters.

Machine learning approaches can overcome these limitations and learn the relationship between signals and tissue parameters as well as generalise to new, unseen signals. More specifically, random forests and neural networks, have shown extremely promising state of the art results in a wide range of medical imaging applications [8, 7, 137]. Therefore, in this work, we propose for the first time to use machine learning for the estimation of microstructure parameters. To demonstrate this idea, we develop and test the feasibility of machine learning based computational models with permeability for white matter imaging. Throughout this work, we will carry out extensive experiments to investigate the performance of such models in simulations and in-vivo, using clinical and preclinical data. For this, we use realistic synthetic white matter tissue substrates, healthy volunteers and MS patients, and a controlled animal model of demyelination. More importantly, we will directly validate part of our model's microstructure estimates using ground truth histological measurements.

Ultimately, this thesis will provide the answer to whether machine learning based computational models constitute feasible techniques for white matter microstructure imaging. In the big picture, the findings from this thesis can contribute towards using machine learning based computational models with permeability as imaging biomarkers in demyelinating pathologies such as MS.

Chapter 3

Sensitivity of DW-MRI sequences to permeability effects: A Simulation Study

In Chapter 2 we discussed the importance of estimating tissue microstructure properties including the intra-axonal exchange time τ_i, and we illustrated how DW-MRI provides a tool with which we can extract this information non-invasively. While recent work looks at the sensitivity of DW-MRI sequences to axon diameter [138], their sensitivity to permeability effects remains poorly understood.

This chapter presents for the first time a study of the sensitivity of DW-MRI sequences to the intra-axonal exchange time τ_i in white matter. Here, we will focus on SDE (previously called 'PGSE') and STEAM sequences since these are the most widely available and commonly used DW-MRI sequences in the clinical and preclinical setting. The experiments in this chapter analyse the effect of different combinations of sequence parameters on the sensitivity to the intra-axonal exchange time τ_i in synthetic substrates representative of white matter tissue. Parts of this work were presented in abstract form at the International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting 2017 in Honolulu, Hawaii.

3.1 Motivation

As many diffusion studies use off-the-shelf sequences, there is a lack of understanding of how different sequence parameters impact the sensitivity to specific microstructural features. While automatic optimisation frameworks exist [72], they do not provide a clear understanding of the output. Numerical simulations, on the other hand, can help address this problem by offering an insight into how the DW-MRI signal varies with different sequence parameters and how these parameters affect the sensitivity to specific microstructural features. Understanding the relationship between the parameters of a sequence and the sensitivity to a microstructure feature of interest can also help maximise a scanner's potential given its hardware constraints, as shown by previous work [72, 56, 139]. While recent work looks at the sensitivity of DW-MRI sequences to axon diameter [138], their sensitivity to the exchange time τ_i has not yet been studied.

In this chapter, we aim to analyse the feasibility of DW-MRI sequences for estimating the intra-axonal water exchange time τ_i in the brain and spinal cord white matter, a parameter inversely related to the membrane permeability. Furthermore, we study the impact that different sequence parameters have on the sensitivity to the exchange time. For this, we use numerical simulations of the DW-MRI signal to explore combinations of sequence parameters covering clinically and preclincally plausible settings. We then calculate the exchange time resolution limit, defined as the largest identifiable exchange time τ_{max} above which one cannot distinguish from identical impermeable substrates, for a range of substrates representative of white matter tissue and under realistic noise conditions.

3.2 Methods

This section describes the model used to represent digital phantoms of white matter tissue, introduces the sequence parameters of interest, and explains the concept of sensitivity to exchange time used in this study.

3.2.1 Monte Carlo simulations

For all experiments in this section, we use Monte Carlo simulations of the DW-MRI signal, generated using the open source Camino simulation framework [140]. Each simulated signal corresponds to a digital phantom that mimics white matter tissue substrates. A digital phantom is defined through a unique combination of five parameters: the mean (μ_R) and standard deviation (σ_R) of the axon radius gamma-distribution, the intrinsic diffusivity of the spins d, the intra-axonal volume fraction f and the intra-axonal water exchange time τ_i. Here, the intra-axonal water exchange time τ_i is defined as the average time a water molecule spends inside the intra-axonal space of the white matter substrate and it is inversely related to the membrane permeability k. In line with previous work [6, 84, 88], the relationship between the permeability k and τ_i is defined through the expression $\frac{R}{2*\tau_i}$ (Eq. A), where R is the axon radius. This and the more general form of the expression are derived in [87].

While the first four parameters can be passed in directly to the Camino simulation framework, τ_i is specified via the probability p of a water molecule stepping through the axonal membrane when it encounters it. The probability p is related to the permeability k through the expression: $p = k\sqrt{6\frac{\delta t}{d}}$ (Eq. B), where d is the intrinsic diffusivity and δt is the temporal resolution of the simulation. This expression is obtained by combining the Monte Carlo step length equation $s = \sqrt{6d\delta t}$ [98] with the transition probability equation $p = \frac{k*s}{d}$ as derived in [141]. By combining equations (A) and (B) above, we obtain $p = \frac{R}{2*\tau_i}\sqrt{6\frac{\delta t}{d}}$ as the direct relationship between the probability p and τ_i. For impermeable substrates ($\tau_i = \infty$), p is set to 0 and the water molecule is reflected back into the intracellular space.

To mimic the structure of white matter tissue, we represent each synthetic substrate as a collection of non-abutting, parallel cylinders with radii drawn from a gamma distribution, a common choice in the literature based on histological measurements [16]. We set the number of gamma-distributed cylinders in each substrate to 100,000 to avoid variation due to sampling of the distribution [142]. The cylinders are packed randomly in the substrates as described in [98] and are aligned along the z direction. We set the mean for the gamma distribution of the axonal radii to $\mu_R \varepsilon \{0.2, 0.75\}$ µm and the standard deviation to $\sigma_R \varepsilon \{0.1, 0.3\}$ µm, corresponding to small and large axons according to histology measurements in the human [16], monkey [18] and rat [71] corpus callosum. In addition to this, we also simulate substrates with $\mu_R = 3.5$ µm and $\sigma_R = 1$ µm, corresponding to larger axons which can be found in the spinal cord according to several post-mortem measurements [143, 144, 145]. We consider two different values for the intra-axonal volume frac-

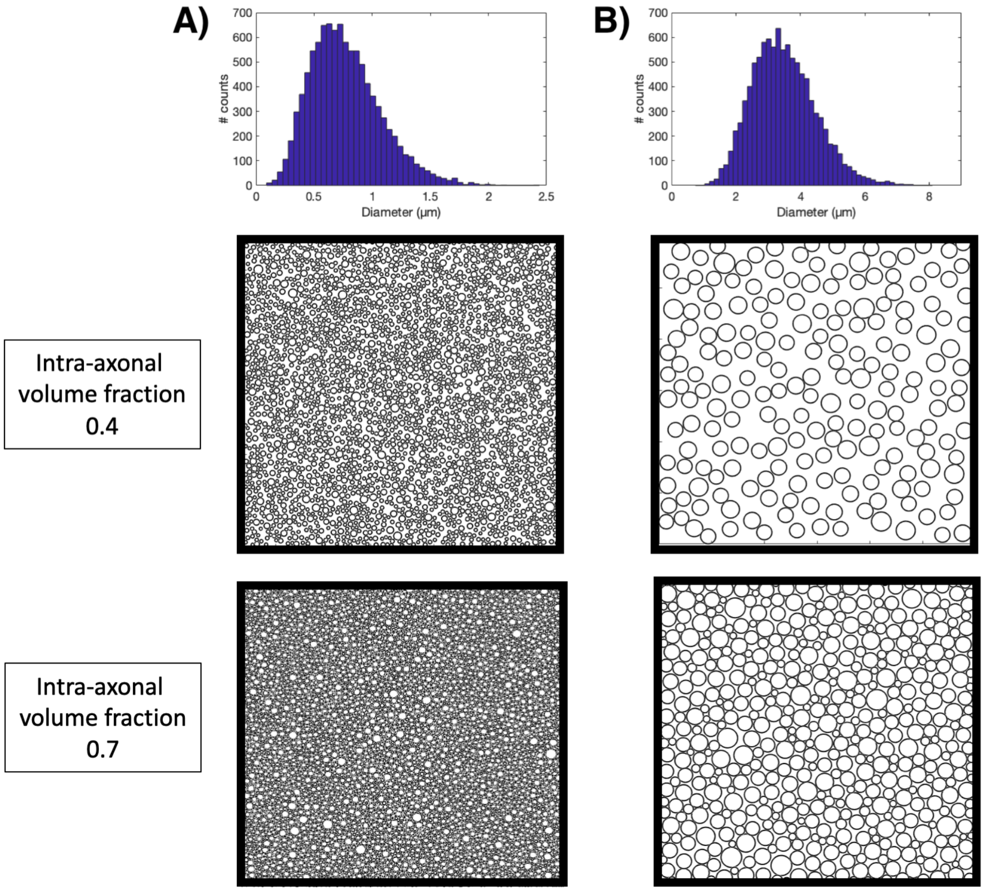

Figure 3.1: Example substrates for Monte Carlo simulations. The histograms in the first row represent the axon diameter distributions used to generate the synthetic substrates in the rows below. Figure **B)** shows a substrate with $\mu_R = 0.75$ µm and $\sigma_R = 0.3$ µm, representative for the brain white matter. Figure **B)** shows a substrate with $\mu_R = 3.5$ µm and $\sigma_R = 1$ µm, representative for the spinal cord white matter.

tion $f \varepsilon \{0.4, 0.7\}$, corresponding to low and high values in the white matter, and we set the intrinsic diffusivity d to $1.7x10^{-9}$ m^{2}s^{-1} [73]. This results in a total of 6 different digital phantoms, examples of which are shown in Figure 3.1. For each substrate, we simulate the signals for exchange times τ_i between 10 ms and 2000 ms, a range which encompasses values of τ_i expected in the in-vivo human and rat white matter tissue according to several literature studies [9, 10, 11, 95, 12].

All Monte Carlo simulations are performed using 100,000 spins and 2,000 time steps in order to ensure that the precision of the unweighted signal (10^{-10}) is several orders of magnitude smaller than the realistic signal noise [98].

3.2.2 Pulse Sequences

Throughout this chapter, we simulate different DW-MRI sequences by varying the following three parameters: the diffusion gradient pulse duration δ, the diffusion time from the first pulsed gradient to the beginning of the final pulsed gradient Δ, and the gradient amplitude of the pulsed gradient G. The diffusion gradients are set perpendicular to the substrate's cylinders in order to maximise sensitivity to the water exchange time. We calculate the b-value for the our pulse sequences as $b = \gamma^2\delta^2G^2(\Delta - \delta/3)^2$.

To calculate practically achievable sensitivities, we include the effect of T1 and T2 relaxation in all our experiments. As our simulations probe long values of Δ, which can be achieved by using STEAM sequences, our signals depend on the echo time (TE) and the mixing time (TM). We include the effect of transverse relaxation by normalising the simulated signal S_0 by $\exp(-\frac{\mathrm{TE}}{\mathrm{T2}})$, where T2 is the transversal relaxation time of white matter. In a STEAM sequence, the spins also undergo T1 relaxation during the mixing time TM between the second and third 90°RF pulses. Consequently, the measurements made using longer Δ experience more relaxation, leading to lower intensities. We include the effect of the T1 relaxation through scaling the signals S_0 by $1/2 * \exp(-\frac{\mathrm{TM}}{\mathrm{T1}})$, where T1 is the longitudinal relaxation time of white matter.

3.2.3 Sensitivity and resolution limit

We design several experiments to investigate the sensitivity of diffusion sequences to the intra-axonal exchange time. Here, we define the sensitivity of a measurement $S(\tau_i)$ to the exchange time as the difference in the DW-MRI signal between $S(\tau_i)$ and the signal for an equivalent impermeable substrate, $S(\tau_i = \infty)$. This difference illustrates whether the exchange time effect can be detected in the signal, i.e. whether it can be distinguished from signals coming from the equivalent impermeable substrate. To compute the sensitivity, we calculate the change in signal for exchange times between 10 ms and 2000 ms for each simulated substrate. As this effect depends on the amount of noise present, we compare $\Delta S(\tau_i)$ with the standard deviation of the noise for a range of SNRs: SNR=10 as a high level of noise, SNR=20 as an expected level of noise for pulse sequences with long Δs, and

SNR=50 as a low level of noise. For all combination of sequence parameters, we use the same TE=680 ms, which we use to calculate the standard deviation for all three SNRs. TM is calculated as the difference between the fixed TE and the Δ and δ for each parameter combination.

To illustrate practically detectable ranges of exchange time values, we compute the exchange time *resolution limit* τ_{max} for different sequence parameters (Δ, G and δ) assuming an SNR of 20. The resolution limit τ_{max} represents the largest detectable τ_i above which one cannot distinguish from identical impermeable substrates (where $\tau_i = \infty$), i.e. the larger the resolution limit, the larger the range of permeabilities that can be detected. For example, given a set of sequence parameters and a particular substrate setting, a resolution limit of $\tau_{max} = 1000$ ms would imply that only substrates with exchange times between 0 and 1000 ms can be distinguished from equivalent impermeable substrates in the presence of noise. The resolution limit is computed by determining the largest τ_i for which the signal difference $\Delta S(\tau_i) = |S(\tau_i = \infty) - S(\tau_i)|$ is above the standard deviation of the noise σ.

3.3 Results

This section presents the main results that investigate the sensitivity of DW-MRI sequences to the intra-axonal exchange time τ_i. The experiments in this section aim to find the range of exchange times to which we are sensitive in both the clinical and the preclincal setting, under realistic noise conditions. Additionally, we aim to identify the key sequence parameters that maximise signal sensitivity to exchange time and evaluate their impact under realistic noise and tissue relaxation conditions. The simulations in this section investigate a range of sequence parameter combinations feasible on current human imaging systems: $G \leq 200$ mT/m, $\Delta \leq 480$ ms and $\delta \leq 40$ ms, corresponding to b-values up to 8,400 $\mathrm{s/mm^2}$, achievable on high performance hardware such as the Connectome scanner or preclinical scanners. The experiments here use the tissue model described in the Methods section and assume $\mathrm{T1} = 830$ ms and $\mathrm{T2} = 70$ ms to match standard values in the white matter at 3T [40].

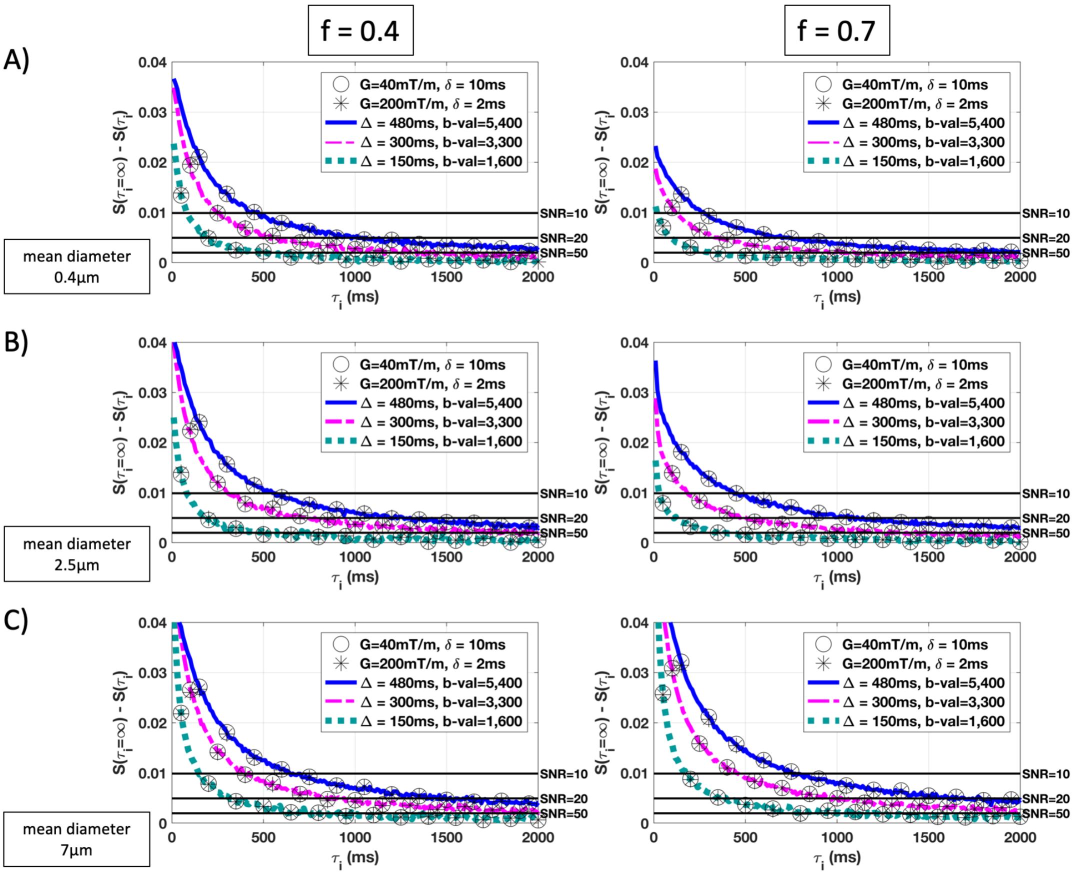

Figure 3.2: Impact of two different G and δ settings on the sensitivity to the exchange time for $\Delta\varepsilon\{150, 300, 480\}$ ms. Figure shows the difference in signal $\Delta S(\tau_i) = |S(\tau_i = \infty) - S(\tau_i)|$ for substrates with different mean diameters ((**A**) $\mu_D = 0.4$ μm, (**B**) $\mu_D = 1.5$ μm, (**C**) $\mu_D = 7$ μm) and different intra-axonal volume fractions (0.4 in the *left column* and 0.7 in the *right column*). The level of signal detectability for the three SNR levels (10, 20, 50) is marked by the black lines.

The first subsection looks at how the sensitivity is affected by the choice of G and δ over a range of different b-values, Δs and SNRs. The second subsection looks at the impact Δ has on the sensitivity range and computes the resolution limit τ_{max} for different Δs and under realistic noise conditions (SNR=20). In the third subsection, we look at how the b-value impacts the sensitivity when Δ is fixed and we compute the resolution limit for different sequence parameter settings. The results in this section provide an indicative sensitivity range that we expect to find in the mammalian brain and spinal cord, under clinical and preclinical scanner settings.

3.3.1 Choice of δ, G

First, in Figure 3.2, we assess the impact of G and δ on the sensitivity to τ_i. For this, we simulate two scenarios. First, we look at values of G and δ commonly used in clinical scanners ($G = 40$ mT/m and $\delta = 10$ ms - represented by the circle markers). For the second scenario, we simulate G and δ values achievable in Connectome or preclinical scanners: $G = 200$ mT/m and $\delta = 2$ ms (represented by the asterisk markers). For each scenario, we simulate a range of Δs achievable in both the clinical and preclinical setting: $\Delta\varepsilon\{150, 300, 480\}$ ms, corresponding to b-values $\{1600, 3300, 5400\}$ s/mm^2. The two combinations of G and δ, ($G = 40$ mT/m, $\delta = 10$ ms) and ($G = 200$ mT/m, $\delta = 2$ ms), have the same b-value weighting (e.g. when computing the b-value of the sequence, both pairs of G and δ contribute with the same amount to the b-value). This allows us to study their effect in isolation from the b-value.

Figure 3.2 shows the results for substrates representing small (row A - $\mu_D = 0.4$ μm) and large axons in the brain (row B $\mu_D = 1.5$ μm). In row C, we show the results for substrates representative of the human spinal cord ($\mu_D = 7$ μm). The *left column* in Figure 3.2 illustrates the results for $f = 0.4$, while the *right column* shows the results for $f = 0.7$, representing small and large values in the white matter. The plots in Figure 3.2 show the change in the simulated DW-MRI signal $\Delta S(\tau_i) = |S(\tau_i = \infty) - S(\tau_i)|$ (*y-axis*) against the exchange time τ_i (*x-axis*). As the detection of permeability effects depends on the amount of noise present in the data, we indicate through black horizontal lines three different noise levels corresponding to SNRs of 10, 20 and 50.

All plots in Figure 3.2 share a common pattern. The difference in the signal between the permeable and impermeable substrates is initially considerably above all three different noise levels, after which it flattens out for values of τ_i above 1000 ms. This suggests that permeability effects are more easily distinguishable for short exchange times ($\tau_i \leq 1000$ ms) than for long exchange times ($\tau_i \geq 1000$ ms). Furthermore, for all substrates and all b-values simulated, we notice that there is no difference in sensitivity between the clinical (*circle* markers) and preclinical (*asterisk* markers) setting, as shown by the overlapping circle and asterisk markers. This

Table 3.1: Resolution limit τ_{max} (ms) for a range of Δs and SNRs.

	μ_D	Δ	**SNR=10**	**SNR=20**	**SNR=50**
f = 0.7	*0.4* µm	*150* ms	10	90	300
		300 ms	110	310	1020
		480 ms	220	610	1900
	1.5 µm	*150* ms	20	110	380
		300 ms	170	450	1300
		480 ms	370	1000	2000
	7 µm	*150* ms	160	330	1010
		300 ms	410	920	2000
		480 ms	640	1400	2000
f = 0.4	*0.4* µm	*150* ms	70	200	510
		300 ms	230	530	1300
		480 ms	400	880	2000
	1.5 µm	*150* ms	80	200	550
		300 ms	290	670	1880
		480 ms	470	1080	2000
	7 µm	*150* ms	140	300	790
		300 ms	350	790	2000
		480 ms	560	1280	2000

The table shows τ_{max} (ms) for a range of Δ and SNRs (10, 20 and 50) and for a b-value of 5,400 s/mm^2. τ_{max} represents the largest exchange time τ_i above which one cannot distinguish from an equivalent impermeable substrate. The underlined values represent sequence parameters for which the full range of exchange times simulated ([10, 2000] ms) is identifiable, given the noise threshold.

shows that the choice of G and δ is not critical for detecting permeability effects, and that we can achieve the same sensitivity in both the clinical and preclinical setting as long as the b-value weighting of G and δ is similar. Therefore, for the rest of the experiments in this chapter, we will ignore the effects of these two parameters. Figure 3.2 also shows that the sensitivity to exchange time effects increases with Δ and the b-value in all simulated substrates. We analyse in more detail how the sensitivity changes according to Δ and the b-value in the next set of experiments.

The sensitivity ranges in Figure 3.2 are different for each substrate and for each set of sequence parameters, resulting in a different resolution limit τ_{max} for each setting. Table 3.1 shows the different resolution limits τ_{max} from Figure 3.2, which we compute as described in Section 3.2.3. τ_{max} represents the largest detectable τ_i

above which one cannot distinguish from identical impermeable substrates (where $\tau_i = \infty$) and depends on the substrate and sequence parameters as well as the noise level. The underlined values in the table represent sequence parameters for which the full range of exchange times simulated ([10, 2000] ms) is identifiable, given the noise threshold.

Table 3.1 reveals two trends across the substrate parameters. Firstly, the resolution limit increases with the mean axon diameter μ_d. For all Δs and SNRs, the τ_{max} for substrates with $\mu_D = 0.4$ µm are smaller than for $\mu_D = 1.5$ µm, which are in turn smaller than for substrates with $\mu_D = 7$ µm. For example, when $\Delta = 480$ ms, $f = 0.7$ and SNR=20, the resolution limit τ_{max} goes up from 610 ms to 1000 ms and then to 1400 ms as the axon diameter increases from $\mu_D = 0.4$ µm to $\mu_D = 2.5$ µm, and to $\mu_D = 7$ µm. The second trend reveals that, for substrates representative of the brain's white matter ($\mu_D \varepsilon \{0.4, 1.5\}$ µm), the resolution limit is higher for small volume fractions ($f = 0.4$) than for large ones ($f = 0.7$). The opposite trend is true for substrates characteristic of the spinal cord ($\mu_D = 7$ µm), where the exchange time has a wider detectable range for large volume fractions ($f = 0.7$) than for small ones ($f = 0.4$). We notice that the sensitivity to exchange time effects increases with Δ and the b-value: the sequence with the maximum sensitivity range is the one with the longest simulated Δ (480 ms) and b-value (5400 s/mm^2). For example, for $\Delta = 480$ ms and an SNR of 50, the resolution limit τ_{max} is 2000 ms, covering the full range of exchange times simulated.

3.3.2 Impact of Δ

In this subsection, we investigate the effect Δ has on the sensitivity to the exchange time when the b-value is fixed. For this, in Figure 3.3, we show the same substrates as in Figure 3.2 and compute the sensitivity to exchange time for $\Delta\, \varepsilon \{150, 300, 480\}$ ms, where each Δ value is represented by the *square, diamond* and *circle* markers respectively. We fix the b-value to 5,200 s/mm^2 for all sequence parameter combinations by varying δ accordingly and setting $G = 40$ mT/m. We observe the same trends in Figure 3.3 as in the previous section: the longer the Δ, the wider the sensitivity range, and the larger the axonal diameter, the higher the resolution limit.

To gain a better understanding of the impact of Δ on the resolution limit τ_{max},

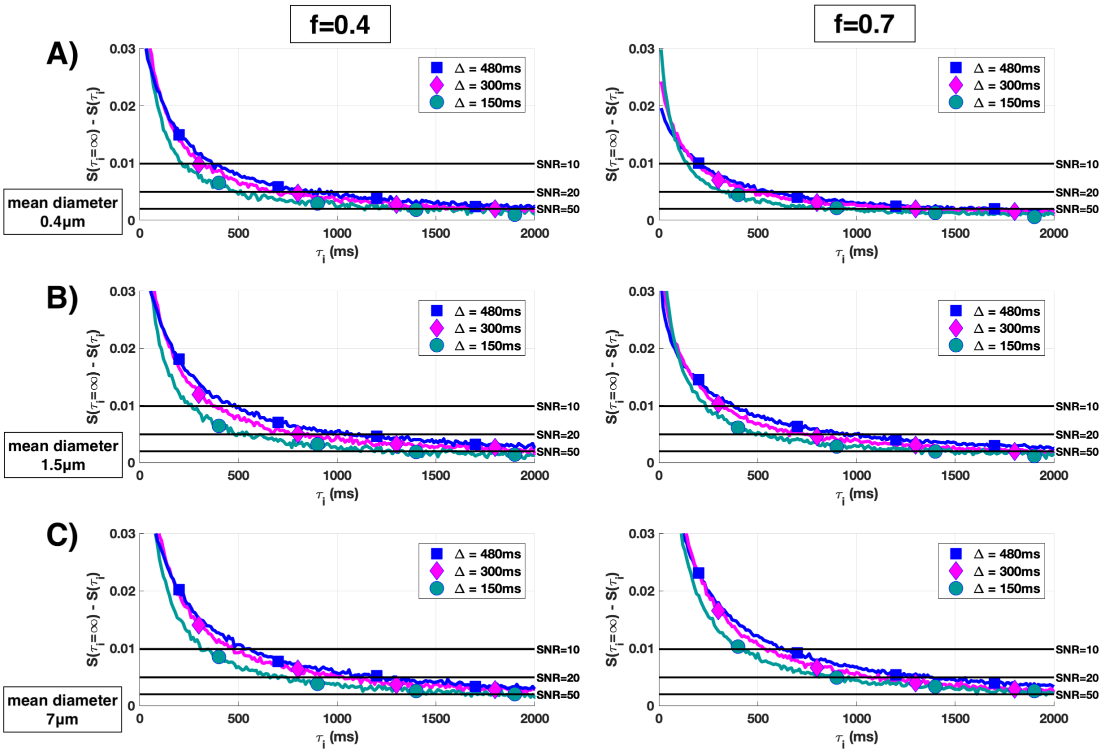

Figure 3.3: Impact of Δ on the sensitivity to the exchange time. Figure shows the difference in signal $\Delta S(\tau_i) = |S(\tau_i = \infty) - S(\tau_i)|$ between the substrates in Figure 3.2 and the identical impermeable substrates. For all three different Δ settings, $G = 40$ mT/m and the b-value is set to 5,200 s/mm^2 by varying δ. The level of signal detectability for the three SNR levels (10, 20, 50) is marked by the black lines.

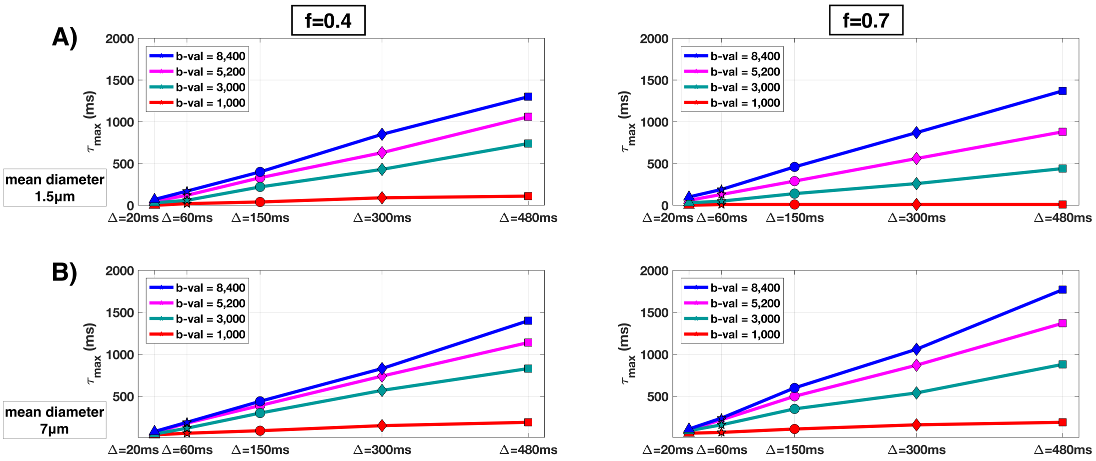

Figure 3.4: Impact of Δ on the resolution limit τ_{max} for a range of b-values under realistic noise conditions. τ_{max} represents the largest exchange time τ_i above which one cannot distinguish from an equivalent impermeable substrate. Figure shows the resolution limits for substrates with two different mean axon diameters - $\mu_D = 1.5$ µm (row **A**) and $\mu_D = 7$ µm (row **B**) - and two different intra-axonal volume fractions (0.4 in *left* column and 0.7 in *right* column).

we calculate the τ_{max} for each Δ value and substrate setting in Figure 3.3. We show these results in Figure 3.4, where we plot τ_{max} against Δ as the *magenta line*. In addition to this, to cover a wider range of sequence parameters, we also run simulations for $\Delta \varepsilon \{20, 60\}$ ms and for three additional b-values: 1,000 $\mathrm{s/mm^2}$ (*red line*) as a low b-value, 3,000 $\mathrm{s/mm^2}$ (*cyan line*) as a moderately high b-value and 8,400 $\mathrm{s/mm^2}$ (*blue line*) as a high b-value, achievable only in preclinical or Connectome scanners. We only show the results for the substrates with $\mu_D \varepsilon \{1.5, 7\}$ µm, but we observe the same trends for the substrate with the smallest axons. All the τ_{max} values in Figure 3.4 are calculated for an SNR of 20, as the level of noise we expect to encounter in clinically viable imaging protocols with long Δs.

Figure 3.4 shows that the sensitivity to exchange time increases with Δ across all substrates and that long Δs are crucial for ensuring a good sensitivity to the exchange time. For example, when Δ is very short (20 ms - *triangle* markers), we cannot detect the effect of exchange times larger than 200 ms, despite b-values as high as 8,400 $\mathrm{s/mm^2}$. We continue to see this even when Δ is longer ($=$ 150 ms), where only exchange times smaller than 500 ms are distinguishable. Increasing Δ from 150 ms to 300 ms doubles this range, with a $\tau_{max} \approx 1000$ ms when the b-value is high ($\geq 5,200$ $\mathrm{s/mm^2}$). For very long Δs (480 ms - *square* markers) we can probe the effect of much longer exchange times, with a resolution limit τ_{max} of 1770 ms for large axons ($\mu_D = 7$ µm) and high b-values (8,400 $\mathrm{s/mm^2}$) (blue square in the left plot of Figure 3.4 B). Figure 3.4 also reveals that increasing Δ does not improve the sensitivity if the b-value is too small, as shown by the flat slopes of the *red* lines. This suggests that sequences with b-values $\leq$ 1000 $\mathrm{s/mm^2}$ (*red* lines) have very limited sensitivity to exchange times despite long Δs ($\tau_{max} \leq$ 190 ms), and that moderately high b-values ($\geq$ 1000 $\mathrm{s/mm^2}$) are necessary to see the effects of maximising Δ.

In addition to this, Figure 3.4 shows that increasing the b-value also improves sensitivity to the exchange time. This effect, however, is very small for short Δs and becomes significant for long Δs. For example, in substrates with $\mu_d = 1.5$ µm, $f = 0.4$, when Δ is very short ($\Delta = 20$ ms), increasing the b-value from 1,000s/mm^2 to 8,400s/mm^2 increases the resolution limit τ_{max} only very slightly, from 10 ms to

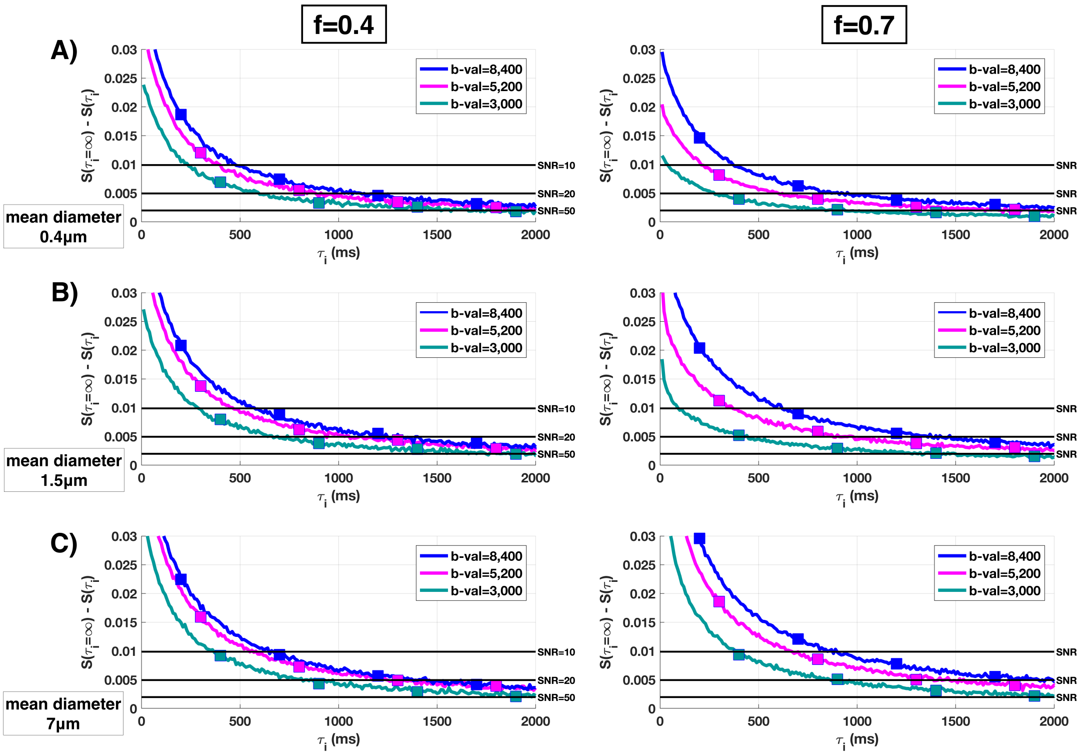

Figure 3.5: Impact of b-value on the sensitivity to the exchange time. Figure shows the difference in signal $\Delta S(\tau_i) = |S(\tau_i = \infty) - S(\tau_i)|$ between the substrates in Figure 3.2 and the identical impermeable substrates. For all three different b-values, we fix $G = 40$ mT/m and $\Delta = 480$ ms and vary δ. The level of signal detectability for the three SNR levels (10, 20, 50) is marked by the black lines.

70 ms (Figure 3.4A - *triangle* symbols in the left plot). On the other hand, when Δ is long (480 ms), increasing the b-value has a much stronger effect: τ_{max} goes up from 110 ms to 1300 ms as the b-value increases from 1,000s/mm^2 to 8,400s/mm^2 (Figure 3.4A - *square* symbols in the left plot). This suggests that increasing the b-values improves the sensitivity significantly only for sequences with long Δ. The next set of experiments looks closer at the effect of the b-value on the resolution limit of a sequence.

3.3.3 Impact of b-value

In this subsection, we investigate the impact of the b-value on the sensitivity to exchange time when Δ is fixed. For this, in Figure 3.5, we fix Δ to 480 ms and we show how the sensitivity changes with three different b-values $\{3000, 5200, 8400\}$ $\mathrm{s/mm^2}$ using the same substrates as in Figure 3.2. We set G to 40 mT/m and we vary

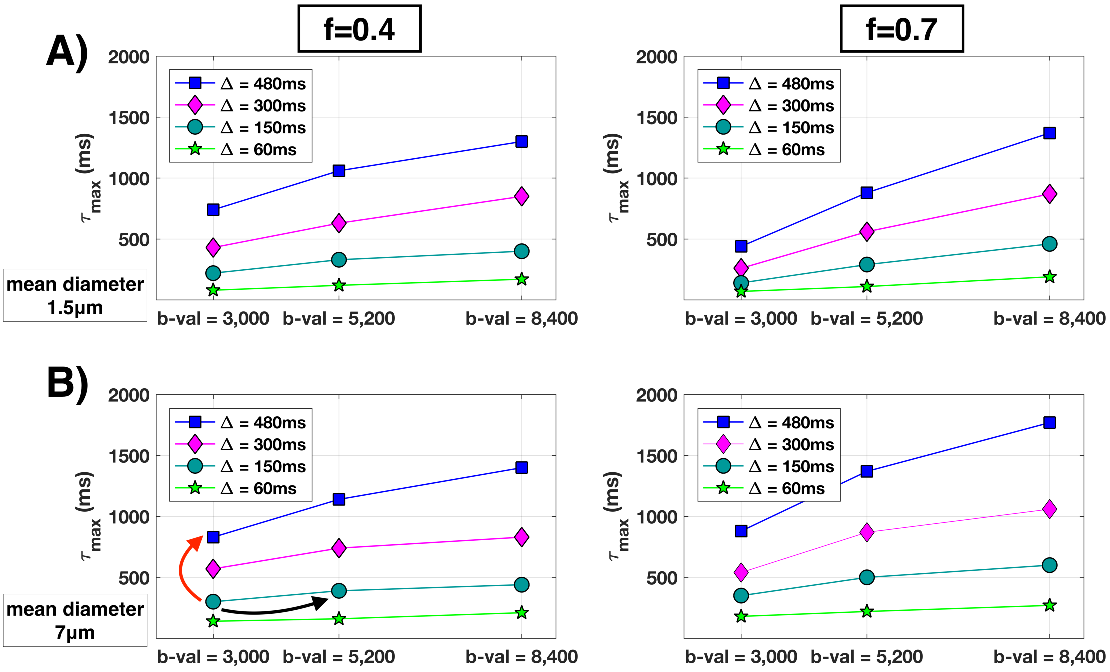

Figure 3.6: Impact of the b-value on the resolution limit τ_{max} for a range of Δs and SNR=20. τ_{max} represents the largest exchange time τ_i above which one cannot distinguish from an equivalent impermeable substrate. Figure shows the resolution limits for substrates with two different mean axon diameters - $\mu_D = 1.5$ μm (row **A**) and $\mu_D = 7$ μm (row **B**) - and two different intra-axonal volume fractions (0.4 in *left* column and 0.7 in *right* column).

δ according to each of the three b-values. The trends we observe are consistent with Figure 3.2: the higher the b-value, the wider the sensitivity range to exchange time effects. This effect is more pronounced for larger axons (row-wise comparison) and high volume fraction (column-wise comparison).

Figure 3.6 quantitatively shows the resolution limit with the detectable range of exchange times for each b-value and substrate in Figure 3.5 (*blue square lines*). In addition to this, to study how the impact of the b-value varies with Δ, we also show the results for three additional fixed Δ values (300 ms - *magenta diamond* lines, 150 ms - *cyan circle* lines, 60 ms - *green star* lines). Figure 3.6 only shows the results for the substrates with $\mu_D \varepsilon \{1.5, 7\}$ μm, but we observe the same trends for the substrate with the smallest axons. As in the previous subsection, we calculate τ_{max} for an SNR=20, as the level of noise we expect to encounter in clinical imaging protocols with long Δs.

Figure 3.6 reveals that with a long Δ (480 ms) and a b-value of 5,200 s/mm^2, achievable on high performance clinical scanners, we can probe the effect of exchange times up to 1370 ms in substrates with large axons ($\mu_D = 7$ µm) and up to 1060 ms in substrates with smaller axons ($\mu_D = 1.5$ µm). By increasing the b-value to values easily reachable on preclinical scanners, but outside the range of clinical scanners (8,400 $\mathrm{s/mm^2}$), we can reach values of up to 1770 ms in substrates with large axons and up to 1370 ms for substrates with smaller axons.

Figure 3.6 also shows that large b-values do not guarantee large sensitivity to exchange time. For example, the maximum resolution limit τ_{max} for $\Delta = 60$ ms (*green* lines) across all substrates is only 270 ms even when the b-value is 8400 $\mathrm{s/mm^2}$ (Figure 3.6B *right* column). Despite a considerable increase in the b-value (from 3,000 $\mathrm{s/mm^2}$ to 8,400 $\mathrm{s/mm^2}$) the resolution limit improves only by a very small amount (90 ms - 33% increase) and the sensitivity remains very limited. This is also reflected in the very flat slope of the *cyan* and *green* lines, which correspond to small Δs. On the other hand, when Δ is long ($\geq$ 300 ms), increasing the b-value improves the sensitivity to exchange time considerably. This can be seen in the much steeper slopes of the *blue* and *magenta* lines in Figure 3.6. In this case, increasing the b-value from 3,000 $\mathrm{s/mm^2}$ to 8,400 $\mathrm{s/mm^2}$ results in a more significant increase in τ_{max}, from 880 ms to 1770 ms (*right* column in Figure 3.6B), a 101% increase. This suggests that increasing the b-value improves the sensitivity by a significant amount only when Δ is already long ($\Delta \geq 300$ ms).

In addition to this, Figure 3.6 also reveals that for clinical scanners with b-values below 5200 $\mathrm{s/mm^2}$, the impact of Δ on sensitivity is greater than that of the b-value. An example of this is illustrated by the left plot in Figure 3.6B) through the *red* and *black* arrows. Here we can see how pushing an already high b-value (3000 $\mathrm{s/mm^2}$) to the limit of what clinical scanners can achieve (e.g. 5,200 $\mathrm{s/mm^2}$) increases the resolution limit τ_{max} from 300 ms to 390 ms, a 30% improvement (*black arrow*). On the other hand, increasing Δ from 150 ms to 480 ms has a much higher impact on the sensitivity, driving the resolution limit from 300 ms to 830 ms, an increase of 176% (*red arrow*). This trend can also be confirmed by a visual comparison between Figures 3.4 and 3.6, where the slopes of the lines in Figures 3.4 are much steeper than those in Figures 3.6. This suggests that although increasing the b-value improves the sensitivity to the exchange time, the impact of Δ is greater when maximising sensitivity to τ_i.

3.4 Discussion

In this chapter, we explore clinical and preclinical combinations of DW-MRI sequence parameters and investigate their sensitivity to the intra-axonal exchange time τ_i. For this, we use Monte Carlo simulations of the DW-MRI signal generated from synthetic substrates mimicking white matter tissue characteristic of the mammalian brain and spinal cord. To calculate practically achievable sensitivities, we include the effect of T_1 and T_2 relaxation in all experiments.

Impact of DW-MRI sequence parameters. Our results show that, when maximising sensitivity to the exchange time, the combination of G and δ is not critical as long as the overall effect of G and δ on the b-value is the same. This suggests that we can obtain a similar sensitivity on a clinical scanner, using a low G (40 mT/m) and a long δ (10 ms), as on a preclinical or Connectome scanner, where G is usually much higher (e.g. 200 mT/m) and δ is shorter (e.g. 2 ms). Furthermore, we find that the longer the Δ and the higher the b-value, the better the sensitivity to τ_i. We show that, under realistic noise conditions (SNR=20) and for all tissue substrates, long Δs ($\geq$ 300 ms) are crucial for reaching resolution limits $\geq$ 1000 ms. We also show that the longest simulated Δ (480 ms) maximises the resolution limit across all substrates. Additionally, we find that increasing the b-value also improves sensitivity to the exchange time, but the effect is significant only for sequences with long Δs ($\geq$ 300 ms). These results suggest that in order to ensure a good sensitivity to the exchange time ($\tau_{max} \geq$ 1000 ms), a combination of long Δ ($\geq$ 300 ms) and moderately high b-values ($\geq 3{,}000$ s/mm^2) are needed.

Sensitivity of different substrates. Our simulations show that the resolution limit τ_{max} (the highest τ_i that can be distinguished from $\tau_i = \infty$) increases with the size of the axons in each substrate. Furthermore, we found that the resolution limit varies differently with the intra-axonal volume fraction, depending on the mean axonal diameter μ_D: for smaller axons ($\mu_D = 0.4, 1.5$ μm), usually found in the brain's white matter, τ_{max} decreases with f, while for larger axons ($\mu_D = 7$ μm), usually found in the spinal cord, τ_{max} increases with f.

Implications for clinical and preclinical applications. The analysis in this chapter reveals that, on clinical scanners, DW-MRI sequences with long diffusion times (480 ms) are sensitive to exchange times up to 1060 ms in substrates representative of the brain white matter, and to exchange times up to 1370 ms in substrates with larger axons, characteristic of the white matter in the spinal cord. On preclinical scanners, which can reach higher b-values, the range of sensitivity is wider: $\tau_{max} = 1370$ ms for substrates characteristic of brain tissue and $\tau_{max} = 1770$ ms for substrates characteristic of spinal cord tissue. Substrates with exchange times greater than τ_{max} are practically indistinguishable from impermeable substrates. We note that here we considered only Δ values up to 480 ms and an SNR of 20, and that the resolution limit is highly dependent on the sequence parameters and the SNR level. We expect longer Δs ($\geq$ 480 ms) to increase the resolution limit even further and simultaneously decrease the SNR due to additional T_1 decay. In-vivo estimates of exchange times are inherently difficult to obtain due to fixation altering the permeability of cell membranes. Nevertheless, most literature studies in rat and human healthy white matter tissue report values of τ_i under 1000 ms [9, 10, 11, 12], with smaller values expected in demyelinated tissue, all within the sensitivity range of the diffusion sequences simulated here. Therefore, our results suggest that carefully tuned clinical and preclinical DW-MRI sequences have sufficient sensitivity to the exchange time τ_i to allow for the probing of values we expect to find in white matter tissue.

Model considerations In this study, we make several assumptions about the geometry of the tissue such as representing axons as non-abutting parallel cylinders due to current limitations of our simulation system. This model does not capture the full complexity of realistic tissue, where potentially confounding effects such as dispersion or myelin water are present. These could have a confounding effect on the estimation of the intra-axonal exchange time, especially in the case of dispersion, which has been shown by Nilsson et al. [1] to result in an underestimation of τ_i. It is possible that the inclusion of such effects would change our results, and more complex simulations that include dispersion and myelin water are subject to ongoing research [146, 147, 148, 86] and form part of future work.

3.5 Conclusions

This chapter presents a simulation study that investigates the sensitivity of DW-MRI sequences to the intra-axonal exchange time τ_i in substrates mimicking white matter in the brain and spinal cord, and aims to find the key parameters that drive this sensitivity. We find that long Δs ($\geq$ 300 ms) are crucial for ensuring a good sensitivity to τ_i ($\geq$ 500 ms). In addition to this, we show that, under realistic noise and tissue relaxation conditions, clinically viable DW-MRI sequences with long Δs ($\geq$ 300 ms) and high b-values ($\geq 3,000\ \mathrm{s/mm^2}$) have a sufficiently good sensitivity to the intra-axonal exchange time to allow for the probing of τ_i values we expect to find in white matter tissue ($\leq$ 1000 ms) [9, 10, 11, 12]. Moreover, we show that DW-MRI sequences with very high b-values ($\geq 5000\ \mathrm{s/mm^2}$), easily achievable on preclinical and Connectome scanners, can probe longer exchange times ($\geq$ 1500 ms), which we expect to find in tissue with larger axons such as in the spinal cord white matter.

Chapter 4

Random forest based computational models with permeability: microstructure parameter estimation in clinical in-vivo human data

The work in Chapter 3 shows that clinically viable DW-MRI sequences with long Δ have a sufficiently good sensitivity to the intra-axonal exchange time to allow for the probing of most τ_i values we expect to find in brain white matter tissue. By developing models of the white matter tissue and relating them to sensitive DW-MRI measurements, we can estimate features such as the intra-axonal volume fraction, the intra-axonal exchange time or the axon diameter. This chapter introduces the first machine learning based compartment model with permeability for white matter microstructure imaging, using a random forest and an optimised STEAM protocol. We test the performance of this novel machine learning based computational framework for microstructure imaging using Monte Carlo simulations of the DW-MRI signal and in-vivo clinical scans from two healthy and two MS patients.

This work is a collaboration study published in *NeuroImage* in 2017, introducing the first machine learning based computational model with permeability for white matter [132]. My contributions to the paper were the following: the optimisation of the machine learning model as presented in the Methods and Appendix of the manuscript, various parts of the synthetic data analysis (correlation analysis, testing the model on a simpler substrate), as well as addressing several of the reviewer's questions (e.g. additional explanations about the probability-permeability link, diffusion step size calculation, random forest regression Appendix). In addition to this, I have recoded, re-simulated and re-analysed the entire pipeline, synthetic database and results of the paper, as well as extended the analysis to further optimise the model and to analyse the feature importances and, therefore, all the scripts, figures and results used in this chapter are my own.

4.1 Motivation

The intra-axonal water exchange time τ_i, a parameter associated with axonal permeability, is an important microstructural property of the tissue, which has been linked with the condition of the myelin sheath surrounding the axons in brain tissue [1, 2, 3]. Several neurological conditions such as MS cause a breakdown of the myelin sheath through a process known as demyelination, which may lead to an decrease in the exchange time as the intra-axonal water molecules encounter less barriers. Changes in permeability have also been linked with pathologies such as Parkinsons disease [4] or cancer [5], leading to a widespread interest in developing permeability-based biomarkers. Due to its sensitivity to the motion of water molecules within tissue, Diffusion-Weighted MRI (DW-MRI) is potentially able to estimate τ_i. However, measuring it has been problematic due to the intractability of the expressions which accurately incorporate τ_i into mathematical models.

So far, mathematical models of white matter ignore or assume simplistic models of permeability [149, 89, 150], while others explicitly incorporate τ_i but rely on assumptions that do not hold in human white matter tissue [90, 85]. The Kärger model [85] is the most widely used analytical model that incorporates permeability due to its compatibility with data acquired using clinically available sequences such as SDE and STEAM [6, 40, 92]. However, the model assumes that the two individual pools of water are well mixed and does not model restrictions, both assumptions which do not hold in white matter. Consequently, the Kärger model fails when applied to highly permeable tissue with fast exchange [84]. An alternative analytical model is the apparent exchange rate (AXR) imaging [91, 95]. However, AXR requires a specialised sequence (double diffusion encoding [52]) and conflates τ_i with the intra-axonal volume fraction f, making it difficult to disentangle the two.

Computational models bypass the need for analytical models and incorporate permeability by creating a mapping between simulations of the DW-MRI signal and the ground truth microstructure parameters. Nilsson et al. [6] use Monte Carlo simulations with known ground truth parameters including permeability to generate a synthetic library of DW-MRI signals. Given a previously unseen signal, they estimate permeability using a nearest-neighbour algorithm which retrieves the closest matching DW-MRI signal and its microstructure parameters from the synthetic library. However, their approach requires new libraries to be generated for each acquisition protocol and fibre orientation, and the nearest-neighbour algorithm in general does not have a good generalisation capacity.

Here, we propose a machine learning based computational framework to construct a mapping between microstructural parameters of interest and rotationally invariant features derived from the DW-MRI data. We use Monte Carlo simulations to generate synthetic signals from a library of histologically relevant microstructure parameters. We then train a random forest to learn the mapping between features derived from synthetic DW-MRI signals and ground truth microstructure parameters, including τ_i, and to generalise smoothly between training examples when estimating parameters from previously unseen data. We test this new approach in simulations and in-vivo, and demonstrate its clinical potential using data from two MS patients.

4.2 Methods

In this section, we first introduce the imaging protocol used to generate the synthetic database and to acquire the in-vivo data in this chapter. Then, we describe the synthetic data generation and feature extraction as a first step of our computational model. Next, we introduce the random forest model and explain how this is integrated within our computational framework. Finally, we outline the synthetic and in-vivo experiments designed to assess our model's suitability as a white matter compartment model with permeability.

4.2.1 Imaging protocol

For both the synthetic and the in-vivo data, we use a rotationally-invariant STEAM protocol optimised for a two compartment model with exchange, according to the framework introduced in [72]. The protocol is optimised assuming biophysically plausible tissue parameters and exchange is incorporated in the optimisation framework via the Karger model [85]. The optimisation was performed by Nedjati et al. and more details can be found in [132]. The resulting protocol has 4 shells, with 108 measurements divided equally between the shells and with 4 b=0 measurements each. STEAM protocols are particularly suited to measuring the exchange time τ_i as they allow for the probing of long diffusion times by exploring the longer T1 relaxation rate compared to the T2 relaxation of spin-echo SDE protocols. Each shell in our STEAM protocol has a different Δ, with values ranging from 102 ms to 412 ms. The full set of parameters of the imaging protocol is shown in Table 4.1. The additional crusher and slice gradients of the STEAM protocol are accounted for in our simulations.

Shell	# b=0	# gradient directions	b (s/mm^2)	$\|G\|$(mTm^{-1})	Δ (s)	δ (s)	TM (s)
1	4	23	1622	62	0.102	0.0077	0.07
2	4	23	1718	62	0.412	0.0039	0.375
3	4	23	3611	62	0.406	0.0057	0.37
4	4	23	4031	62	0.169	0.0094	0.135

Table 4.1: STEAM protocol parameters, optimised for a two-compartment model with exchange using the framework in [72]

4.2.2 Synthetic data

4.2.2.1 Monte Carlo Simulations

The first step in our computational model is to generate a database of Monte Carlo simulations of the DW-MRI signal from digital phantoms mimicking a wide range of plausible white matter substrates. For this, we use the imaging protocol in Table 4.1 and the open source Camino ([140]; *http://camino.cs.ucl.ac.uk*) simulation framework [98].

Each digital phantom is represented by a unique tissue substrate, characterised as a unique combination of five parameters: the mean μ_R and the standard deviation

σ_R of the axon radii distribution, the intra-axonal volume fraction f, the intrinsic diffusivity d and the intra-axonal exchange time τ_i. The exchange time τ_i represents the average time a water molecule spends inside the intra-axonal space of the white matter substrates. Here, the exchange time τ_i is the parameter through which we include the effects of the cell membrane permeability k. The intra-axonal exchange time is inversely related to the membrane permeability through the expression $k = \frac{R}{2 * \tau_i}$, where R is the axon radius [87, 6, 84, 88]. While the first four substrate parameters (μ_R, σ_R, f and d) are input directly into the Camino software, τ_i is specified via probability parameter p through the expression: $p = \frac{R}{2 * \tau_i}\sqrt{6\frac{\delta t}{d}}$, where d is the intrinsic diffusivity and δt is the temporal resolution of the simulation (detailed in Chapter 3.2.1). p expresses the probability that a spin steps through a cell membrane when it encounters it (instead of always being reflected backwards as it is the case for impermeable substrates). For impermeable substrates ($\tau_i = \infty$), p is set to 0 and the water molecule is reflected back into the intracellular space.

To model the structure of white matter brain tissue, we build the synthetic substrates as a collection of 100,000 non-abutting, parallel cylinders, with radii sampled from a gamma distribution, a common choice in the brain literature [16]. We choose 100,000 non-abutting parallel cylinders to avoid the variation due to sampling of the gamma distribution, as shown by *Hall et al.* in [142]. To mimic the in-vivo human brain data used in this chapter, we construct a database of unique tissue substrates and their corresponding DW-MRI signals by randomly sampling the five parameters from biophysically plausible ranges: $\mu_R \varepsilon [0.2, 5]$ µm and $\sigma_R \varepsilon [min(0.1, \mu_R/5), \mu_R/2]$ µm (to ensure that the distributions have a non-zero mode, matching the distributions observed in histology [16]), $f \varepsilon [0.4, 0.7]$, $\tau_i \varepsilon [2, 1000]$ ms and $d \varepsilon [0.8, 2.2]$ $\mu\text{m}^2\text{ms}^{-1}$. The cylinders are randomly packed in the substrates as described in [98], with example substrates shown in Figure 4.1. For this range of parameters, the diffusion step size is calculated as in [98] and is between 0.5 µm and 0.9 µm, depending on the diffusivity of each substrate. To ensure the convergence and the high precision of the simulated signals, we generate our synthetic database using 100,000 spins and 2,000 time steps [98]. This setting provides a precision of the simulated signal of 10^{-10}, several order of magnitude smaller than the realistic signal noise

[98].

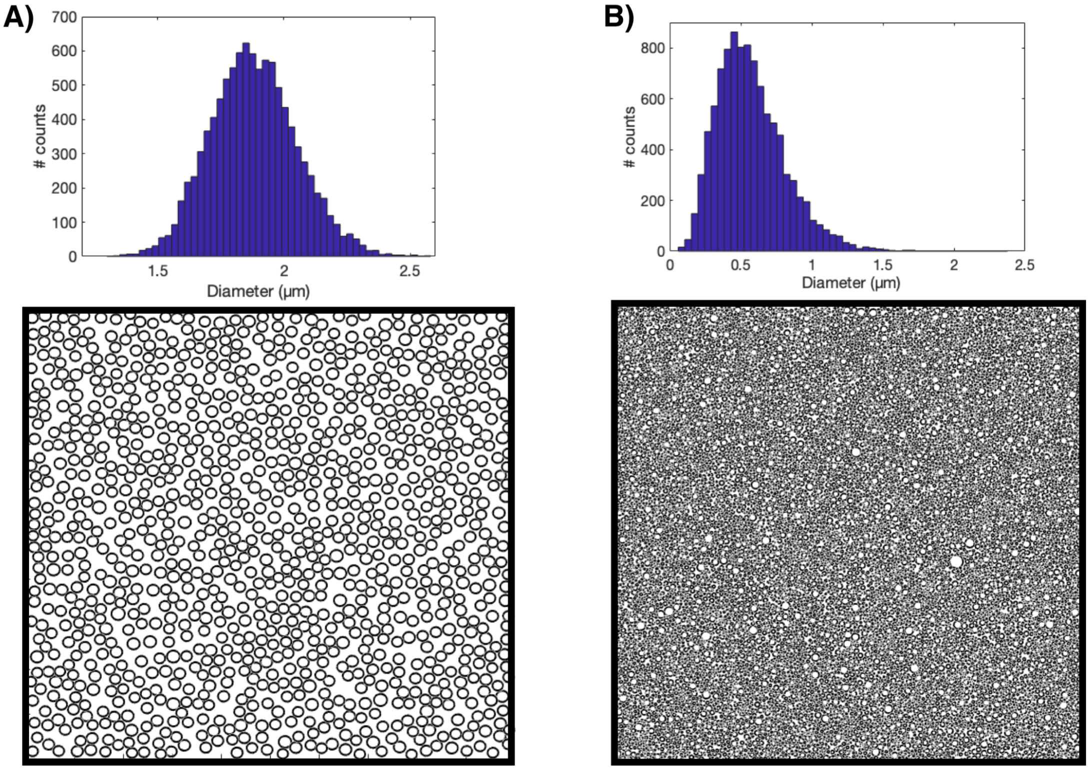

Figure 4.1: Example substrates for Monte Carlo simulations, showing variations in both axonal size distribution and the intra-axonal volume fraction. The histograms in the first row represent the axon diameter distributions used to generate the synthetic substrates in the second row. The parameters for the substrate in Figure **A)** are $\mu_R = 1.88$ μm, $\sigma_R = 0.17$ μm and $f = 0.44$. The parameters for the substrate in Figure **B)** are $\mu_R = 0.58$ μm, $\sigma_R = 0.24$ μm and $f = 0.64$.

To account for the noise present in the in-vivo data in this chapter, we generate an additional noisy database. As spins undergo T1 relaxation during the mixing time TM between the second and third 90°RF pulses, measurements made using longer Δ (and so longer TM) experience more relaxation leading to lower signal intensities and signal to noise ratios (SNR). For the noisy dataset we scale the signals by $\exp(\mathrm{TM}/\mathrm{T1})$ using T1 = 0.832 s for white matter at 3T [40]. We then add Rician noise, choosing the standard deviation of the noise σ so that the SNR of the b=0 images with Δ=0.102 s is 20, which reflects our in-vivo data. This results in two databases of signals for each geometry: noise-free and noisy (SNR=20).

Each simulated signal, associated with a unique combination of ground truth microstructure parameters, forms one entry in our synthetic database. During the next step of our computational model, we extract a set of rotationally invariant features from each simulated signal in our noise-free and SNR=20 databases.

4.2.2.2 Feature extraction

In order to achieve independence from the fibre orientation, we extract a set of 15 rotationally invariant features from each shell of the protocol. The first five features are derived by fitting a DTI model to the DW-MRI signal, while the final ten are derived by fitting a 4th order spherical harmonics (SH) model to the same DW-MRI signal. The 15 rotationally invariant features are presented in more detail in Table 4.2.

Here, we extract the DTI and SH features using the Camino toolkit. This results in an additional signal-derived and rotationally invariant database, which we use to train the random forest. As our STEAM protocol has 4 shells, each entry in our database has 60 features, 15 for each shell. The advantage of this approach is that the training database is independent of the imaging protocol as long as the TE and the b-values of each shell match.

Each resulting feature vector, together with the corresponding ground truth microstructure parameters, represents one entry in the training database for our machine learning based computational model. The rotationally invariant feature vector is used for training, while the ground truth tissue parameters act as labels during the supervised learning task.

4.2.3 In-vivo data

In this chapter, we analyse the data from two healthy volunteers and two MS patients. The first healthy volunteer is a 32 year old male and the second healthy volunteer is a 30 year old female. Both healthy volunteers are scanned twice, with the rescan taking place within a month of the initial scan. The first MS patient is a 22 year old female with relapsing remitting MS (RRMS), a disease duration of 2 years and expanded disability status scale (EDSS) of 2. The second MS patient is a 63 year old male with secondary progressive MS (SPMS), a disease duration of 25 years and EDSS of 6. Both MS patients were scanned once. None of the MS patients recruited into the study experienced a relapse or a course of corticosteroids three months prior to imaging. Written informed consent was obtained for

Ftr No.	Feature/ Model	Feature Information
1	λ_1 (DTI)	First eigenvalue of the diffusion tensor, representing the first main direction of diffusion. Obtained as the first output of the *dteig* command in Camino.
2	λ_2 (DTI)	Second eigenvalue of the diffusion tensor, representing the second main direction of diffusion. Obtained as the fifth output of the *dteig* command in Camino.
3	λ_3 (DTI)	Third eigenvalue of the diffusion tensor, representing the third main direction of diffusion. Obtained as the ninth output of the *dteig* command in Camino.
4	MD (DTI)	Mean diffusivity, an estimate of the overall diffusion in a voxel, computed as $\lambda_1 + \lambda_2 + \lambda_3$.
5	FA (DTI)	Fractional anisotropy, an estimate of the anisotropic diffusion in a voxel. It takes values between 0 and 1 and is computed as: $\frac{3}{2}\sqrt{\frac{\Sigma(\lambda_i - MD)^2}{\Sigma\lambda_i}}$
6	I_0 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=0 and index i, calculated as $I_0 = \sum_{i=-k}^{k} \lvert a_{k,i}\rvert^2$, where k=0.
7	I_2 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=2 and index i, calculated as $I_2 = \sum_{i=-k}^{k} \lvert a_{k,i}\rvert^2$, where k=2.
8	I_4 (SH)	A combination of the SH coefficients $a_{k,i}$ of order k=4 and index i, calculated as $I_4 = \sum_{i=-k}^{k} \lvert a_{k,i}\rvert^2$, where k=4.
9	mean ADC (SH)	This feature is computed by calculating the values of the spherical functions f of the voxel at a set of evenly distributed sample points on a unit sphere S and taking the mean of these. Obtained directly from the *sfpeaks* command in Camino.
10	peak ADC (SH)	The maximum value of the spherical functions f over the points of the unit sphere S (see feature 9 for more details). Obtained as the 10th output of the *sfpeaks* command in Camino.
11	λ_1 (SH)	The first eigenvalue of the Hessian matrix at the peak.
12	λ_2 (SH)	The second eigenvalue of the Hessian matrix at the peak.
13	anisotropy (SH)	The anisotropy of the ADC profile f, obtained using the $sfanis$ command in Camino.
14	skewness (SH)	The skewness of the ADC profile f, obtained using the $sfskewness$ command in Camino.
15	kurtosis (SH)	The kurtosis of the ADC profile f, obtained using the $sfkurtosis$ command in Camino.

Table 4.2: Table shows the 15 DTI and SH (4^{th} order) features extracted from the DW-MRI signal and used for training the random forest regressor.

all participants, and the study was approved by our local research ethics committee.

All scans were acquired and preprocessed by the first authors in [132] using the STEAM protocol in Table 4.1, on a 3 T Philips Achieva scanner, with a 32-channel receive-only RF coil, SENSE acceleration factor of 2.5, full k-space acquisition, and receiver bandwidth of 1948 Hz/pixel. The imaging parameters are: TE=0.068 s, TR=12 s, FOV=256 mm x256 mm, matrix size=128x128, slice thickness=4 mm, number of slices=40 (except the first scan of the first volunteer, with 30 slices). The total imaging time is approximately 40 minutes. For the MS patients, a high resolution (1 mm x 1 mm x 3 mm) T2-weighted scan is also acquired for the purposes of creating a lesion mask.

The images are corrected for eddy currents and motion using the eddy tool in FSL [151]. The white matter voxels are selected by computing maps of linearity and planarity from the diffusion tensor fit to the lowest b-value shell. The SNR of the white matter region in b=0 image of the shell with $\Delta = 102$ ms is approximately 19 for all subjects. In order to create a synthetic database that closely resembles our in-vivo data, we add a similar level of Rician noise to the noise-free synthetic database.

4.2.4 Machine learning

4.2.4.1 Random Forest for microstructure parameter estimation

The model introduced in this chapter uses a random forest to learn a mapping between microstructure parameters and features of the DW-MRI signal. Due to their interpretability, robustness to noise and easiness of tuning [110] random forests are widely used as regression or classification techniques in the medical field [152, 112, 114, 8]. A random forest is formed of a collection of decision trees and its output is a weighted average of the estimates produced by each individual decision tree. Here, each decision tree in the random forest learns a mapping between feature vectors derived from Monte Carlo simulations of the DW-MRI signal and their corresponding ground truth microstructure parameters. The final microstructure parameter estimation of the random forest is obtained by averaging the estimations from all trees in the forest, as illustrated in Figure 4.2.

Each decision tree (shown in Figure 4.3) is formed of internal (grey circles)

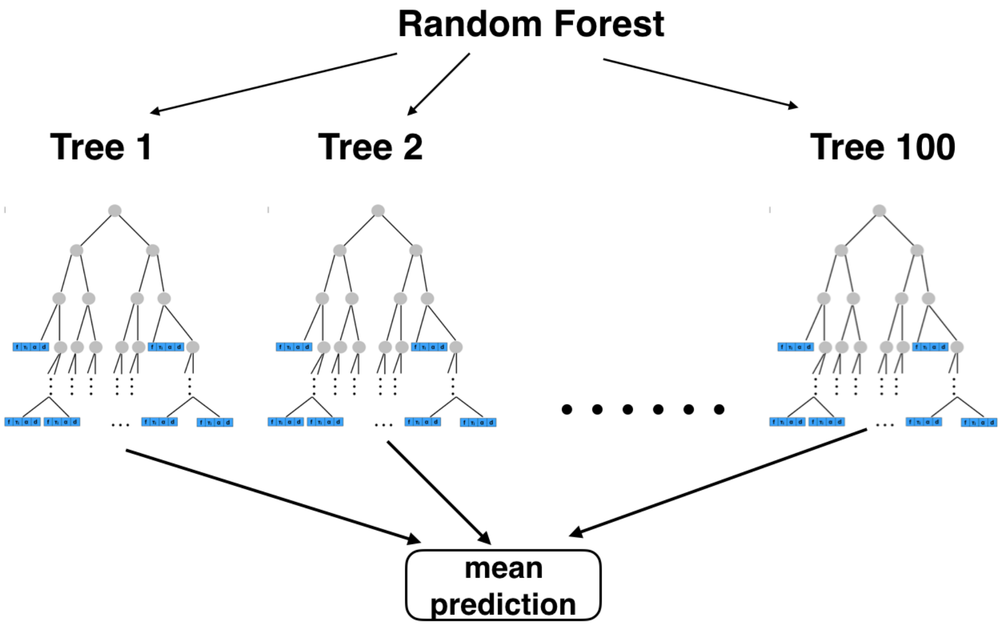

Figure 4.2: Random forest regressor. The regressor used here is formed of 100 trees and the final estimation of the tissue parameters is obtained by averaging over the estimations of each individual decision tree.

and terminal nodes (blue rectangles), connected in a hierarchical fashion by edges (black lines). Each internal node has one incoming and two outgoing edges and stores a test or split function that is applied to the incoming data, which, here, is a multi-dimensional vector with rotationally invariant features of the DW-MRI signal. According to the result of the test function, the internal node sends the data along one of its outgoing edges to the nodes on the level below. This process is repeated until the data reaches a terminal node. Finally, the terminal node takes the incoming feature vector and outputs estimates of the microstructure parameters of interest using a predictor function stored inside the node.

Random forest regression has two stages. The first stage is the training stage, in which the random forest learns a mapping between the input data (here, feature vectors derived from the diffusion MRI signals) and the parameters to be estimated (here, tissue microstructure parameters). The second stage is the testing stage, in which the random forest uses the mapping learned in the training stage to estimate microstructure parameters from unseen feature vectors.

During the first stage, each decision tree in the random forest is trained simultaneously and independently of each other. Each tree in the forest is trained

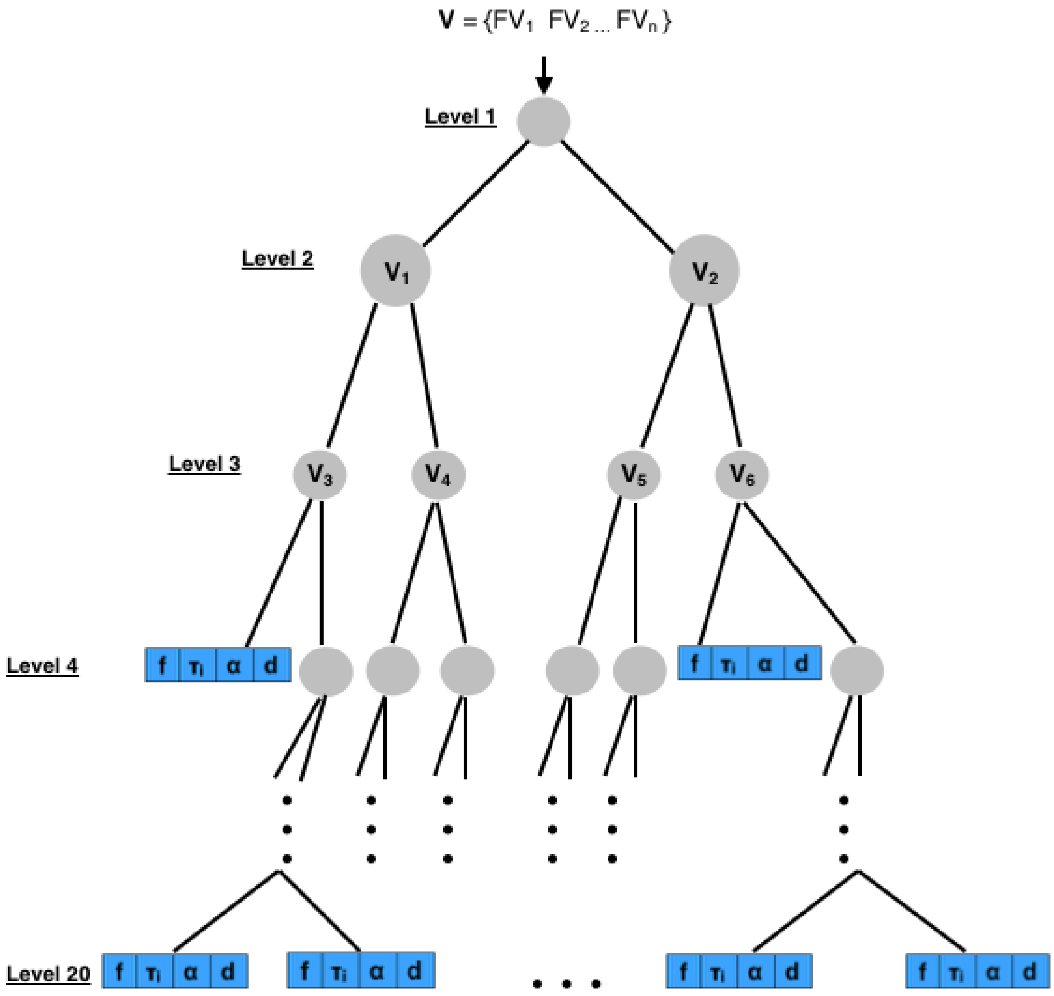

Figure 4.3: *Regression decision tree.* During training, each tree receives a random subset of labelled feature vectors V, which are used to optimise the parameters of the tree. Each branch of the tree splits the feature vector into smaller subvectors, V_i, until the terminal nodes on the lower level are reached. The terminal nodes store the predictor that relates the incoming data to the estimated microstructure parameters of interest.

as follows (see Figure 4.3 for reference): the parent node of the tree performs an initial linear regression to find a relationship between the input feature vector ($V = FV_1, FV_2, ..., FV_n$) and the microstructure parameters of interest (e.g. f, τ_i, α, d). Then, the decision tree looks for a partition of the feature vector V into two subvectors such that having a separate linear regression on each subvector improves the parameter estimation. The best partition is chosen as the one that maximises the information gain over the single partition at the root node. If such a partition exists, the feature vector V is split into two subvectors (V_1 & V_2) and two child nodes are added to the original node. This process is then continued for the child nodes, and all subsequent nodes, until the estimation cannot be improved any further. Finally, when the addition of child nodes does not improve the parameter estimation, the node becomes a terminal node (blue rectangles), and it encodes a piece-wise linear mapping between the feature vectors and the parameters to be estimated. When this stage is reached for all trees in the random forest, the training finishes. Moreover, in order to improve the performance and robustness of the random forest regressor, randomness is typically introduced during training [8]. This can be done either through bagging [110], which trains each decision tree on a random subset of data, or through using only a subset of features to search for the best partition at each node [116].

During the testing stage, the random forest estimates new tissue parameters from unseen feature vectors by generalising from the mapping learned during training. At this stage, we can assess how well the random forest has learned to generalise to new feature vectors that were not seen during training.

4.2.4.2 Random forest implementation

The random forest regressor used here was implemented using the open source scikit-learn toolkit [153], a widely used software machine learning library for Python. We use bagging to inject randomness into the random forest and we run a series of preliminary experiments to determine the number and depth of decision trees, which we present in the next section. More in-depth implementation details are available on the scikit-learn website (*www.scikit-learn.org*). In Figure 4.4, we show how the random forest is integrated within our computational model's pipeline. The figure shows a schematic overview of how training, testing and estimation are done.

We train two separate random forest regressors, one on the noisy data set (SNR=20) and another on the noise-free data set. Following preliminary experiments shown in the next section, the random forests are trained on 10,000 of the 12,500 feature vectors, with the remaining 2,500 previously unseen feature vectors used for testing. The testing phase allows us to evaluate how well the random forest performs by directly comparing the parameter estimations to the known ground truth values. When estimating from the noise-free feature vectors, we use the ran-

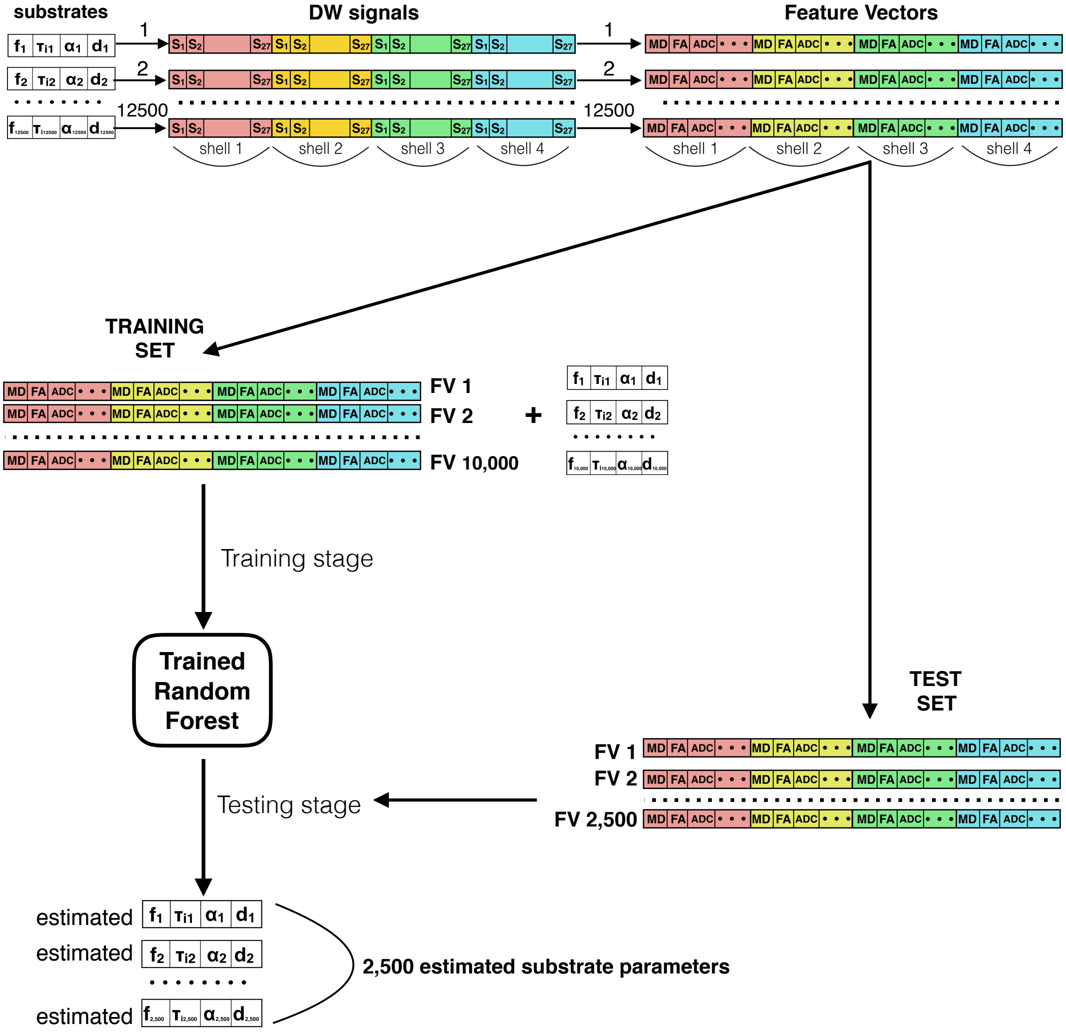

Figure 4.4: A schematic overview of how training, testing and parameter estimation are integrated within our computational framework.

dom forest trained on noise-free data. When estimating from the noisy feature vectors, we use the random forest trained on the noisy data. When estimating on the in-vivo data, which is affected by Rician noise, we use the random forest trained on the SNR=20 database, as it has similar noise characteristics to the actual in-vivo data.

We train the random forest to estimate parameters that describe the underlying microstructure properties of the tissue. For the noise-free database, we estimate the intra-axonal exchange time τ_i, the intra-axonal volume fraction f, the intrinsic diffusivity d, as well as a single volume-weighted radius index α [73], which combines the mean axonal radius μ_R with the standard deviation of the axonal radius σ_R

of the substrate. For the database with SNR=20 as well as the in-vivo data, we estimate only f, τ_i and d and exclude α due to the lack of sensitivity to this parameter in the presence of noise and for the gradient strength used, as shown in [138].

4.2.5 Experiments

4.2.5.1 Synthetic experiments

Random forest hyperparameter tuning

It is generally known that the size and the choice of the training set can influence the performance of the machine learning model. To investigate the optimal size, we generate a database of 25,000 unique tissue substrates and train the random forest on training sets of different sizes, ranging from 5,000 to 22,500 and using the remaining 2,500 as the test set. Experiments in [132] show that the choice of the training set does not bias the estimations of the random forest as long as there is sufficient coverage of the parameter range, which we ensure in all our experiments.

To determine the combination of random forest parameters that maximises the performance of our model, we run a set of preliminary experiments. As accuracy generally increases with the number of trees [108], we first search for the number of trees that maximises the performance of the random forest. In order to determine this, we train different random forests, each with a different number of trees, and we evaluate their performance by computing the Pearson correlation coefficient R^2 between the estimated and the ground truth parameters. Similarly, to determine the optimal tree depth of each decision tree, we sample a wide range of tree depths and study how this affects the correlation coefficient.

Parameter estimation

To assess how well the random forest model estimates match the known ground truth values in the test set, we show correlation scatter plots and compute the Pearson correlation coefficient R^2 between the estimations of the model and the ground truth parameter values. R^2 measures the linear relationship between two datasets and takes values between -1 and $+1$, with ± 1 implying an exact positive/negative correlation and 0 implying no correlation. This is calculated as:

$$R^2 = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{4.1}$$

where cov is the covariance of the two variables, and σ_X and σ_Y are the standard deviations of the two variables. Furthermore, to investigate any general bias in the estimations, we show Bland-Altman graphs, which plot the mean of the ground truth and estimated values against their difference. All points on the correlation graphs and the Bland-Altman plots are colour-coded with colour bars showing the percentage error. For better visibility, we limit the colour bar to –50 to 50%, with points outside this range taking the same colour as the two maximum points.

In addition to this, we also rank the features in both the the noise-free and SNR=20 synthetic database (DTI and SH features) according to how informative they are. We analyse the feature importance to determine which features are the most informative and how this ranking is affected by the presence of noise.

4.2.5.2 In-vivo experiments

Healthy subjects

We generate estimates of the three parameters of interest (f, τ_i and d) across the white matter voxels of the scans and rescans for the two subjects. We show parameter maps of f, τ_i and d across representative axial, coronal and sagittal slices of both scan and rescan data.

We use the scan-rescan data from the two healthy subjects to investigate intra-subject reproducibility at the region of interest (ROI) level. For this, we manually define regions in the splenium (CC-S) and genu (CC-G) of the corpus callosum (CC), the left (ALIC-L) and right (ALIC-R) anterior limbs of the internal capsule and the left (CST-L) and right (CST-R) corticospinal tracts on the scan and rescan FA maps for both volunteers. The matrix size for these images was 128x128 with 40 slices each except the first scan of the first volunteer, for which 30 slices were acquired. For full details of the other imaging parameters, see section 4.2.3. We calculate the mean parameter estimates from the random forest in each ROI for both the scan and rescan data and compute correlation plots for f, τ_i and d individually. To do this, for each parameter, we pool together the scan and rescan parameter estimates for both volunteers in each region.

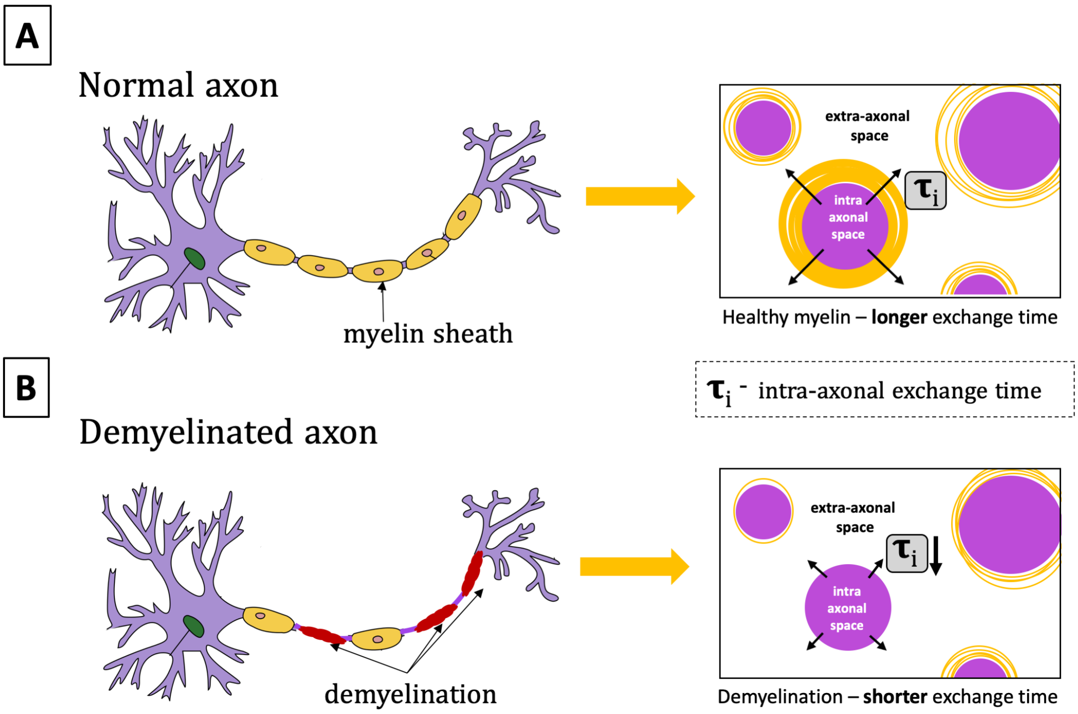

Figure 4.5: The effect of demyelination on the intra-axonal exchange time. In **A** we show a normal axon, with an exchange time τ_i. In **B**, we show a demyelinated axon, where the myelin sheath (*in yellow*) is damaged, resulting in less barriers encountered by the water molecules as they leave the intra-axonal space (*in purple*). This is hypothesised to result in a shorter exchange time τ_i of the demyelinated axon.

MS subjects

To investigate the sensitivity of the random forest to tissue damage in MS lesions, an expert clinical researcher marked the lesions on the high resolution T2-weighted images. We also mark additional ROIs in the contralateral normal appearing white matter (NAWM) for comparison. The ROI and lesion masks are then registered to the mean b=0 image from shell 1 of the diffusion-weighted data using FLIRT [154]. We use this shell for registration as it has the shortest diffusion time and so provides the best contrast. The matrix size of the MS subjects is 128x128 with 40 slices each. For full details on the other imaging parameters, see section 4.2.3. In the first subject, in the early stages of MS, two of the marked lesions overlap the white matter mask completely and are investigated further. For the second subject, in the late stage of MS, the lesions are much more widespread, overlapping most of the white matter mask used to select voxels for analysis. However for this subject, our analysis is limited by the availability of contralateral NAWM for comparison and thus we only use one lesion in the CST. We then calculate the mean parameter estimates from both models in the lesions and control ROIs and compare. We then assess the statistical significance of the difference in estimates between the lesion and the NAWM using a t-test.

As explained in more detail in the Background chapter, the axons in MS lesions undergo a process known as demyelination, which is characterised by a breakdown of the myelin sheath surrounding the axons. This is illustrated in Figure 4.5, and is intuitively expected to lead to a decrease in the intra-axonal exchange time τ_i as the water molecules encounter less barriers when moving from the intra-axonal to the extra-axonal space (see Figure 4.5 B).

4.3 Results

4.3.1 Synthetic experiments

4.3.1.1 Random forest hyperparameter tuning

Params / Training set size	5,000	9,000	9,500	10,000	10,500	11,000	12,500	15,000	20,000
f	0.875	0.878	0.880	0.883	0.883	0.880	0.878	0.878	0.880
τ_i	**0.873**	**0.910**	**0.943**	**0.945**	**0.933**	**0.928**	**0.916**	**0.884**	**0.851**
α	0.800	0.807	0.812	0.815	0.812	0.814	0.808	0.810	0.814
d	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

Table 4.3: Average correlation coefficient (R^2) for training sets of different sizes. For example, a training set size of 5,000 implies that the random forest regressor is trained on 5,000 different randomly sampled unique combinations of substrate parameters and their associated synthetic signals.

Table 4.3 shows the correlation coefficient R^2 for f, τ_i, α and d for noise-free training sets of different sizes. Each R^2 coefficient was computed by first training the random forest on dataset of the size indicated by in each column heading and, once training is finished, calculating the R^2 score on a test set of 2,500 previously unseen noise-free DW-MRI synthetic signals. The results in Table 4.3 show that for all four parameters of interest, the optimal training set size is 10,000 ($f : R^2 = 0.883, \tau_i : R^2 = 0.945, \alpha : R^2 = 0.815,\ d : R^2 = 0.999$). As the training set size

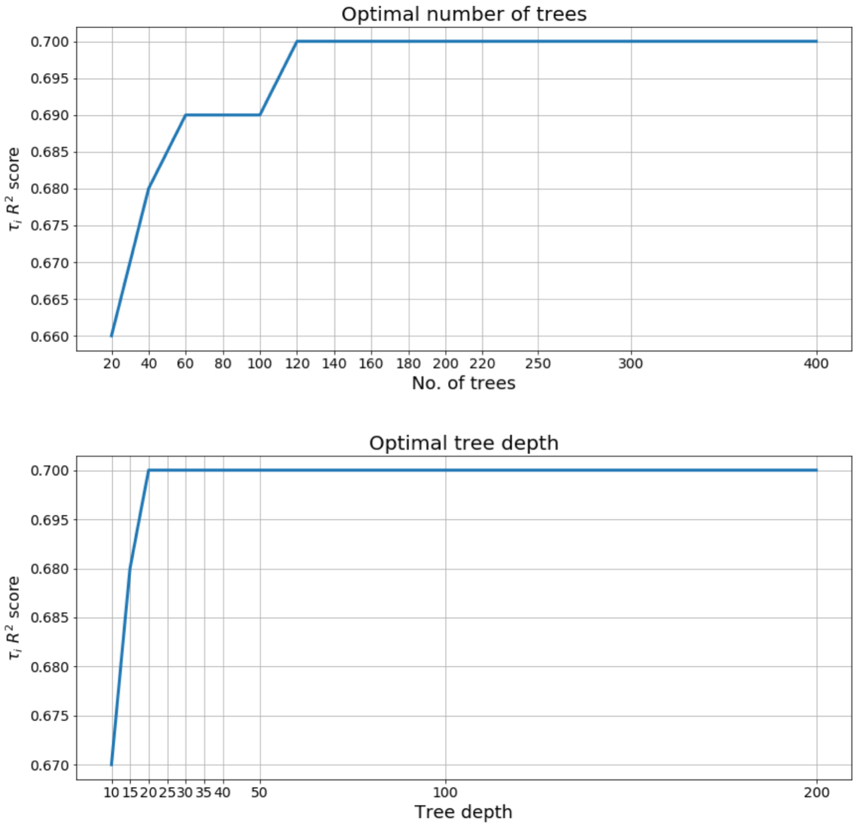

Figure 4.6: *Random forest hyperparameter tuning*. The top graph shows the correlation coefficient between the ground truth and the estimated τ_i (*y-axis*) against the number of trees in the random forest (*x-axis*). The bottom graph shows the correlation coefficient between the ground truth and the estimated τ_i (*y-axis*) against the tree depth of each decision tree in the random forest (*x-axis*).

increases from 5,000 to 10,000 we see marginal improvements in the performance of the random forest, after which the coefficients start going down again. Following from this, all experiments in this chapter are ran using a training database of 10,000 synthetic signals.

The second set of experiments in this section focuses on the optimal number of decisions trees in the forest and the optimal tree depth. As we are interested in optimising the random forest for parameter estimation on the in-vivo data, we run this set of experiments on the SNR=20 database, matching the noise level of our in-vivo data. Figure 4.6 shows the R^2 score for τ_i, as our main parameter of interest, against the number of trees (*top row*) and against the tree depth (*bottom row*). The R^2 score improves by 0.04 when increasing the number of trees from 20 to 120. Above this value, we see diminishing returns, while the computational complexity and, consequently, the estimation time increase. For this reason, we set the number of trees to 120 as the optimal one. For the tree depth, R^2 score improves by $\approx$0.03 when increasing the depth from 10 to 20, reaching a maximum of $R^2_{\tau_i} = 0.97$, after which it plateaus. These results suggest that the optimal hyperparameter setting for the random forest is 120 decision trees, each of maximum depth 20, a setting which we will use for the remaining of the experiments in this chapter.

4.3.1.2 Parameter estimation

Figure 4.7 shows correlation scatter plots (*top row*) and the corresponding Bland-Altman plots (*bottom row*) of f, τ_i, α, and d for the random forest regressor for the noise-free database. The data points are colour-coded according to how close the estimates are to the actual values and the percentage error is shown on the colour bars. The solid black line in the Bland-Altman plot indicates the mean difference between ground truth and estimated parameters and the dashed lines show the 95% limits of agreement.

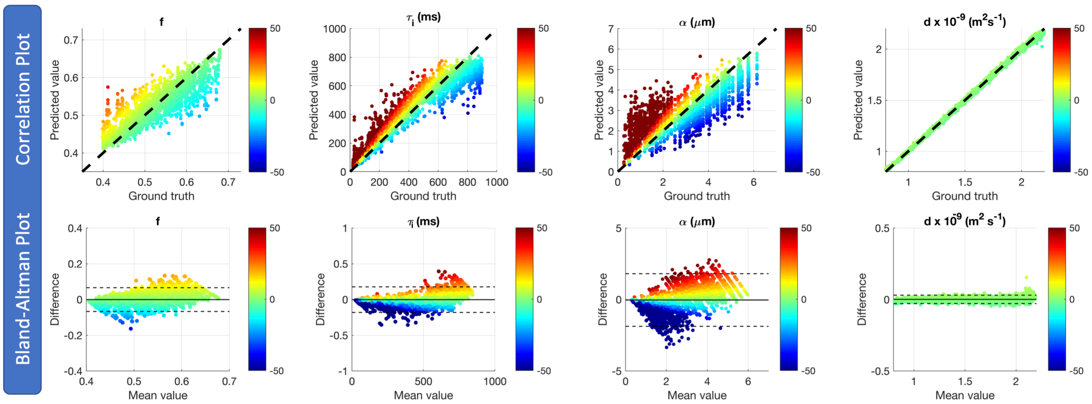

Figure 4.7: *Random forest performance on the noise-free database.* The *top row* shows the correlation scatter plots comparing the ground truth values of f, τ_i, α and d (*x-axis*) with estimations from the random forest (*y-axis*) for the synthetic noise-free database. The *bottom row* shows the equivalent Bland-Altman plots for f, τ_i, α and d. The *y-axis* shows the difference between the ground truth and the estimations of a parameter, and the *x-axis* shows the mean of those two values.

The correlation scatter plots in the first row show a strong correlation between the ground truth and the estimated parameter values: $f : R^2 = 0.88, \tau_i : R^2 = 0.95, \alpha : R^2 = 0.82$ and $d : R^2 = 0.99$. In the corresponding Bland-Altman plots, points are clustered about the zero difference line, indicating low bias. However, despite the absence of noise, the recovery of the parameters is not perfect. In particular, for f, τ_i and α, there is some bias in the estimated values which depends on the ground truth value. For example, large values of τ_i are consistently underestimated, while small values are overestimated.

The noise-free performance provides a benchmark for the best estimation we can achieve given the data available. Furthermore, these results also show that the random forest model can estimate, even though less accurately than for other parameters, the volume-weighted axon radius index α ($R^2 = 0.82$) in the absence of noise. This shows the potential of such a machine learning based approach to be extended to other microstructure parameters if there is sensitivity in the data. Nevertheless, we exclude α from our analysis onwards as it has been shown in [138] that for the gradient strength (G=62 mT/m) and level of noise present in our noisy synthetic and in-vivo data (SNR=20), there is no sensitivity to axon diameters below 6 µm, a range which covers the majority of axons in the brain's white matter [16].

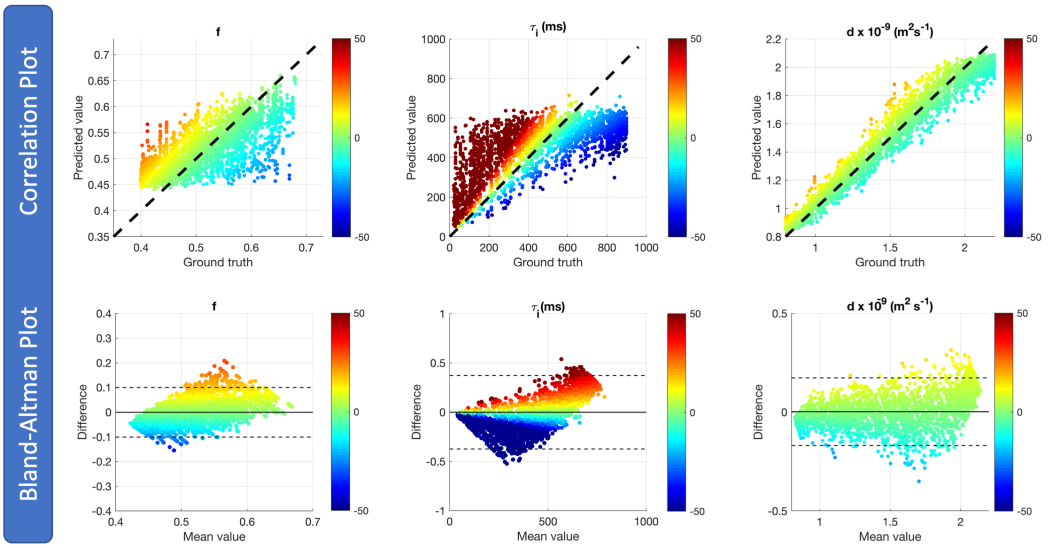

Figure 4.8: *Random forest performance on the SNR=20 database.* The *top row* shows the correlation scatter plots comparing the ground truth values of f, τ_i, α and d (*x-axis*) with estimations from the random forest (*y-axis*) for the synthetic SNR=20 database. The *bottom row* shows the equivalent Bland-Altman plots for f, τ_i, α and d. The *y-axis* shows the difference between the ground truth and the estimations of a parameter, and the *x-axis* shows the mean of those two values.

Figure 4.8 shows similar plots, but for the synthetic database with SNR=20.

The results for all parameters are consistent with those obtained from the noise-free data, although the 95% limits of agreement are slightly wider. Although the mean difference lines are mostly close to zero, again we see that there is some bias in our estimates, which depends on the ground truth values. For f ($R^2 = 0.71$), we see that larger volume fractions are slightly underestimated, whereas low f values are slightly overestimated. This is also the case for τ_i ($R^2 = 0.70$). Exchange times of up to approximately 600 ms are estimated well, after which the estimates level off. d is again very well estimated ($R^2 = 0.98$), with no significant under or overestimation bias.

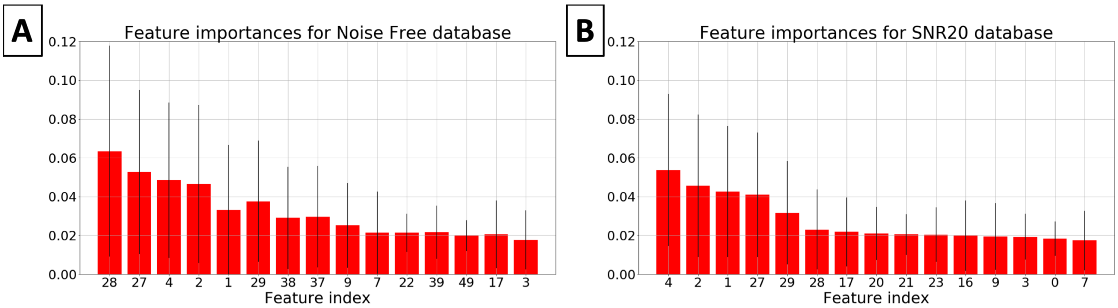

Figure 4.9: *Random forest feature importance*. Figure shows the 15 most informative features for the random forest when trained on the noise-free database (graph **A**), and when trained on the SNR20 database (graph **B**). The features are numbered from 1 to 60, the first 15 corresponding to the first shell of the protocol in Table 4.1, the next 15 to the second shell and so on. The features in each shell are arranged according to Table 4.2 . The higher the red bars on the *y-axis*, the more important the feature.

Figure 4.9 shows the feature importance results for the noise-free (Fig. *A*) and SNR=20 (Fig. *B*) training scenarios. The feature importances are computed using the *sklearn* toolkit in Python. We rank the 60 individual features of the training set (in our case the DTI and SH features of the DW-MRI signal) according to how informative they are during the estimation of the microstructure parameters. The higher the feature importance value on the x-axis, the more important the feature is, with all feature importance values adding to 1. We number the features from 0 to 59, the first 15 corresponding to the first shell, the next 15 to the second, and so on. For better visibility, Figure 4.9 only shows the 15 most informative shells. For both noise scenarios, we find that all features contribute to the parameter estimation by at least 1%, and that there are no features that contribute to the estimation significantly. For example, even the most informative feature contributes by only ≈0.060 in the noise-free case and by ≈0.055 in the noisy case. This suggests that there are no significant features driving the parameter estimation. The black lines, which indicate the standard deviation of the feature importances among the 120 decision trees, are very wide. This is expected, as the random forest is trained using only a subset of features to look for the best partition at each node, as discussed in the Methods section.

The most informative 6 shells in the noise-free scenario are three DTI features from the shell with the shortest Δ and three SH features from the shell with the longest Δ (Fig. 4.9A). The three DTI features are the second and third eigenvalue of the diffusion tensor and the fractional anisotropy, and the tree SH features are the anisotropy, skewness and kurtosis. In the presence of noise (Fig. 4.9B), the six most informative features remain the same, but their ranking changes, with the features from the short Δ shell being ranked ahead of those from the long Δ shell. This result can be explained by the presence of noise, which affects the shell with the shortest Δ the least and has a stronger effect on the shell with the longest Δ. Overall, these findings suggest that each feature plays a role in the parameter estimation, with no overall dominant feature. Moreover, as the most informative 6 features belong to the longest as well as the shortest Δ shells, this suggests that it is necessary to include a combination of both long and short Δ shells in our imaging protocol.

4.3.2 In-vivo experiments

4.3.2.1 Healthy subjects

Figure 4.10 shows scan and rescan parameter maps estimated using the random forest across representative sagittal, coronal and axial slices for the first healthy volunteer. An initial visual inspection suggests good agreement between the scan and rescan parameters, with similar trends observed across all the major white matter tracts. The values estimated for all parameters are within plausible ranges. Estimates of volume fraction f are in the range 0.44-0.65. The upper bound is slightly lower than expected, but as suggested by the simulation results in Figure 4.8, large f tends to be underestimated by the random forest when the data is noisy. However, we still see the expected high-low-high trend in f across the mid-sagittal CC.

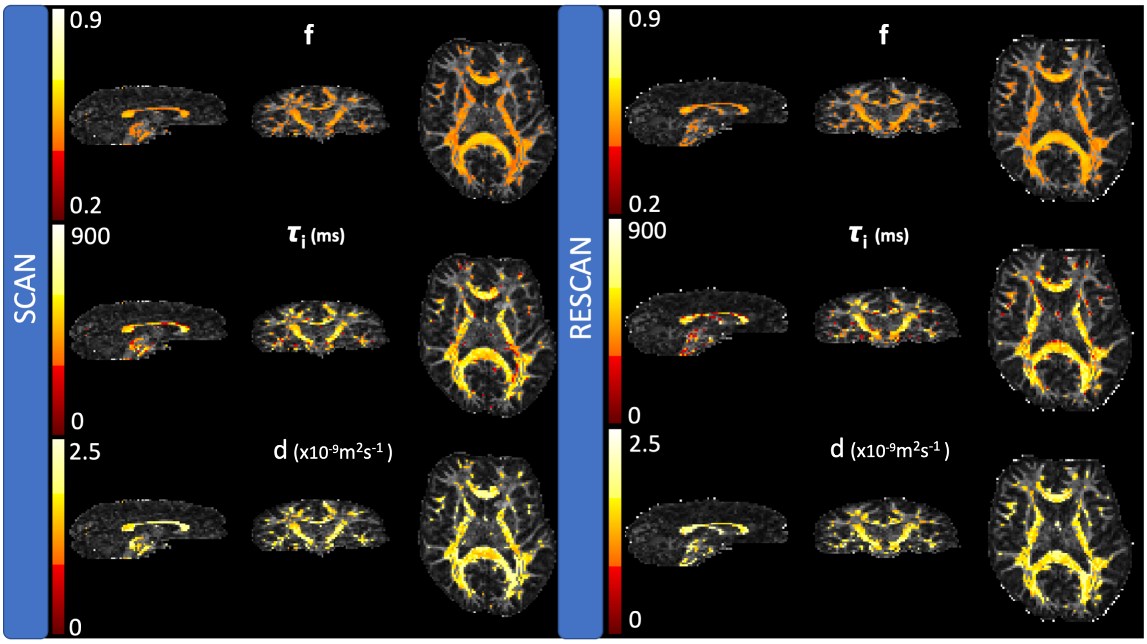

Figure 4.10: Scan (*left*) and rescan (*right*) parameter maps of f, τ_i and d estimated by the random forest model across representative sagittal, coronal and axial slices of data from volunteer 1.

Estimates of τ_i are consistently in the range 400-500 ms across the major tracts. The scan and rescan maps of d are also highly consistent, with estimations for most voxels in the range 1.4-1.8$x10^{-9}$ m^{2}s^{-1}.

Figure 4.11 shows equivalent parameter maps and scatter plots for volunteer 2. As for subject 1, the scan and rescan maps for the random forest model show good visual similarities for all parameters. We see the high-low-high trend in f, and the estimates of τ_i and d are consistent between scans as well as between the two subjects. The values estimated for all parameters are within plausible ranges and within the same range as for subject 1. Additional Bland-Altman plots of the scan-rescan parameters published in [132] confirm that there is no bias between scan and rescan estimates for either of the two volunteers.

In Figure 4.12, we analyse the reproducibility of the estimates across six ROIs. The ROIs are manually defined in the splenium (CC-S) and genu (CC-G) of the corpus callosum, the left (ALIC-L) and right (ALIC- R) anterior limbs of the internal capsule and the left (CST-L) and right (CST-R) corticospinal tracts on the scan and rescan data for both volunteers. The bottom row in Figure 4.12 shows an example ROI mask overlaid on the FA image of the scan data for volunteer 2. We

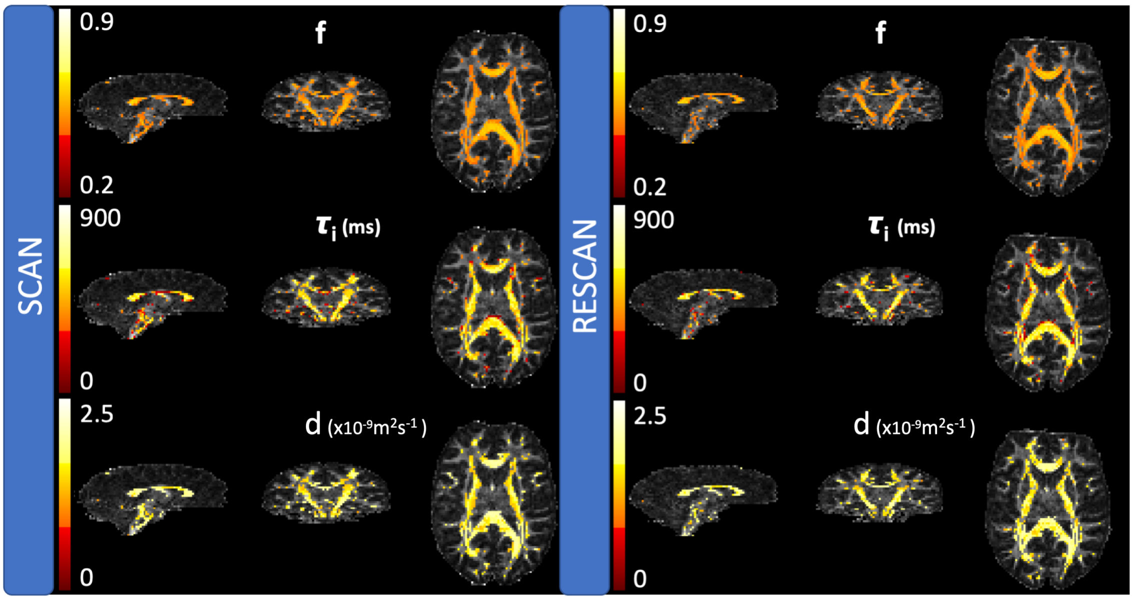

Figure 4.11: Scan (*left*) and rescan (*right*) parameter maps of f, τ_i and d estimated by the random forest model across representative sagittal, coronal and axial slices of data from volunteer 2.

calculate the mean of the parameter estimates from the random forest in each ROI for the scan and rescan for both subjects. This results in 24 measurements: 12 from the 6 ROIs in the scans of the two subjects and 12 from the 6 ROIs in the rescans of the two subjects. We plot the 12 scan mean parameter estimates against the 12 rescan mean parameter estimates in the top row in Figure 4.12. In addition to this, we compute the linear correlation coefficient ρ between the scan and rescan mean estimates. The scan-rescan scatter plots for f, τ_i and d in Figure 4.12 show high reproducibility for f ($\rho_f = 0.96$) and d ($\rho_f = 0.81$), and a good reproducibility for τ_i ($\rho_f = 0.60$) at the ROI level.

Finally, we pool together the scan and rescan parameter estimates for both subjects across the six regions of interest and calculate the mean μ and the standard deviation σ of each parameter in each region, which we show in Table 4.4. Due to the small number of voxels and subjects, we calculate the standard deviation in each region over all patients and scan/rescan altogether without accounting for the within/between subject variance. We find that the means across the 6 ROIs for f are in between 0.49 and 0.59. For τ_i, the ROI means lie in the range 490–606 ms, and for d, the ROI means are in between 1.6–1.95$x10^{-9}$ m^{2}s^{-1}. The parameter ranges

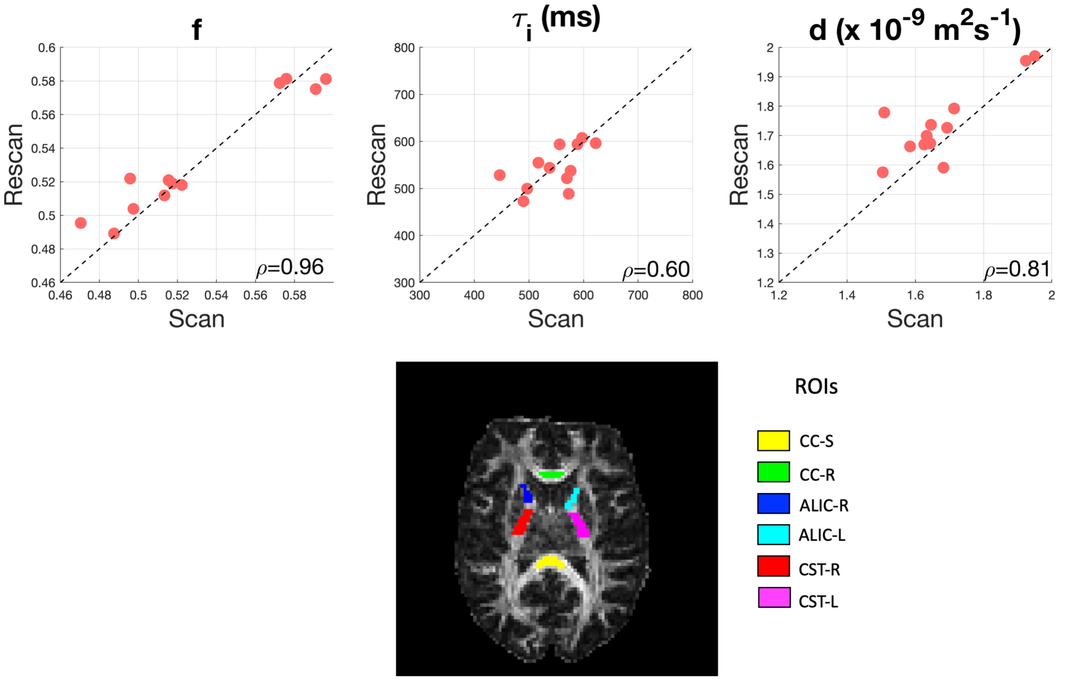

Figure 4.12: The *top row* shows the correlation of scan-rescan estimates of f, τ_i and d across 6 ROIs for both healthy volunteers. The image in the *bottom row* shows the six ROIs overlaid on the FA map of volunteer 2.

are all plausible for human white matter, as discussed further in the next section.

	f	τ_i (ms)	d (x10^{-9} m^{2}s^{-1})
ROI	$\mu(\sigma)$	$\mu(\sigma)$	$\mu(\sigma)$
CC-S	0.59(0.02)	564(54)	1.69(0.22)
CC-G	0.58(0.02)	606(44)	1.95(0.10)
ALIC-L	0.50(0.03)	513(124)	1.60(0.18)
ALIC-R	0.49(0.03)	490(119)	1.63(0.18)
CST-L	0.52(0.03)	564(80)	1.70(0.16)
CST-R	0.52(0.04)	541(85)	1.66(0.15)

Table 4.4: Mean and standard deviation of the random forest estimates of f, τ_i and d in all six ROIs across the scan and rescan of both healthy volunteers.

4.3.2.2 MS subjects

Figure 4.13 shows the parameter maps of the two MS subjects. The top row shows the estimates for the early stage MS subject, while the bottom row shows the estimates for the late stage MS subject. First from the left is an FA image with the overlaid lesion mask (red areas in squares) and respective normal appearing white matter tissue (NAWM) (green areas in circles). The remaining three columns are

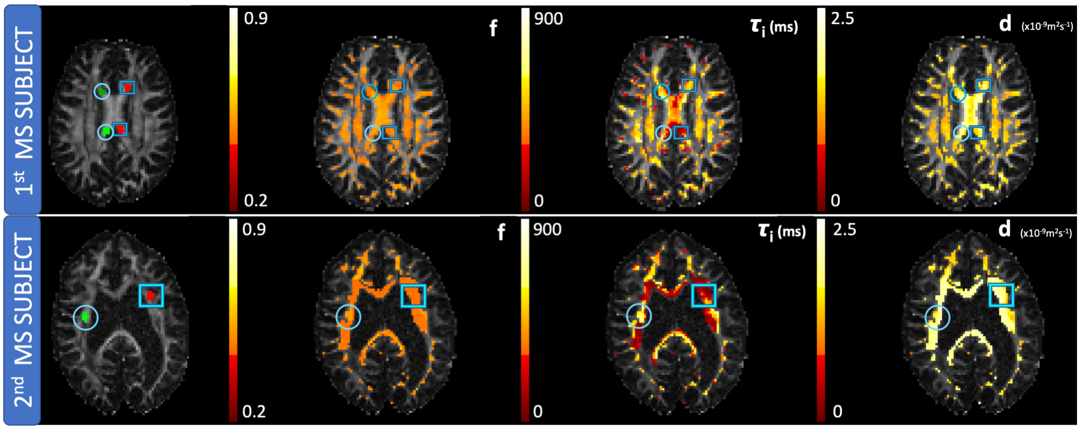

Figure 4.13: FA image (1^{st} column) and parameter maps of f, τ_i and d ($2^{nd}, 3^{rd}, 4^{th}$ columns) estimated by the random forest model across representative axial, coronal and sagittal slices of two MS patients. The scan of MS subject 1 in the *top row* shows two lesions (red areas in blue squares) in the genu (top of the scan) and the splenium (bottom of the scan) of the CC. The scan of MS subject 2 in the *bottom row* shows one lesion in the CST (red area in the square). For each lesion, its respective contralateral NAWM mask is shown as a green area in a circle.

parameter maps from the same slice estimated using the random forest model.

In the case of the early MS subject (top row), an initial visual inspection of parameter maps suggests that the random forest does not detect any obvious parameter differences in the lesion in genu (top area in a square), compared to the respective NAWM tissue (top area in a circle). In contrast, for the lesion in splenium (bottom area in a square), an initial visual inspection reveals differences when compared to the NAWM (bottom area in a circle), especially in the parameter maps for τ_i. τ_i estimated using the random forest is reduced relative to the NAWM area, and the volume fraction is also slightly lower when compared to the contralateral NAWM region.

In the case of the late MS subject (bottom row), the lesions are much more widespread, overlapping most of the white matter mask used to select voxels for analysis, including the genu and the corticospinal tracts. An initial visual inspection of the whole white matter mask shows more dramatic changes in the parameter maps compared to the first early stage MS subject. We notice a reduction in f and τ_i in the lesion area (red area in a square) compared to the contralateral NAWM (green area in a circle).

To provide a more quantitative analysis, we calculate the mean and the standard

	MS subject 1				MS subject 2	
	CC: Genu		*CC: Splenium*		*CST*	
	Lesion (9)	*NAWM (7)*	*Lesion (11)*	*NAWM (10)*	*Lesion (43)*	*NAWM (40)*
f	0.48(0.03)	0.49(0.03)	0.48(0.02)	0.50(0.03)	0.46(0.01) *	0.50(0.04)
τ_i	537(66)	491(83)	154(85) *	302(130)	138(92) *	595(50)
d	1.73(0.12) *	1.52(0.16)	1.64(0.21)	1.74(0.23)	1.99(0.07) *	1.77(0.17)

Table 4.5: Mean and standard deviation of the random forest estimates of f, τ_i (ms) and d ($x10^{-9}$ m^{2}s^{-1}) in MS lesions compared to contralateral NAWM of both MS patients. Lesion values which are statistically different ($p < 0.02$) from values in the corresponding NAWM are marked with the '$*$' symbol. Row 3 of the table indicates between brackets the number of voxels in each ROI.

deviation of the parameter estimates in the lesions shown in Figure 4.13 (red areas in squares), and their respective contralateral NAWM ROIs (green areas in circles). We show these in Table 4.5. In the genu lesion of MS subject 1, d is higher than in the NAWM. The difference between f and τ_i is not statistically significant. For the lesion in the splenium of MS subject 2, the random forest estimates a mean τ_i of 154 ms, compared to 302 ms in the contralateral NAWM. For the lesion in the CST of subject 2, the quantitative analysis confirms the initial visual inspection. The mean random forest estimates of f drops from 0.50 to 0.46, and τ_i is reduced from 595 ms in the NAWM to 138 ms in the lesion. We also notice an increase in d, although for d the change is the same order of magnitude as the standard deviation across the ROI. Lesion parameter values for which there is a statistically significant difference (p<0.02) to the values in NAWM are marked with stars.

4.4 Discussion

In this chapter, we have introduced a random forest based computational model that builds a mapping between features derived from DW-MRI signals and tissue microstructure parameters, including the intra-axonal exchange time τ_i. First, we use synthetic data to maximise the performance of the random forest by choosing the optimal number of decision trees, tree depth and training set size. Next, we establish a benchmark performance for our model by testing it on noise-free simulations, where we find strong correlations between the estimated and ground truth parameters (f: $R^2 = 0.88$, τ_i: $R^2 = 0.95$, d: $R^2 = 0.99$). We find that for realistic levels of noise (SNR=20), as present in our in-vivo data, the simulation performance is affected, however, the parameters are still well estimated (f: $R^2 = 0.71$, τ_i: $R^2 = 0.70$, d: $R^2 = 0.98$). Furthermore, we apply our model to in-vivo data from two healthy subjects and show that the random forest estimates sensible microstructure parameters. Finally, we generate parameter maps from scans of two MS subjects. In both subjects, we observe reductions in the random forest estimates of f and τ_i in lesions compared to the contralateral NAWM, which is consistent with expectations from the pathology of MS lesions. We expect τ_i to reduce due to myelin damage and f to decrease due to the breakdown of myelin as well as potential axonal loss, both changes characteristic of the MS pathology.

Simulations. We find that the setting which maximises the performance of the random forest is 120 trees of maximum depth 20, together with a training set of 10,000 unique DW-MRI feature vectors and their associated ground truth microstructure parameters. We show that the random forest model estimates in the noise-free scenario have very strong correlations with the ground truth values, providing an excellent benchmark performance for our model and imaging protocol (f: $R^2 = 0.88$, τ_i: $R^2 = 0.95$, α: $R^2 = 0.82$, d: $R^2 = 0.99$). We show that the addition of noise with SNR=20, matching our in-vivo data, does not affect much the estimation of d ($R^2 = 0.98$), however, it has a stronger effect on the estimation of τ_i and f (f: $R^2 = 0.71$, τ_i: $R^2 = 0.70$). Finally, we analyse the feature importances and show that all features contribute in small proportion, without any dominant features to drive the estimation. A potential explanation for this is that the features are not mutually independent and contain overlapping information (see Table 4.2), and, therefore, their importance is also shared. Moreover, we show that three DTI features from the shell with the shortest Δ (0.102 s) and three SH features from the shell with the longest Δ (0.412 s) are the six most informative shells. A potential explanation for this could be that the DTI features in the $\Delta = 102$ s shell are informative for the estimation of f and d, as they have a better SNR and because the estimation of f and d does not require long Δs, while the SH features in the $\Delta = 0.412$ s shell are more informative for the estimation of τ_i because it allows the probing of longer exchange times.

In-vivo data. The scan-rescan experiments in healthy subjects show that the random forest model has good reproducibility: we observe the same trends across the scan/rescan parameter maps (e.g. the high-low-high trend in f across the mid-sagittal CC), together with good linear correlation ρ between scan and rescan parameters ($\rho_f = 0.96, \rho_{\tau_i} = 0.60, \rho_d = 0.81$). The reproducibility of the random forest model is thoroughly compared with that of the Karger model [85], in [132]. The results show that our random forest model has better reproducibility, in simulations and in-vivo, at both voxel and ROI level.

The absolute values estimated by the random forest are inherently difficult to validate. Accurate estimates of τ_i are not obtainable via histology as tissue fixation is known to alter cell membrane properties. Nevertheless, an existing study of intra- to extra-axonal water exchange across the whole in-vivo rat brain using relaxometry and contrast agents suggests a mean τ_i of approximately 550 ms [11], aligning with our estimates. Furthermore, studies on sphingomyelin membranes (which are found in axonal membranes) estimate permeabilities of $\approx 1\ x10^{-6}\text{ms}^{-1}$ [9], which correspond to exchange times of 300-600 ms for axons with radii of 0.5-1 µm, a range within which the majority of our τ_i estimates lie (400-500 ms in the healthy subjects). The maps of f reveal the characteristic high-low-high trend across the mid-sagittal CC consistent in both scan and rescan maps, and the estimated values for this parameter lie in the range [0.4, 0.65], a plausible range for white matter tissue [73]. The upper bound is slightly lower than expected, but as suggested by the simulation results, large f tends to be underestimated when the data is noisy. The scan and rescan maps of d are highly consistent. The estimations for most voxels lie in the range 1.4-1.8 $x10^{-9}$ m^2s^{-1}, which is plausible for white matter tissue.

In MS subjects, we observe reductions in the random forest estimates of f and τ_i in lesions compared to NAWM. This is intuitive as the breakdown of myelin in lesions is likely to make the exchange time shorter compared to the NAWM (as illustrated in Figure 4.5). The reduction in f is also intuitive as MS lesions are often characterised by axonal loss [155, 156, 157]. This, together with the breakdown of the myelin sheath would lead to an increase in the extra-axonal space and a decrease in the overall intra-axonal volume fraction f. When comparing the NAWM in MS subjects to the same ROIs in healthy subjects, we observe that the trend continues, i.e. we observe reductions in f and τ_i in the NAWM of MS subjects compared to the healthy tissue in the two volunteers. For example, we find that in the CC-S of the healthy subjects, the estimated τ_i is 564±54 ms, while in the CC-S of the MS patient τ_i is 302±130 ms in the NAWM and 154±85 ms in the lesion. It has been shown previously that NAWM in MS subjects is somewhat damaged compared to the normal tissue [158] and our results support this suggestion. We also find almost no difference between the parameters in the lesion and the NAWM tissue for the first MS subject. This could be because the lesion was relatively new compared to the other lesion and the damage has not yet fully developed. These are just preliminary results because the number of both MS and healthy subjects is quite small. However, the trends are plausible and the estimated parameters sensible. We show some promising new results, however a larger study of MS patients is necessary in order to investigate whether the parameters of the random forest model, in particular τ_i, could act as suitable biomarkers for detecting and tracking changes in MS pathology.

Limitations. One of the limitations of the work in this chapter is that the training uses a simplified model of white matter tissue. Specifically, the model uses long straight circular cylinders that mimic axon bundles and does not account for myelin water, curvature of axons, dispersion or crossing fibres. Nevertheless, the general idea we present here, of computational models based on simulations and machine learning extends naturally to more complex simulations. Once more realistic simulations that account for more effects (e.g. undulation [159], dispersion [146, 147] or myelin water [148, 86]) become available, they can easily be incorporated in the machine learning framework used in this thesis and form part of future work.

A further limitation of the analysis in this chapter is the size of the healthy and diseased cohorts. For example, the ROI standard deviation reported in Table 4.4 does not take into account the within/between subject variance and treats the sample as being independent, despite some of the estimates coming from repeated measurements. While this choice is motivated by a small number of subjects and therefore estimates per ROI, this problem can be avoided by increasing the number of subjects.

Another limitation is that the machine learning procedure is tested on the same type of substrates that it was trained on. We note, however, that although the substrate model is the same, the model parameters between the training and the test data are different so the results demonstrate generalisation to some extent. Once more complex simulations become available, we can extend the testing of our model to more diverse simulations. Another important future improvement which accounts for unfamiliar test data is to incorporate measures of uncertainty as in [129].

The performance of our computational framework is dependent on the sensitivity of the DW-MRI protocol to the parameters of interest. This can be seen in the simulation results, where for τ_i greater than 600 ms, the estimation performance decreases. Indeed, the findings in Chapter 3 show that for the Δ, G and the b-value of our STEAM protocol, we can only distinguish exchange time effects for τ_i around 650-700 ms in substrates representative of the brain white matter tissue (see Figure 3.4A). Another parameter that only weakly influences the DW-MRI signal is the volume-weighted axon radius index α. We show that in the noise-free scenario our random forest model can estimate the axon radius index α reasonably well ($R^2 = 0.82$). However, the addition of noise (SNR=20) severely affects the protocol's sensitivity to this parameter [138], and *Nedjati et al.* [132] show that the random forest is only able to distinguish between small and large axons in the presence of noise. The parameter maps of α are extremely noisy and the reproducibility is very low. The sensitivity of the signal and hence the reproducibility of our parameter estimates can potentially be improved by optimising the acquisition protocol further, e.g. by increasing the maximum diffusion time or gradient strength to provide better contrast (also subject to SNR constraints) or using the random forest model in the protocol optimisation instead of the Karger model. This limitation, though critical for the performance of our model, is intrinsic to the sensitivity of the data rather than to the machine learning based computational framework introduced in this thesis.

4.5 Conclusions

In this chapter, we introduce the first machine learning based computational model with permeability for white matter microstructure imaging, using a random forest regressor that builds a mapping between rotationally invariant features of DW-MRI signals and microstructure parameters. We demonstrate the idea by testing the model's ability to estimate the intra-axonal exchange time τ_i, the intra-axonal volume fraction f and the intrinsic diffusivity d in simulations and in clinical data from two healthy and two MS patients. Using realistic simulations, we show that our model has a very good benchmark performance and, even though the performance is affected by the presence of noise, the parameters continue to be well estimated. Using in-vivo data from the healthy subjects, we show that our computational model has good scan-rescan reproducibility and provides parameter estimates within the plausible range for human white matter tissue. Using the in-vivo data from the two MS subjects, we show that the random forest estimates in MS lesions are in line with the expected MS pathology (significant decrease in f and τ_i and a small increase in d), demonstrating the clinical potential of this new technique.

The computational modelling approach we use here opens doors to estimating a wide range of other parameters for which mathematical models are intractable such as undulation or properties of the extracellular space. Although the mapping we learn is specific to randomly packed, parallel, non-abutting cylinders and a STEAM protocol, the approach easily extends to other tissue configurations and pulse sequences as well as to other machine learning algorithms. In the next chapter, we explore our machine learning computational framework's extensibility to neural networks.

Chapter 5

Neural network based computational models with permeability: microstructure parameter estimation in clinical in-vivo human data

In the previous chapter we introduced the first machine learning based computational model with permeability for white matter microstructure imaging. To demonstrate the idea, we use a random forest and test the model on two healthy and two MS patients. An advantage of the computational framework introduced in this thesis is that it can easily be adapted to other machine learning approaches. In this chapter, motivated by the promising results shown by neural networks in many domains including medical imaging [7], we extend our computational framework to incorporate a neural network model, which we test on the synthetic and in-vivo clinical data in Chapter 4.

5.1 Motivation

While random forests have many advantages such as interpretability, easiness of tuning and robustness to noise [108], recent advances in deep learning [7] have shown that neural networks can outperform conventional machine learning algorithms in medical imaging tasks such as Image Quality Transfer [129]. Prompted by these recent advances, we aim to test the feasibility of neural networks for microstructure parameter estimation. For this, we implement a simple and widely used neural network model, a multilayer perceptron (MLP), and assess its performance on the synthetic and in-vivo clinical human data in Chapter 4.

One of the advantages of neural networks over other machine learning models is that their architecture can easily be adapted to include other optimisation objectives in order to enhance the overall performance of the model. For example, a neural network architecture can be modified to include local information (in convolutional neural networks), which, for our task, is available when estimating from in-vivo scans. In addition to this, neural network architectures can be modified to perform domain adaptation. This is usually employed in scenarios when there are differences between the training and the test data. This is particularly relevant for our task, as our model is trained on simplified simulations of the in-vivo data. Therefore, neural networks can potentially be used to bridge the gap between simulations and in-vivo data. In this thesis, we start by developing a simple neural network model and by assessing its performance in the field of microstructure parameters estimation. Extending the analysis to more complex architectures such as domain adaptation and convolutional neural networks, which are undergoing continuous development, forms part of future work.

5.2 Methods

In this section, we first provide a brief outline of the imaging protocol and the datasets used for the experiments in this chapter. Next, we describe the architecture of our neural network model. Finally, we outline the synthetic and in-vivo experiments designed to test the performance of the model.

5.2.1 Imaging protocol and data

For both the synthetic and the in-vivo data, we use the STEAM imaging protocol introduced in the previous chapter (Ch. 4). The protocol is optimised for a two compartment model with exchange and includes 4 b-shells with the Δ ranging from 102 ms to 412 ms. Full details of the protocol can be found in Table 4.1.

For the data analysis, we use the synthetic and the in-vivo datasets described in Chapter 4. Our synthetic data is formed of 12,500 unique rotationally invariant signal-derived feature vectors and their associated ground truth microstructure parameters. The feature vectors are extracted from Monte Carlo simulations of the DW-MRI signal from substrates mimicking human white matter. More details about the dataset are provided in Section 4.2.2. We generate two synthetic databases, one from the noise-free DW-MRI simulations and another one by adding Rician noise with a standard deviation σ corresponding to an SNR of 20, matching the level of noise present in our in-vivo data. Our in-vivo dataset is formed of scan-rescan data from two healthy volunteers and scan data from two MS patients. For more details regarding the data acquisition and preprocessing, see Section 4.2.3.

5.2.2 Machine learning

5.2.2.1 Neural networks for microstructure parameter estimation

The neural network used in this chapter is a multilayer perceptron (MLP), which is a feed-forward neural network, formed of an input layer, a number of middle layers called *hidden layers* and an output layer as illustrated in Figure 5.1. In our case, the input layer takes as an input the rotationally invariant feature vector of the DW-MRI signal, and then passes it on to the hidden layers for processing until the output layer is reached. The output layer is formed of 3 neurons, each one providing an estimation of one of the three tissue parameters of interest (f, τ_i and d).

During training, the network learns a mapping between microstructure parameters and features of the DW-MRI signals. The neural network learns the correct mapping during training by minimising the sum of the mean squared error (MSE) between the ground truth microstructure parameters y and the estimated parameters $\hat{y}$ over the training set: $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, where n is the size of the training set. To minimise the MSE, the neural network parameters are adjusted after each epoch of training by computing the gradient of the error with respect to each parameter. The network stops learning once the MSE stops decreasing after each epoch of training. Once the training is finished, we use the optimised neural network for estimating tissue microstructure parameters from unseen feature vectors.

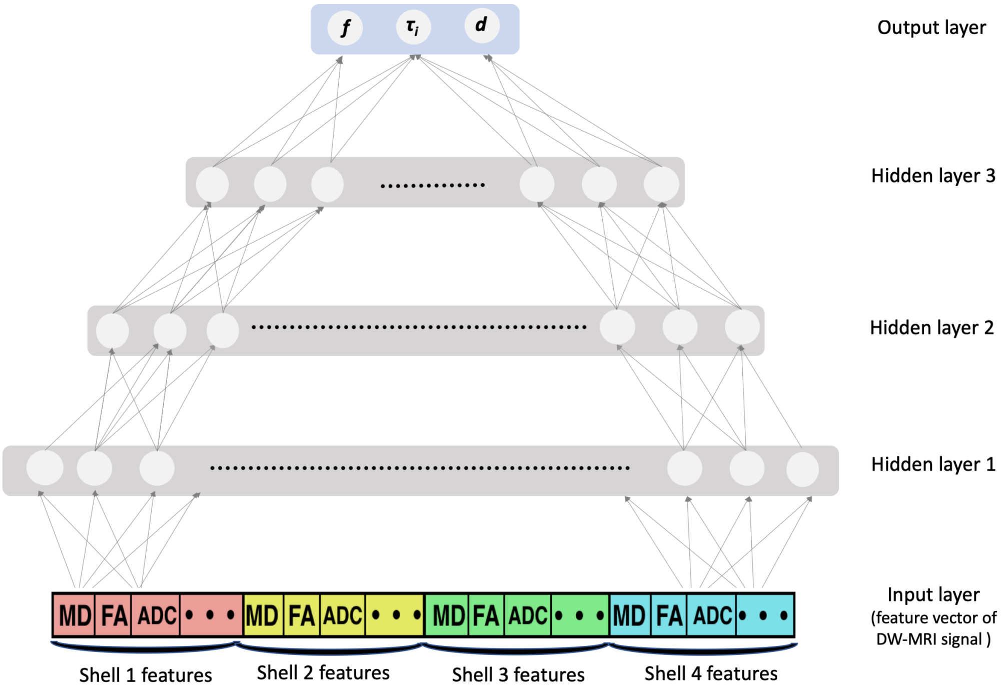

Figure 5.1: Schematic illustration of the neural network used in this chapter. The neural network is formed an input layer, three hidden layers and an output layer. The input layer takes as input the feature vector, which is then passed on and processed by the hidden layers. The output layer is formed of three neurons, one for the estimation of each of the three tissue parameters of interest (f, τ_i and d).

5.2.2.2 Implementation details

We implement the neural network using the open source Keras toolkit [160] with Tensorflow backend [161]. We employ a common protocol for the training of the network. We minimise the loss using stochastic gradient descent for 100 epochs with learning rate 0.07, decay 1e-6, and momentum 0.9. We use a 3 hidden layer architecture with 150, 75 and 75 hidden neurons with 'tanh' (hyperbolic tangent) activations and dropout 0.7. The number of layers was chosen empirically, by starting with one layer and increasing the number of layers until no improvement in performance on the validation set could be seen. This is common practice in the machine learning field due to the absence of a theoretical framework for choosing the number of layers of a neural network. We select the best performing architecture as the one that minimises the mean squared error on a validation set formed of 2,000 DW-MRI synthetic signals.

We train the neural network to estimate three microstructure parameters of interest: the intra-axonal volume fraction f, the intra-axonal exchange time τ_i and the intrinsic diffusivity d. For this, we use our synthetic database of 12,500 signal-derived feature vectors and their associated ground truth tissue parameters acting as labels during the supervised learning task. We use 10,000 feature vectors for training and 2,500 for testing. We train two neural networks, one using the noise-free database and another using the SNR=20 database. For the in-vivo estimation, we use the neural network on the SNR=20 database as the one that most closely resembles the in-vivo data.

5.2.3 Experiments

5.2.3.1 Synthetic experiments

To assess how well the neural network parameter estimates match the known ground truth values in the test set, we show correlation plots and compute the correlation coefficient R^2 between the model estimations and the ground truth parameter values. The R^2 score measures the linear relationship between two datasets and takes values between -1 and $+1$, with ± 1 implying an exact positive/negative correlation and 0 implying no correlation. This is calculated as in Equation 4.1. Furthermore, to investigate any general bias in the estimations, we show Bland-Altman graphs, which plot the means of the ground truth and estimated values against the difference between them. Points on the correlation and Bland-Altman plots are colour-coded with colour bars showing the percentage error. For better visibility, we limit the colour bar to –50 to 50%, with points outside this range taking the same colour as the two maximum points.

We compute the correlation scatter plots and the Bland-Altman plots for both the noise-free and SNR=20 databases. The noise-free scenario provides a benchmark for the best estimation we can achieve given the data available, while the SNR=20 performance provides an indication of the estimation under noise conditions matching those in the in-vivo data.

5.2.3.2 In-vivo experiments

Healthy subjects

We generate estimates of three microstructure parameters of interest (f, τ_i and d) across white matter voxels in the scans and rescans for the two healthy subjects.

We show parameter maps of f, τ_i and d across representative axial, coronal and sagittal slices of both scan and rescan data. We use the scan-rescan data from the two healthy volunteers to investigate intra-subject reproducibility for six manually defined ROIs: the splenium (CC-S) and genu (CC-G) of the corpus callosum, the left (ALIC-L) and right (ALIC-R) anterior limbs of the internal capsule, and the left (CST-L) and right (CST-R) corticospinal tracts. We calculate the mean parameter estimates from the neural network in each ROI for both the scan and rescan data and show correlation scatter plots and Bland-Altman plots for f, τ_i and d individually. For highly reproducible parameters the points on a Bland-Altman plot should be closely clustered about the zero difference line. In addition to this, we show the equivalent mean parameter estimates from the random forest model in Chapter 4 to assess comparatively the reproducibility of the two machine learning algorithms. Finally, we show neural network and random forest parameters maps of f, τ_i and d across the same axial slice of the second volunteer.

MS subjects

To investigate the sensitivity of the neural network to tissue damage in MS lesions, we analyse the model estimates in lesions marked by an expert clinical researcher. We also mark additional ROIs in the NAWM for comparison. In the first subject, in the early stages of MS, two of the marked lesions overlap the white matter mask completely and are investigated further. For the second subject, in the late stage of MS, the lesions are much more widespread, overlapping most of the white matter mask used to select voxels for analysis. However for this subject, our analysis is limited by the availability of contralateral NAWM for comparison and thus we only use one lesion in the CST. For both subjects, we assess the statistical significance of the difference in estimates between the lesion and the NAWM using a t-test. Finally, we compare the mean parameter estimates from both machine learning models in the lesions and NAWM ROIs.

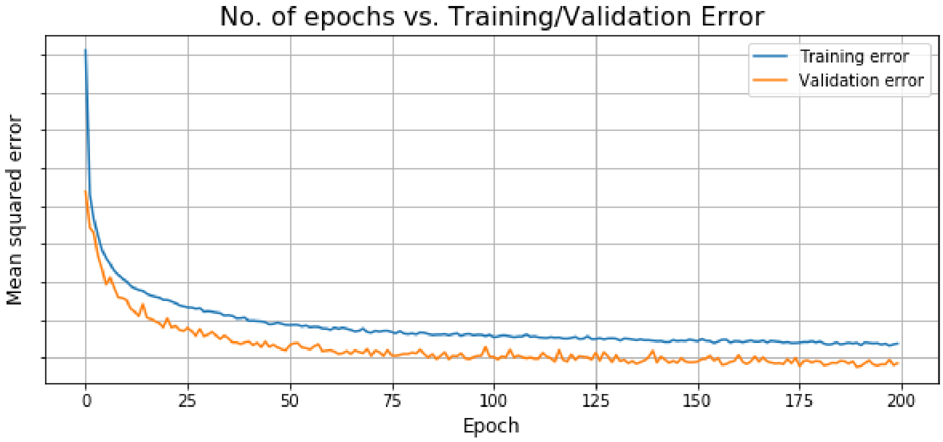

Figure 5.2: Figure shows how the training and the validation error change with each epoch of training.

5.3 Results

5.3.1 Synthetic Experiments

Figure 5.2 shows how the training and the validation mean squared errors change with each epoch of training. The training error (blue curve) shows that the error drops significantly during the first 50 epochs of training, after which it continues to decrease, but by smaller amounts, suggesting that the neural network is still learning. The validation error is commonly used to asses whether the network learns to generalise from the training data or whether it is overfitting. Since the validation set is unseen during training, if the validation error decreases we can say that the model generalises better to new data. Conversely, if the validation error starts to increase, this suggests that the model is learning the training data too closely and is unable to generalise to new unseen data, a process known as overfitting. The validation error curve (orange) in Figure 5.2 drops significantly during the first 50 epochs, after which it decreases mildly and finally flattens out around epoch 100. This suggests that the first 100 training epochs improve the performance of the neural network, after which the performance stays constant. Consequently, we set the number of training epochs to 100.

Figure 5.3 shows correlation scatter plots (*top row*) and the corresponding Bland-Altman plots (*bottom row*) of f, τ_i and d for the neural network in the noise-free scenario. The data points are colour-coded according to how close the es-

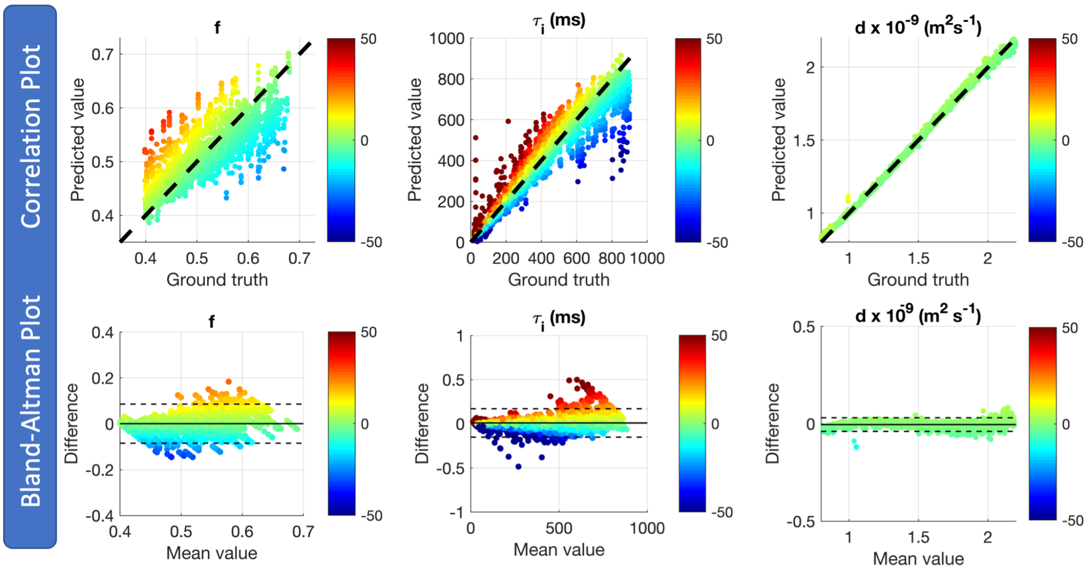

Figure 5.3: Neural network performance on the noise-free database. The *top row* shows correlation scatter plots comparing the ground truth values of f, τ_i and d with the neural network estimations. The *bottom row* shows the equivalent Bland-Altman plots for f, τ_i and d. The *y-axis* shows the difference between the ground truth and the estimated values and the *x-axis* shows their mean.

timates are to the actual values and the percentage error is shown on the colour bars. The solid black line in the Bland-Altman plot indicates the mean difference between the ground truth and the estimated parameters and the dashed lines show the 95% limits of agreement. The correlation scatter plots in the first row show a strong correlation between the ground truth and the estimated parameter values: $f : R^2 = 0.81, \tau_i : R^2 = 0.95$ and $d : R^2 = 0.99$. In the corresponding Bland-Altman plots, points are clustered around the zero difference line, indicating low bias. However, despite the absence of noise, the recovery of the parameters is not perfect. In particular, for f and τ_i, there is some bias in the estimated values which depends on the ground truth value. Similar to the random forest model, the neural network overestimates small values of τ_i and f, and underestimates large values. The noise-free performance provides a benchmark for the best estimation we can achieve given the data available.

Figure 5.4 shows similar plots, but for the synthetic database with SNR=20. The results for all parameters are consistent with those obtained from the noise-free data, however, the 95% limits of agreement are slightly wider due to the presence of noise. Although the mean difference lines are very close to zero, we again see

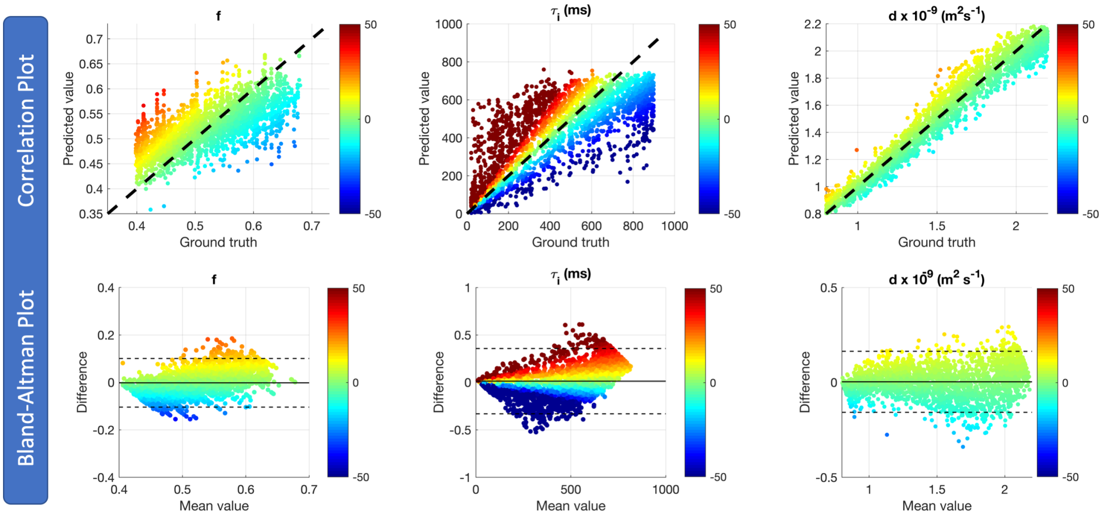

Figure 5.4: Neural network performance on the SNR=20 database. The *top row* shows correlation scatter plots comparing the ground truth values of f, τ_i and d with the neural network estimations. The *bottom row* shows the equivalent Bland-Altman plots for f, τ_i and d. The *y-axis* shows the difference between the ground truth and the estimated value and the *x-axis* shows their mean.

the same bias as in the noise-free case, but more pronounced. For f ($R^2 = 0.71$), we see that large volume fractions tend to be underestimated slightly whereas low volume fractions are slightly overestimated. This is also the case for τ_i ($R^2 = 0.74$). Exchange times of up to approximately 700 ms are estimated well, after which the estimates level off. The 700 ms upper limit is 100 ms higher than for the random forest, for which the estimates level off around 600 ms. This is reflected quantitatively in the R^2 score for τ_i, which is 0.04 greater than the random forest equivalent. d is again very well estimated ($R^2 = 0.98$), with no significant under or overestimation bias.

	SNR=20			SNR=∞		
	f	τ_i	d	f	τ_i	d
NN	0.71	0.74	0.98	0.88	0.95	0.99
RF	0.71	0.70	0.98	0.81	0.95	0.99

Table 5.1: Neural network (NN) and random forest (RF) correlation coefficients (R^2 scores) for the synthetic databases (noise free and SNR=20).

Table 5.1 summarises the performance of the neural network and random forest (from the previous chapter) on the noise-free and SNR=20 databases. In the noise-free scenario, the neural network and the random forest have the same performance in the case of τ_i and d. The neural network estimates f better than the random forest in the noise-free scenario, however, this difference in performance is cancelled by the presence of noise (SNR=20).

5.3.2 In-vivo Experiments

5.3.2.1 Healthy subjects

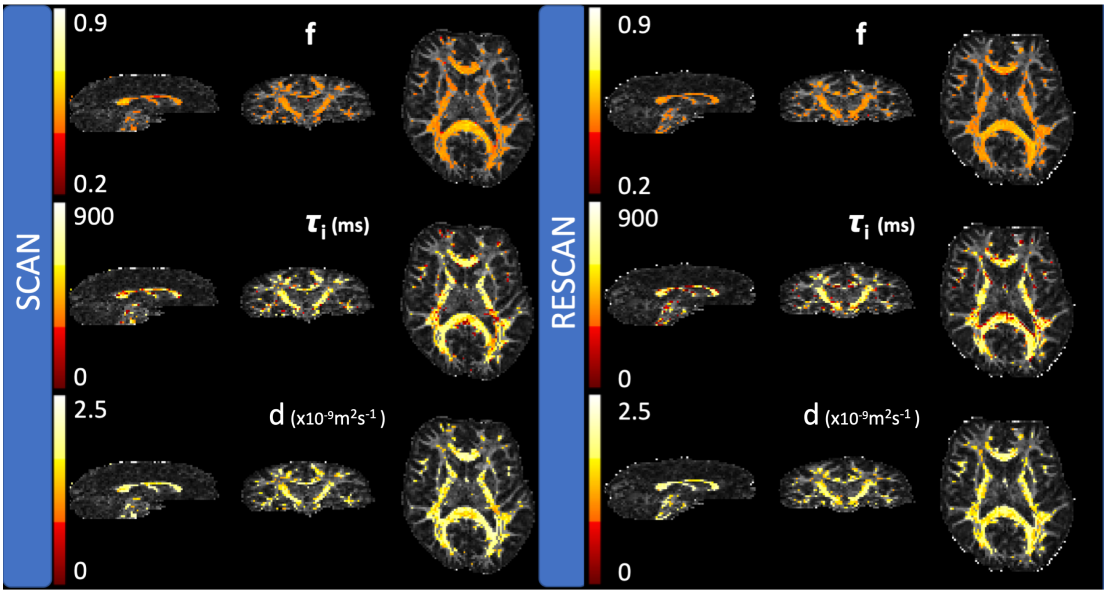

Figure 5.5: Scan (*left*) and rescan (*right*) parameter maps of f, τ_i and d estimated by the neural network model across representative sagittal, coronal and axial slices of data from volunteer 1.

Figure 5.5 shows scan and rescan parameter maps estimated by the neural network across representative sagittal, coronal and axial slices for the first healthy volunteer. The parameter maps reveal a good agreement between the scan and rescan estimations and very similar trends across the major white matter tracts. The values estimated for all parameters are within plausible ranges for the human white matter. Estimates of the volume fraction f are in the range 0.4-0.65. The upper bound is slightly lower than expected, but, as suggested by the simulation results in Figure 5.4, large f tends to be underestimated when the data is noisy. However, we still see the expected high-low-high trend in f across the mid-sagittal CC. Estimates of τ_i are consistently in the range 450-650 ms across the major tracts. The scan and rescan maps of d are also highly consistent, with estimations for most voxels in the range 1.4-2 $x10^{-9}$ m^{2}s^{-1}.

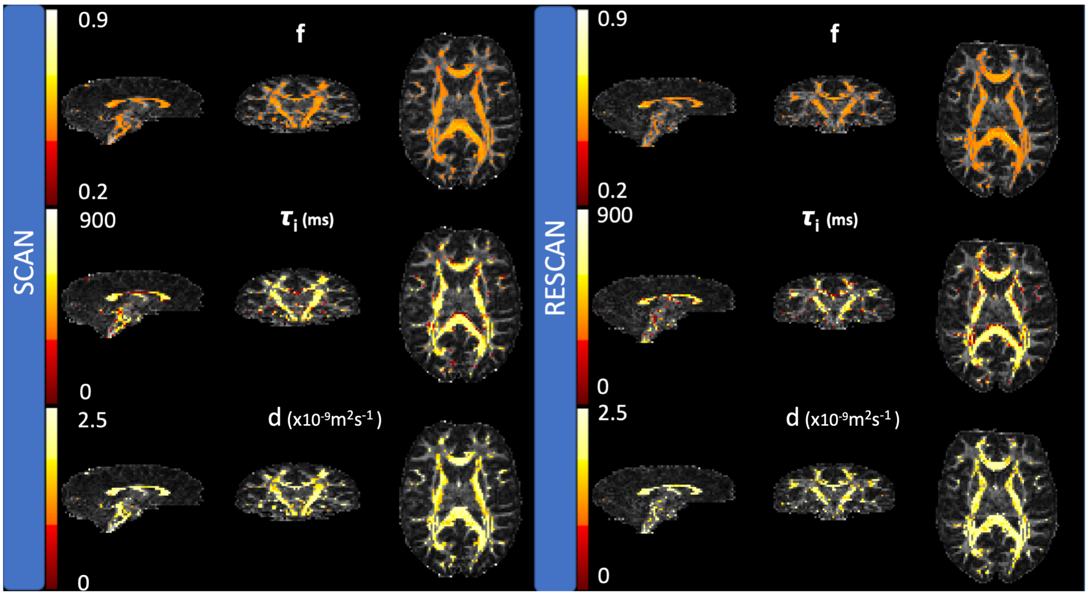

Figure 5.6: Scan (*left*) and rescan (*right*) parameter maps of f, τ_i and d estimated by the neural network model across representative sagittal, coronal and axial slices of data from volunteer 2.

Figure 5.6 shows equivalent parameter maps and scatter plots for volunteer 2. As for subject 1, the scan and rescan maps for the neural network model show good visual similarities for all parameters. We see the high-low-high trend in f, and the estimates of τ_i and d are consistent between scans as well as between the two subjects. The values estimated for all parameters are within plausible ranges, and within the ranges of subject 1 estimates.

In Figure 5.7, we analyse the reproducibility of the estimates across the six ROIs in the white matter. We calculate the mean of the parameter estimates from the neural network in each ROI for the scan and rescan for both subjects. This results in 24 measurements: 12 from the 6 ROIs in the scans from the two subjects and 12 from the 6 ROIs in the rescans of the two subjects. The top row of Figure 5.7A shows scatter plots of scan versus rescan estimates of f, τ_i and d. The bottom row of Figure 5.7A shows the equivalent Bland-Altman plots, which show the mean of the scan and rescan parameter estimates against their difference. Dashed lines indicate the 95% limits for each method. Results from the neural network model (*blue circles*) and the random forest in the previous chapter (*pink circles*) are shown on the same plots.

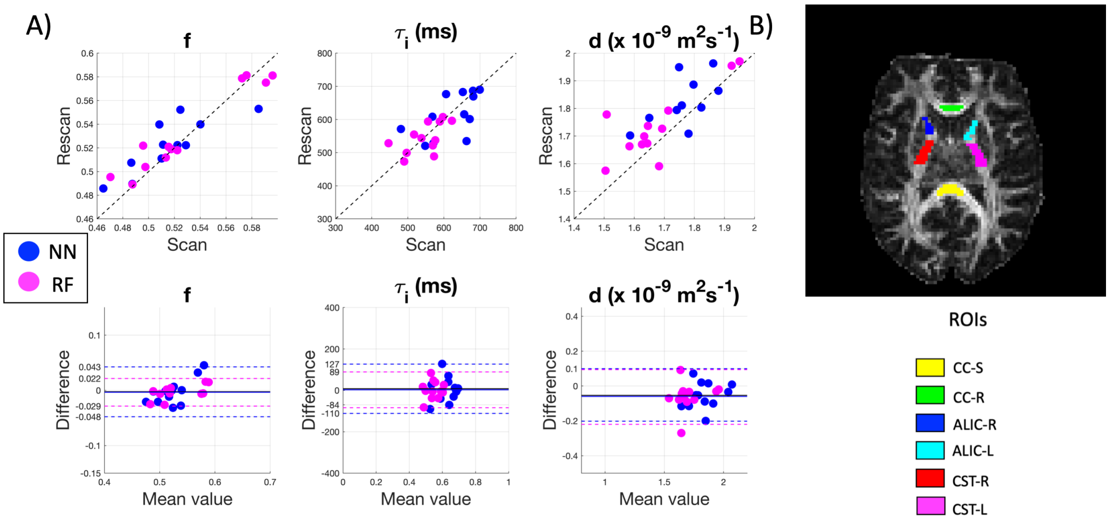

Figure 5.7: **A)**: The *top row* shows the correlation of scan-rescan estimates of f, τ_i and d across 6 regions of interest for both healthy volunteers. The *bottom row* shows the equivalent Bland-Altman plots. **B)** The six regions of interest used in the analysis overlaid on the FA map of volunteer 2.

The neural network shows good reproducibility for all parameters at the ROI level. There is good agreement between the estimates of the two machine learning approaches. The neural network estimates of τ_i are slightly higher than those of the random forest. However, the differences are all within one standard deviation from the ROI mean (as computed in Table 5.2 below) and in line with the synthetic data results which show that the neural network estimates of τ_i level off around 700 ms while those of the random forest level off around 600 ms. For f and d, the range of estimated values is the same for the two machine learning approaches.

For f and τ_i, the data points for the random forest are closer to the zero line than the data points for the neural network model. For f, the 95% limits of agreement between scan-rescan are [-0.048, 0.043] for the neural network and [-0.029, 0.022] for the random forest, narrower by 44%. For τ_i, the scan-rescan 95% limits of agreement are [-110, 127] ms for the neural network, while those of the random forest are 27% narrower ([-89, 84] ms). These results indicate that the scan-rescan random forest estimates of f and τ_i show better agreement than those of the neural network. For d, the data points from both the neural network and random forest show a very similar spread about the zero line, indicating that the methods have equivalent reproducibility.

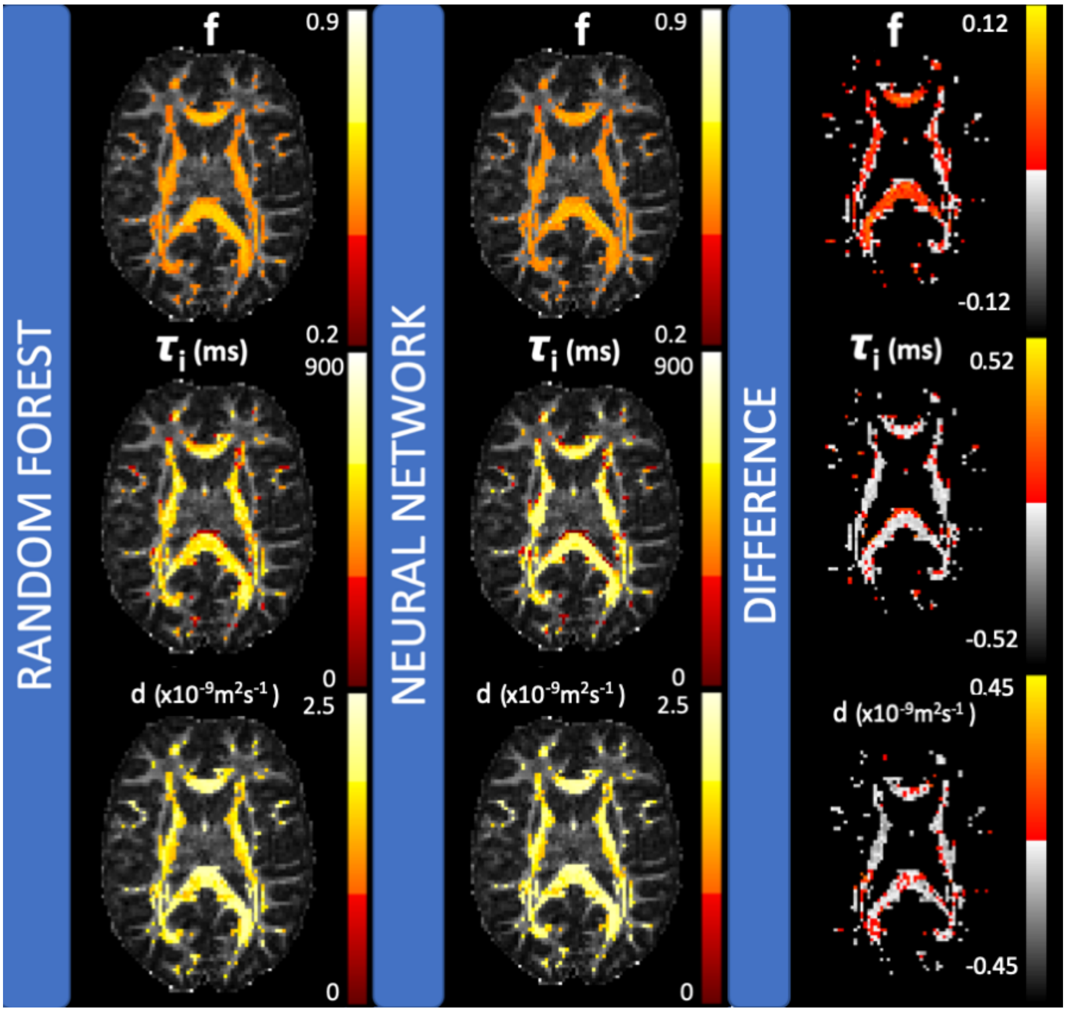

Figure 5.8: Parameters maps showing the estimates of f, τ_i and d across a representative axial slice of subject 2, generated using the random forest (*left column*) and the neural network (*middle column*). The *right* column shows the difference between the random forest and the neural network parameter map. This column reveals that, on average, the random forest estimates higher values for f and lower values for τ_i and d than the neural network.

Figure 5.8 shows parameter maps across the same axial slice of volunteer 2 for the random forest (*left column*) and for the neural network (*middle column*). A visual inspection reveals that there is a very good agreement between the estimates of two approaches as well as between the detected white matter trends (i.e. low-high-low f trend in the splenium of the CC). The τ_i maps reveal that both approaches estimate very fast exchange (red) in the same regions at the edge of the white matter mask, where we expect partial volume effects. On average, the random forest estimates higher values of f and lower values of τ_i and d compared to the neural network.

Finally, we pool the scan and rescan parameter estimates for both subjects across the six ROIs, across which we calculate the mean μ and the standard deviation σ of each parameter. We show these results in Table 5.2. We find that the means across the 6 ROIs for f are between 0.49 and 0.59. For τ_i, the ROI means lie in the range 541–651 ms. For d the ROI means are between 1.64–2.01 $x10^{-9}$ $\mathrm{m^2s^{-1}}$. The parameter ranges are all plausible for human white matter, as discussed further in the next section.

	f	τ_i (ms)	d ($x10^{-9}$ $\mathrm{m^2s^{-1}}$)
ROI	$\mu(\sigma)$	$\mu(\sigma)$	$\mu(\sigma)$
CC-S	0.59(0.05)	606(101)	1.79(0.21)
CC-G	0.55(0.06)	651(39)	2.01(0.09)
ALIC-L	0.51(0.09)	571(175)	1.64(0.20)
ALIC-R	0.49(0.09)	541(144)	1.69(0.19)
CST-L	0.53(0.08)	623(95)	1.82(0.17)
CST-R	0.52(0.09)	610(112)	1.76(0.17)

Table 5.2: Mean and standard deviation of the neural network estimates of f, τ_i and d in all six regions of interest across the scan and rescan of both healthy subjects.

5.3.2.2 MS subjects

Figure 5.9 shows the parameter maps of the MS subjects. The top row shows the estimates for an early stage MS subject, while the bottom row shows the estimates for a late stage MS subject. First from the left is an FA image with the overlaid lesion mask (red areas in squares) and the contralateral NAWM (green areas in circles). The remaining three columns are parameter maps from the same slice estimated using the neural network model.

In the case of the early MS subject (top row), an initial visual inspection of the parameter maps suggests that the neural network does not detect any obvious parameter differences in the lesion in the genu (top area in a square), compared to the respective NAWM tissue (top area in a circle). In contrast, for the lesion in the splenium (bottom area in a square), an initial visual inspection reveals differences in f and τ_i when compared to the NAWM (bottom area in a circle), especially in the parameter maps for τ_i. The neural network estimates of τ_i are reduced relative to the NAWM area, and the volume fraction is also slightly lower when compared

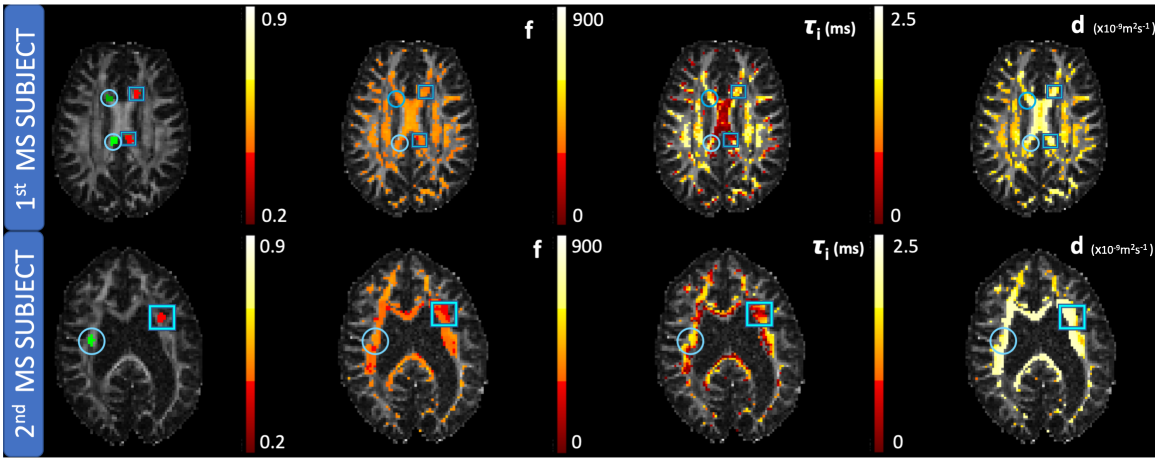

Figure 5.9: FA image (1^{st} column) and parameter maps of of f, τ_i and d ($2^{nd}, 3^{rd}, 4^{th}$ columns) estimated by the neural network model across representative axial, coronal and sagittal slices of two MS patients. The scan of MS subject 1 in the *top row* shows two lesions (red areas in blue squares) in the genu (top of the scan) and the splenium (bottom of the scan) of the CC. The scan of MS subject 2 in the *bottom row* shows one lesion in the CST (red area in the square). For each lesion, its corresponding contralateral NAWM mask is shown as a green area in a circle.

to the contralateral NAWM region. In the case of the late MS subject (bottom row), an initial visual inspection of the whole white matter mask reveals more dramatic changes in the parameter maps compared to the early stage subject. It can be noticed that τ_i estimates in the body of the CC are lower than in other regions of the MS subjects. This could be explained by partial volume effects due to the CSF in the adjacent ventricles, especially as the ventricles become enlarged as can be seen in the late MS subject in the bottom row. We notice a reduction in f and τ_i in the lesion area (red area in a square) compared to the contralateral NAWM (green area in a circle). For both subjects, the trends are in perfect agreement with those estimated by the random forest model in the previous chapter.

To provide a more quantitative analysis, we calculate the mean and the standard deviation for all parameter estimates in the lesions shown in Figure 5.9 and their respective contralateral NAWM, which we show in Table 5.3. In addition to this, to aid comparison between the estimates of the two machine learning approaches, we include the results of the random forest model (from Table 4.5). For the CC-G lesion in the first MS subject, the only statistically significant difference estimated is in d, which is higher in the lesion than in NAWM. The differences in f and τ_i

	MS subject 1				**MS subject 2**	
	CC: Genu		*CC: Splenium*		*CST*	
NN	*Lesion*	*NAWM*	*Lesion*	*NAWM*	*Lesion*	*NAWM*
f	0.48(0.02)	0.49(0.02)	0.47(0.04)	0.49(0.05)	**0.45(0.01)** *	0.50(0.01)
τ_i	531(78)	497(92)	**91(84)** *	267(180)	**204(180)** *	616(50)
d	**1.88(0.13)** *	1.61(0.19)	1.74(0.19)	1.83(0.26)	**2.14(0.03)***	1.87(0.20)

RF	*Lesion*	*NAWM*	*Lesion*	*NAWM*	*Lesion*	*NAWM*
f	0.48(0.03)	0.49(0.03)	0.48(0.02)	0.50(0.03)	**0.46(0.01)** *	0.50(0.04)
τ_i	537(66)	491(83)	**154(85)** *	302(130)	**138(92)** *	595(50)
d	**1.73(0.12)** *	1.52(0.16)	1.64(0.21)	1.74(0.23)	**1.99(0.07)** *	1.77(0.17)

Table 5.3: Mean and standard deviation of the neural network (*top table*) and random forest (*bottom table*) estimates of f, τ_i and d in MS lesions compared to contralateral NAWM of both MS patients. Lesion values which are statistically different ($p < 0.02$) from values in the contralateral NAWM are marked with the '$*$' symbol and highlighted in *bold*.

are not statistically significant. For the CC-G lesion of the same subject, the mean τ_i estimate in the CC-S lesion is 91 ms, compared to 267 ms in the contralateral NAWM, confirming the initial visual inspection of the parameter maps. For the second, late-stage MS subject, the neural network estimates a reduction in f from 0.50 to 0.45, and a significant drop in τ_i, from 616 ms in the NAWM to 204 ms in the lesion. We also notice an increase in d, although for d the change is the same order of magnitude as the standard deviation across the ROI. Lesion parameter values for which there is a statistically significant difference ($p<0.02$) to the values in NAWM are marked with stars.

Both the neural network and the random forest detect the same pairs of lesions/NAWM with statistically significant differences. The absolute values of the mean estimates are also very close between the two machine learning approaches, with only very minor differences, all smaller than the standard deviation of the mean. We also observe no significant differences in the magnitude of the standard deviations between the neural network and the random forest, for the majority of parameters across both subjects. There is only one instance in the CST lesion, where the standard deviation of the neural network estimates of τ_i is twice as large as that of the random forest. However, this could be explained by the small number of voxels in this particular region.

5.4 Discussion

Synthetic experiments We find that the neural network estimates in the noise-free scenario have very strong correlations with the ground truth values, providing an excellent benchmark performance for our model and imaging protocol f: $R^2 = 0.81$, τ_i: $R^2 = 0.95$, d: $R^2 = 0.99$. We show that the presence of noise with SNR=20, matching that of our in-vivo data, has a negligible effect on the estimation of d ($R^2 = 0.98$), and a stronger effect on the estimation of τ_i and f (f: $R^2 = 0.71$, τ_i: $R^2 = 0.74$). Nevertheless, the performance of the neural network is still sufficiently good, with values of τ_i up to approximately 700 ms estimated well and centred around the diagonal line (Figure 5.4).

In-vivo experiments. The scan-rescan experiments in healthy subjects show that the neural network has good reproducibility with highly consistent trends across the white matter (e.g. the high-low-high trend in f across the mid-sagittal CC). As discussed in the previous chapters, the absolute values of the estimations are inherently difficult to validate, especially in the case of τ_i. Existing studies report values of τ_i between 300 and 600 ms for axons with radii of 0.5–1 μm. The mean estimated τ_i in the healthy subjects' ROIs lie at the higher end of this range, with the CC-S and CST-L means above 600 ms (CC-S: $\mu = 651 \pm 39$ ms, CST-L: $\mu = 623 \pm 95$ ms). This could be explained by the presence of axons larger than 1 μm in the actual human white matter tissue [16], in which we expect longer values of τ_i, as the water molecules take longer to leave the intra-axonal space. The maps of f reveal the characteristic high-low-high trend across the mid-sagittal CC, consistent in both scan and rescan maps. The estimated values for this parameter lie in the range [0.4, 0.65], a plausible range for white matter tissue [73]. Similar to the random forest results, the upper bound is slightly lower than expected, but as suggested by the simulation results, large f tends to be underestimated when the data is noisy. The scan and rescan maps of d are highly consistent. The estimations for most voxels lie in the range 1.4-2 $x10^{-9}$ $\mathrm{m}^2\mathrm{s}^{-1}$, which is plausible for white matter tissue.

In MS subjects, we observe trends in line with the expectations of demyelinating pathologies: a significant reduction in f and τ_i in the lesions compared to the NAWM. We expect to see a decrease in the exchange time as the water molecules encounter less barriers when leaving the intra-axonal space due to the destruction of the myelin layer that surrounds the axons. In addition to this, we also expect to see a decrease in f due to the destruction of the myelin sheath and potential axonal loss characteristic of MS lesions [155, 156, 157]. When comparing the NAWM in MS subjects to the same ROIs in healthy subjects (presented in Table 5.2) we observe that the trends continue and the neural network estimates a decrease in f and τ_i. This is in line with expectations that there is underlying damage in the NAWM compared to healthy tissue [158]. For example, we find that in the CC-S of the healthy subjects, the mean estimated τ_i is 606 ms, while in the CC-S of the MS patient the mean τ_i is 267 ms in the NAWM and 91 ms in the lesion. In line with the results in Chapter 4, we find negligible differences between the parameters in the lesion and the NAWM tissue for the first MS subject which can potentially be explained by the early stage of the disease. As in the case of the random forest, these are just preliminary results because the number of both MS and healthy subjects is small. However, the trends are plausible and the estimated parameters sensible.

Random forest comparison. The performance of the neural network and random forest on synthetic data with realistic levels of noise (SNR=20) is very similar. Both approaches show no general estimation bias, and a local overestimation/underestimation bias for small/large values of f and τ_i. We observe no difference between the correlation coefficients for f and d between the two machine learning approaches. In the case of τ_i, the R^2 score of the neural network is slightly higher than that of the random forest: $R^2 = 0.74$ for the neural network and $R^2 = 0.70$ for the random forest. By visually inspecting the correlation plots, we can see that the neural network estimates well exchange times up to 700 ms, after which the estimates level off (Figure 5.4). The estimates of the random forest, on the other hand, level off around 600 ms (see Figure 4.8). These results suggest that the neural network learns the relationship between the noisy synthetic DW-MRI signals and the exchange time slightly better than the random forest.

In our in-vivo analysis, we find a very good agreement between the estimates of the neural network and the random forest. For the scan-rescan data of the healthy subjects, the mean f and d estimates of the two approaches are within the same range. For τ_i, the neural network estimates are slightly higher in some ROIs than those of the random forest. The neural network estimates mean τ_i values up to 700 ms, while the random forest estimates mean values up to 600 ms. This can be explained by our simulation results, which show that the estimates of the random forest level off around 600 ms, while those of the neural network level off at 700 ms. Nevertheless, the differences between the two machine learning approaches are within one standard deviation of the mean. Moreover, a visual inspection of the parameter maps for the two approaches reveals a very good agreement between the two (e.g. the same areas of very fast exchange in the axial slice of the brain).

For the MS subjects, the two approaches detect the same statistically significant changes between the lesions and NAWM for both subjects. In addition to this, there is a very good agreement between the ROI mean estimates of the neural network and of the random forest, with all differences within a standard deviation from the mean. When comparing the reproducibility of the two approaches (Figure 5.7), we find that the two methods have equivalent reproducibility for d, while for f and τ_i, the random forest has a better reproducibility. For τ_i, the 95% agreement lines in the Bland Altman plots are slightly narrower for the random forest (by 27%). A potential explanation for this could be that the random forests estimates cannot be outside the parameter values in the training data, while the estimations of the neural network are unbounded. This results in wider ranges over which the neural network scan-rescan estimates can differ, potentially reflected in wider 95% agreement lines.

5.5 Conclusions

In this chapter, we extend our computational framework in Chapter 4 to a neural network approach. We demonstrate the idea by testing the model's ability to estimate f, τ_i, and d in simulations and in clinical data from two healthy and two MS patients. First, using simulations mimicking white matter tissue, we show that the neural network has a very good benchmark performance, and that despite the presence of noise, the parameters are still well estimated. For τ_i, the neural network estimates well exchange times up to 700 ms, providing a slight improvement over the equivalent random forest results in the presence of noise (R^2=0.74 for the neural network vs R^2=0.70 for the random forest). Next, using in-vivo scans of healthy and MS subjects, we show that our neural network model provides parameter estimates within the plausible range for human white matter tissue and in line with expectations from the MS pathology (a significant decrease in f and τ_i and a mild increase in d). The in-vivo estimations of the neural network are in very good agreement with the random forest estimations, with all differences within a standard deviation from the mean. Furthermore, we show that the two machine learning approaches have similar reproducibility when estimating d, while in the case of f and τ_i, the random forest has a better reproducibility than the neural network.

These are just preliminary results because the number of both MS and healthy subjects is quite small. However, the trends measured by both machine learning approaches are plausible and the estimated parameters sensible. We show some promising new results, however a larger study with available histological data is necessary in order to investigate whether the parameters estimated by our computational framework, in particular τ_i, could act as suitable biomarkers for detecting and tracking changes in MS pathology. We aim to address this issue in the next chapters, by testing our computational framework on a controlled mouse model of demyelination with available histological data.

Chapter 6

Random Forest based computational models with permeability: Mouse model of demyelination - Sequence sensitivity, optimisation and synthetic testing

The previous two chapters introduced machine learning based white matter models with permeability, using a random forest (Chapter 4) and a neural network (Chapter 5) approach. We demonstrate, using Monte Carlo simulations and in-vivo healthy and MS patient data, that this is a promising new approach for permeability estimation, providing good reproducibility and plausible parameter estimates for human white matter tissue.

In order to be able to quantitatively investigate the performance of our computational framework, we design experiments that use a preclinical cuprizone mouse model of demyelination, which is extensively used in the MS literature due to its similarity to the processes occurring in MS lesions [162]. We then compare the estimates of our random forest and neural network models to available histological data.

These experiments are presented in the following three chapters. The first chapter (Chapter 6) presents the synthetic experiments which investigate the sensitivity of the preclinical mouse imaging protocol, optimise the random forest performance with respect to the protocol, and test the simulation performance of the random forest given the new protocol and synthetic dataset. The second chapter (Chapter 7) experimentally investigates the in-vivo performance of the random forest through a direct comparison with histological electronmicroscopy (EM) data from the same mice. The third chapter (Chapter 8) presents an experimental study of the synthetic and in-vivo performance of the neural network model using the preclinical mouse data.

The overall purpose of the three chapters is to experimentally investigate machine learning based computational models with permeability using a controlled mouse model of demyelination through a direct comparison with histology.

Parts of this work are featured in the recently submitted paper to *NeuroImage*: *Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination I. Hill et al. [163].*

6.1 Motivation

In Chapter 4 we introduce for the first time a machine learning approach using a random forest and a database of rotationally invariant features for the in-vivo estimation of axonal permeability. The study in [132], covering part of the results in Chapter 4, shows that our random forest computational model outperforms the most widely used analytical model with exchange [92, 6, 40], the Kärger model [85], in both synthetic and in-vivo human data, by providing more reproducible and robust estimates of τ_i. In addition to this, the model also improves on previous computational models with permeability by being independent of fibre orientation and through using a machine learning approach capable of generalising to new unseen data. However, the new approach is tested in-vivo only qualitatively, using just two healthy and two MS patients, and needs further validation. Furthermore, the study hypothesises that τ_i is linked with demyelination in MS lesions, and does not show whether other underlying processes such as axonal swelling or orientation dispersion affect the estimates. The work in this Chapter, together with Chapter 7, aims to address these issues and experimentally validate the random forest approach introduced in Chapter 3.

The aim of this chapter is threefold. Firstly, we aim to investigate the sensitivity to the exchange time of the SDE protocol used to acquire the mouse in-vivo data in order to ensure there is enough information in the data. Secondly, as the SDE protocol uses an explorative range of sequence parameters, we aim to select the most informative shells (i.e. b-values and directions) with respect to the exchange time to optimise the performance of our model. We also investigate how the performance of the random forest is affected by the level of noise and by the number of shells used for training. Thirdly, we test the performance of the random forest in the noise-free and SNR=40 case using a previously unseen synthetic test set. While the noise-free dataset provides a benchmark performance for our model, the SNR=40 case provides an indicative performance of the model on the in-vivo mouse data, which is affected by similar levels of noise. To run the experiments in this chapter, we use Monte Carlo simulations mimicking white matter substrates representative of the mouse in-vivo data we analyse in Chapter 7.

6.2 Methods

This section introduces the diffusion imaging protocol and synthetic data used in this chapter. Next, it outlines the sensitivity analysis experiments, the shell selection procedure and the training and testing of the random forest.

6.2.1 Diffusion imaging protocol

We use the same SDE protocol for the synthetic data in this chapter as for the in-vivo data in the next two chapters. We do not use a STEAM protocol as in the previous two chapters as the preclinical dataset in the remaining chapters was aquired using a protocol that was originally designed to maximise signal reconstruction accuracy under realistic time constraints [164]. Our imaging protocol has 25 shells, each with one b=0 measurement and a different combination of diffusion gradient strength G and diffusion gradient separation Δ, as summarised in Table 6.1. The resulting protocol has 345 measurements in total, diffusion gradient duration $\delta = 5$ ms, $|G_{max}| = 500$ mTm^{-1} and shell b-values as shown in Table 6.1. Additional protocol details are as follows: $TE = 33.6$ ms, $TR = 2$ s, $FOV = 16x16$ mm, matrix size = 160x160, number of slices = 5, slice thickness = 0.5 mm.

Δ (ms) **G (mT/m)**	**10.8**	**13.1**	**15.4**	**17.7**	**20**	**# grad dirs**
150	358	445	533	620	707	**16**
200	620	775	930	1086	1241	**16**
300	1384	1733	2083	2432	2781	**8**
400	2489	3110	3731	4352	4973	**11**
500	3892	4862	5833	6803	7773	**13**

Table 6.1: SDE parameters with the corresponding nominal b-values in s/mm^2.

6.2.2 Synthetic data

To construct our computational model, we first generate two synthetic databases: one formed of simulated DW-MRI signals and one formed of rotationally invariant features derived from the simulated signals. Each entry in the database corresponds to a unique digital phantom which mimics the in-vivo data and for which the ground truth microstructure parameters are known. Each synthetic database is used to train a machine learning algorithm, here a random forest, to build a mapping between the signal or features and the corresponding ground truth microstructure parameters.

6.2.2.1 Synthetic signals database

We use Monte Carlo simulations of the DW-MRI signal to build our synthetic training database. The signals are generated using the open source Camino ([140]; *http://camino.cs.ucl.ac.uk*) simulation framework [98] together with the imaging protocol in Table 6.1. Each simulated signal corresponds to a digital phantom which mimics in-vivo mouse brain white matter. The digital phantoms are represented by synthetic substrates that model white matter as a collection of 100,000 non-abutting, parallel cylinders with gamma-distributed radii, a common choice in the brain literature [16]. The cylinders are randomly packed in the substrates as described in [98], with example substrates shown in Figure 6.1. We construct a database of 11,000 unique tissue substrates and their corresponding DW-MRI signals by randomly sampling from a range of histologically plausible substrate parameters for white matter tissue [16, 71]. A white matter synthetic substrate is defined through five parameters: the mean $\mu_R \varepsilon [0.2, 1]$ µm and the standard deviation $\sigma_R \varepsilon [min(0.1, \mu_R/5), \mu_R/2]$ µm of the axon radii distribution, the intra-axonal volume fraction $f \varepsilon [0.4, 0.7]$, the intra-axonal exchange time $\tau_i \varepsilon [2, 1000]$ ms and the intrinsic diffusivity $d \varepsilon [0.8, 2.2]$ $\mu m^2 ms^{-1}$. To ensure the convergence and the high precision of the simulated signals, we generate our synthetic database using 100,000 spins and 2,000 time steps [98]. The permeability of a substrate is specified within the Camino simulation framework via the probability parameter p. This parameter expresses the probability that a spin steps through a membrane encountered during the random walk (instead of always being reflected backwards as it is the case for impermeable substrates). Similarly to the other Chapters, the relationship between τ_i and p is given by the expression $p = \frac{R}{2 * \tau_i} \sqrt{6 \frac{\delta t}{d}}$, where d is the intrinsic diffusivity, R is the axon radius and δt is the temporal resolution of the simulation.

To maximise the performance of our machine learning regressor, we aim to build a training database that resembles as closely as possible the in-vivo data. For this, we generate an additional set of synthetic signals to account for the noise present in the in-vivo data. We add Rician noise with a standard deviation σ corresponding to an SNR of 40, which reflects the noise level of the b=0 images with the longest Δ.

6.2.2.2 Synthetic features database

In addition to the signals database, we construct a rotationally invariant feature database as introduced in Table 4.2 from Chapter 4. We compute the DTI and the 4th order SH fit for each shell from the synthetic signals using the Camino toolkit [140]. We then derive 15 rotationally invariant features for each shell and build an equivalent rotationally invariant synthetic database. The first five signal-derived features are calculated from the DTI fit and are the three eigenvalues $\lambda 1, \lambda 2, \lambda 3$, the mean diffusivity (MD) and the fractional anisotropy (FA). The remaining ten features are derived from the SH fit: the mean, peak, anisotropy, skewness and kurtosis of the apparent diffusion coefficient together with the peak dispersion and simple combinations of the first, second and fourth order spherical harmonics.

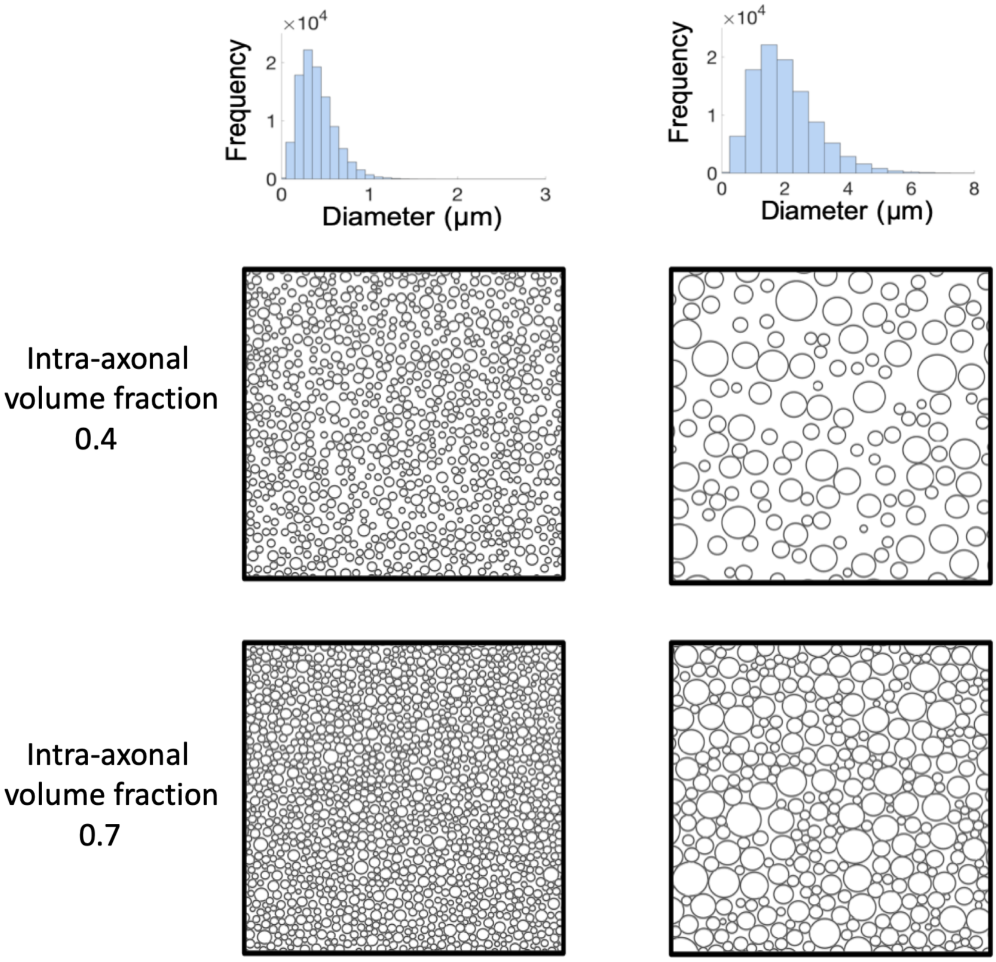

Figure 6.1: Histograms of two example axon diameter distributions used to generate synthetic substrates for our Monte Carlo simulations (first row). Figure also shows four digital tissue substrates corresponding to the two example axon diameter distributions and two different intra-axonal volume fractions: 0.4 (second row) and 0.7 (third row).

6.2.3 Machine learning

6.2.3.1 Random Forest

Following from Chapter 4, we build a random forest regressor that learns a mapping between the synthetic training database of DW-MRI signals/features and the ground truth microstructure parameters of the corresponding substrates. The mapping is learnt through a greedy splitting process of the input space (the synthetic signals/features) guided by the associated tissue parameters provided as labels during training.

There are two important parameters that need to be optimised to improve the learning performance of a random forest: the number of trees and the maximum tree depth. The number of trees determines the smoothness of the decision boundary, and the tree depth parameter specifies the maximum levels that each decision tree can have. Too large a value can lead to overfitting while too low a value can lead to underfitting, depending on the complexity of the data. Here, we run preliminary experiments and optimise these two parameters for our task in order to maximise the performance of our model.

6.2.3.2 Training and testing

We implement a random forest regressor using the scikit-learn open source Python toolkit [153]. Following preliminary experiments, we build a random forest with 200 trees of maximum depth 20 and bagging, as the setting the maximises the performance of the model. More general implementation details can be found at *http://scikit-learn.org/*. We train the random forest for a multi-parameter regression task: we estimate the intra-axonal exchange time τ_i together with the intra-axonal volume fraction f and the intrinsic diffusivity d. We do not fit the axon radius index [73] due to the lack of sensitivity of the signal to this parameter for our imaging protocol [138].

The dimensionality of our synthetic databases is 345 for the each signal in the database and 375 for each feature vector. We set the size of the training set to 10,000 as we did not find any improvements in performance above this number. The length of each synthetic training sample is reduced further during training according to the number of shells selected in each training scenario. We train and test the random forest on the synthetic databases using the associated ground truth parameters as labels for the supervised regression task. When predicting the parameter maps for the in-vivo data, we train the random forest using the noisy databases as they are a more accurate representation of the in-vivo data. We split our synthetic database into a training set of 10,000 randomly selected signal/feature vectors and a test set formed of the remaining previously unseen 2,500 signal/feature vectors. As shown in [132], the random forest is not biased by the random selection of the training data as long as there is sufficient coverage of the parameter range, which we also ensure.

We adopt two different training scenarios for our experiments. The first one uses the raw signal database for training, and the second one uses the rotationally invariant feature database obtained as in Section 6.2.2.2. While the first approach builds a direct mapping between the raw signals and the ground truth microstructure parameters, the second approach introduces an additional step of model fitting and constructs a mapping between the DTI and SH features of the raw signals and the microstructure parameters of interest.

6.2.4 Experiments

6.2.4.1 Sensitivity analysis

To ensure there is enough information in the data, we investigate the sensitivity of our SDE protocol to the intra-axonal exchange time by looking at the range of τ_i values for which the DW-MRI signal can be distinguished from that of an impermeable substrate. For this, we consider two synthetic substrates representative of mouse white matter tissue, with the following properties: the mean axonal diameter $\mu_D = 0.4$ μm and $\mu_D = 2$ μm, mimicking small and large axons in the CC [71], the intra-axonal volume fraction f =0.7 [71], and the intrinsic diffusivity $d = 1.2$ $\mu\text{m}^2\text{ms}^{-1}$ [165]. These substrates are a good representation of our in-vivo mice data, as shown by the histological measurements of μ_D in the next Chapter, all within the range of the gamma-distributions above. Using the Camino toolbox, we generate synthetic signals for each substrate and different values of δ, Δ, G, corresponding to the shells in our SDE protocol. The diffusion gradients are set perpendicular to the cylinders in the substrate to maximise sensitivity to τ_i. We investigate whether exchange time effects can be detected in the signal by looking at the difference in the normalised DW-MRI signal between impermeable ($\tau_i = \infty$) and permeable substrates. Moreover, we analyse the effect of noise by looking at a range of different SNRs: SNR=∞, SNR=40 and SNR=20, where SNR=40 corresponds to the level of noise present in our in-vivo data. By using synthetic substrates representative of our in-vivo data and the same imaging protocol, we expect the analysis in this section to provide an indicative range of exchange time values for which there is reasonable sensitivity in our in-vivo data.

6.2.4.2 Shell selection

As our imaging protocol uses an explorative range of imaging parameters, we select the shells that maximise the performance of our random forest model with respect to τ_i. For this, we evaluate the performance of our random forest model for every possible combination of 4, 9 and 16 shells out of the 25 in our protocol. We first evaluate combinations of 4 shells using as a benchmark the number of shells of the STEAM protocol in Chapter 4 optimised [72] for a two-compartment model with exchange and biophysically plausible tissue parameters. As there are 12,650 possible combinations of 4 shells, we train the random forest 12,650 times, once on each different shell combination. Then, for each training scenario corresponding to a unique combination of shells, we compute the correlation coefficient R^2 for f, τ_i and d between the ground truth and the estimated values in the test set. Finally, we sort the different shell combinations according to their R^2 score for τ_i, and choose the combination with the highest score as the one that maximises the performance of the model.

Furthermore, we investigate the effect of increasing the number of shells used for training. For this, we also look at combinations of 9 shells, as the minimum number of shells required to sample independently every unique G and Δ value in our SDE protocol. Additionally, we look at combinations of 16 shells as a middle value between the 9-shell and the full protocol scenario. For this analysis, we use the synthetic feature-based dataset described in Section 6.2.2.2. Finally, we investigate the effect of noise on the performance of our model. For this, we look at a range of different SNRs: SNR=∞, SNR=40 and SNR=20, where SNR=40 corresponds to the level of noise present in our in-vivo data.

6.2.4.3 Synthetic experiments

To assess the quality of the random forest estimates after training is completed, we compute the Pearson correlation coefficient R^2 between the ground truth values and the random forest estimates of the parameters in the previously unseen test set. To evaluate any potential bias in the estimates, we use Bland-Altman plots showing the mean of the ground truth and estimated parameter values against their difference. We first analyse the performance of the model on the noise-free synthetic databases to establish a benchmark given our data and imaging protocol. Next, we apply our machine learning model to the SNR=40 database as an indicative in-vivo model performance. For each experiment, we analyse both training scenarios outlined in Section 6.2.2 (signal-based and feature-based) to test whether there are any differences in performance between the two approaches.

6.3 Results

6.3.1 Sensitivity Analysis

Figure 6.2 shows the range of exchange time values for which the DW-MRI signal S(τ_i) can be distinguished from that of an impermeable substrate S($\tau_i = \infty$) in the presence of noise. For this, we calculate the change in signal $|S(\tau_i = \infty) - S(\tau_i)|$ between a permeable $S(\tau_i)$ and an equivalent impermeable $S(\tau_i = \infty)$ substrate. To illustrate practically achievable sensitivities, we plot this difference against three noise levels, denoted by the black plane: SNR=∞ (1st column), SNR=40 (2nd column) and SNR=20 (3rd column). Figure 6.2A illustrates the results for a substrate mimicking large axons in the white matter ($\mu_D = 2$ µm), while Figure 6.2B corresponds to a substrate with smaller axons ($\mu_D = 0.4$ µm). The second column shows that, for substrates with large axons (row A) and an SNR of 40, matching that of our in-vivo data, it is possible to distinguish exchange time effects for $\tau_i \leq 400$ ms. For substrates with small axons (row B), we can distinguish only permeable substrates with exchange times up to $\tau_i \leq 250$ ms. As expected, when the SNR drops to 20, it becomes harder to distinguish between impermeable and permeable substrates. This trend can be observed in the 3rd column, where the range for distinguishable permeable substrates narrows from $\tau_i \varepsilon [0,400]$ ms to $\tau_i \varepsilon [0,200]$ ms for large axons, and from $\tau_i \varepsilon [0,250]$ ms to $\tau_i \varepsilon [0,140]$ ms for small axons.

6.3.2 Shell selection

As our original 25-shell SDE protocol uses an explorative range of imaging parameters, we optimise it with respect to the exchange time by choosing the shells most sensitive to τ_i (see Section 6.2.4.2 for further details). In Figure 6.3, each point on the x-axis represents one unique shell combination and the corresponding y-axis value indicates the R^2 score when the random forest is trained on that particular

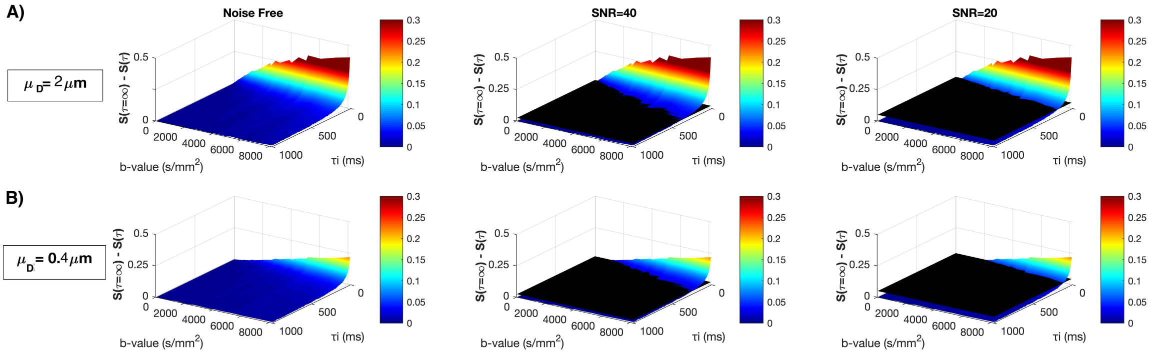

Figure 6.2: Exchange time ranges over which impermeable and permeable substrates can be distinguished for different noise levels and tissue substrates. Figure shows the difference in the DW-MRI normalised signal between impermeable ($\tau_i = \infty$) and the equivalent permeable ($\tau_i \varepsilon [2, 1000]$ ms) substrates. Figure 6.2A) shows results for a substrate with intra-axonal volume fraction f=0.7 and mean axonal diameter $\mu_D = 2$ μm, representing large axons in the mouse brain. Figure 6.2B) illustrates the results for a substrate with f=0.7 and $\mu_D = 0.4$ μm, mimicking small axons in the brain. The level of signal detectability is displayed for three SNR levels (∞, 40 and 20), represented by the black planes, below which any change in signal is undetectable.

shell combination. For example, the x-axis in Figure 6.3A will have 12,650 points, each one corresponding to one of the 12,650 unique 4-shell combinations. As we are interested in the performance of the model with respect to τ_i (1st column), we rearrange the shell combinations in increasing order according to their R^2 for τ_i. This results in a monotonically increasing curve for τ_i, as seen in the first column. For f (2nd column) and d (3rd column), we keep the x-axis ordering consistent with the results for τ_i in the 1st column.

The R^2 scores curves in the 1st column of Figure 6.3 show that only a limited number of shell combinations have a good correlation coefficient and are optimal for estimating τ_i, while the R^2 scores in the 2nd and 3rd column show that the majority of shell combinations provide good estimates of f and d. For example, in the noise-free (blue curves) 4-shell case in the top row, we notice that the difference in R^2 score for τ_i between the best and the worst performing shell combinations is approximately 0.5. In contrast, this difference is much narrower for f and d: 0.2 for f and 0.1 for d. We observe the same trends for SNR=40 (orange) and SNR=20 (green), however, the differences in R^2 scores between the worst and the best performing shell combinations are smaller due to the presence of noise.

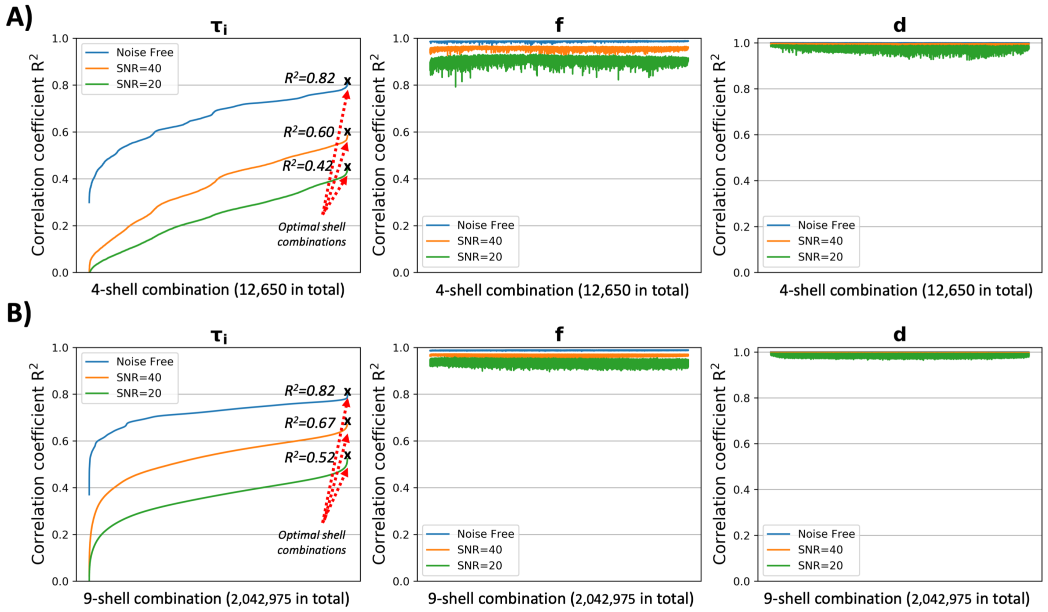

Figure 6.3: Performance of the random forest model trained on different combinations of 4 (A) and 9 (B) shells. In the first column, each curve shows the R^2 score (y-axis) of the random forest trained on a different combination of shells (x-axis). The shell combinations are sorted in increasing order according to their R^2 score. The second and third columns show the R^2 scores for f and d, which are sorted according to the order of the first column. We show the results for three levels of noise: SNR $= \infty$ (blue curve), SNR=40 (orange curve) and SNR=20 (green curve). The R^2 score for τ_i is calculated only for values ≤ 400 ms as this is the range over which we are sensitive to this parameter (see Section 6.3.1).

By comparing the best R^2 scores on the blue curves in Figure 6.3A and Figure 6.3B, we can see that there is no difference in performance in the noise-free scenario between using the best combination of 4 or 9 shells. However, this changes with the addition of noise. For example, for SNR=40 (orange curves) the R^2 score of the best 9-shell combination is 0.67, 0.07 higher than for the best 4 shells. This trend is similar for SNR=20 (green curves), with a difference of 0.1 between 9 and 4 shells. For the 16-shell scenario, we found no improvement in performance over using 9 shells.

Figure 6.3 also shows the effect of noise on the estimation of each parameter. As expected, the addition of noise results in lower R^2 scores, a trend that holds for all parameters and across the 4 and 9-shell case. However, the estimation of τ_i is the most affected by the presence of noise: the maximum correlation coefficient drops from 0.82 in the noise-free case to 0.67 for SNR=40 and even further to 0.52 for SNR=20. For f (2nd column), the effect of noise is considerably smaller: R^2

drops from 0.99 for SNR=∞ to 0.94 for SNR=20. The estimation of the intrinsic diffusivity d is very robust to noise: the correlation coefficients remaining very high (0.99) even when training the model on the SNR=20 dataset. Furthermore, we found that all the top 100 combinations contain the two highest b-value shells (6,803 smm^{-2} and 7,773 smm^{-2}) with the two longest Δs. Additionally, we find that high b-value shells only maximise the performance of the random forest in combination with low b-value shells (775 and 930 smm^{-2}). For SNR=40 (orange curves), which we use when predicting on the in-vivo data, we found that the best combination of 9 shells sorted by b-value is [620, 775, 930, 1241, 1384, 2489, 4973, 6803, 7773] smm^{-2} with an R^2 score of 0.67, and the best combination of 4 shells is [775, 930, 6803, 7773] smm^{-2} with an R^2 score of 0.60. Since we are looking to optimise our framework for in-vivo estimation on the mouse data, we will run the in-vivo experiments using the best 9-shell combination in the SNR=40 scenario, as the noise level which matches our in-vivo data.

6.3.3 Synthetic experiments

Figure 6.4 shows the random forest results obtained using the feature (top row) and the signal (bottom row) noise-free databases. To assess the quality of our fit, we display the results using Bland-Altman plots and colour each data point according to how close the estimates are to the ground truth values. To aid visual interpretation, we cap the percentage error at $\pm 50\%$. The mean difference between the ground truth and the estimated values is shown by the black line and the 95% upper and lower limits of agreement by the dashed lines. For all three parameters of interest, we observe no overall estimation bias as the estimates are spread equally around the zero-difference black line. However, for τ_i, the parameter recovery is not perfect and the Bland-Altman plots show an overestimation bias for small values of τ_i and an underestimation bias for large values. The R^2 scores show a strong correlation between the estimates of our model and the ground truth parameter values: $R^2_{\tau_i} = 0.82/0.84$ (features/signals database), R^2_f=0.99 (both databases), and $R^2_d = 0.99$ (both databases). When assessing the models performance with respect to the two training databases (features/signals), we observe no significant difference between the two approaches. The R^2 scores remain unchanged for f and d

and show only a minor difference for τ_i: $R^2_{features} = 0.82$ / $R^2_{signals} = 0.84$. The advantages of each approach are discussed further in the next section. The noise-free results in Figure 6.4 provide a benchmark performance of the model given our data and imaging protocol.

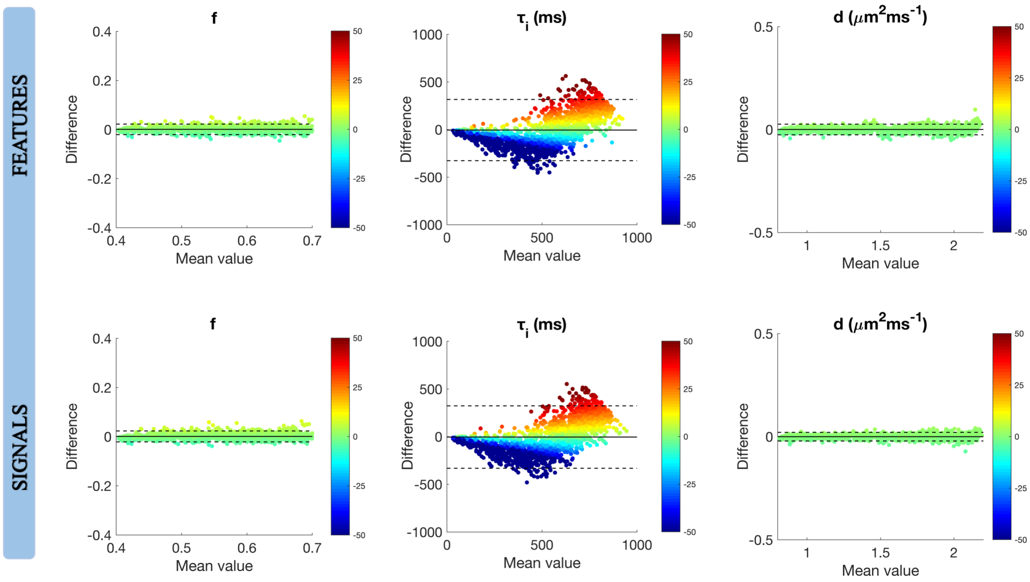

Figure 6.4: Bland-Altman plots for the random forest estimates of f, τ_i and d using the features (top row) and signals (bottom row) noise-free simulated database. To aid visual interpretation, the plots are colour-coded with the percentage error capped at $\pm 50\%$.

Figure 6.5 shows the equivalent results for SNR=40. The presence of noise results in wider limits of agreement and affects differently the prediction of each parameter. The mean difference lines for all three parameters remain at zero, showing no general bias in the estimates. Intra-axonal volume fraction and diffusivity continue to be very well estimated and their correlation coefficients are only very mildly affected by the presence of noise: $R^2_f = 0.97$ and $R^2_d = 0.99$, equal for both training databases. In contrast to this, the presence of noise has a stronger effect on the estimation of τ_i, resulting in a lower R^2 score and a more pronounced overestimation/underestimation bias for small and large values respectively. Despite this, we find that the random forest works well within the sensitivity range computed in Section 6.3.1, with a sufficiently good correlation coefficient between the models estimates and ground truth for $\tau_i \leq 400$ ms (R^2=0.68). Outside this indicative sensitivity range, the correlation coefficient is very weak: R^2= 0.07 for $\tau_i \geq 400$ ms. In line with the noise-free case, we continue to see no significant difference between the signal and the feature approach: $R^2_{features}$ = 0.67 / $R^2_{signals}$ = 0.68.

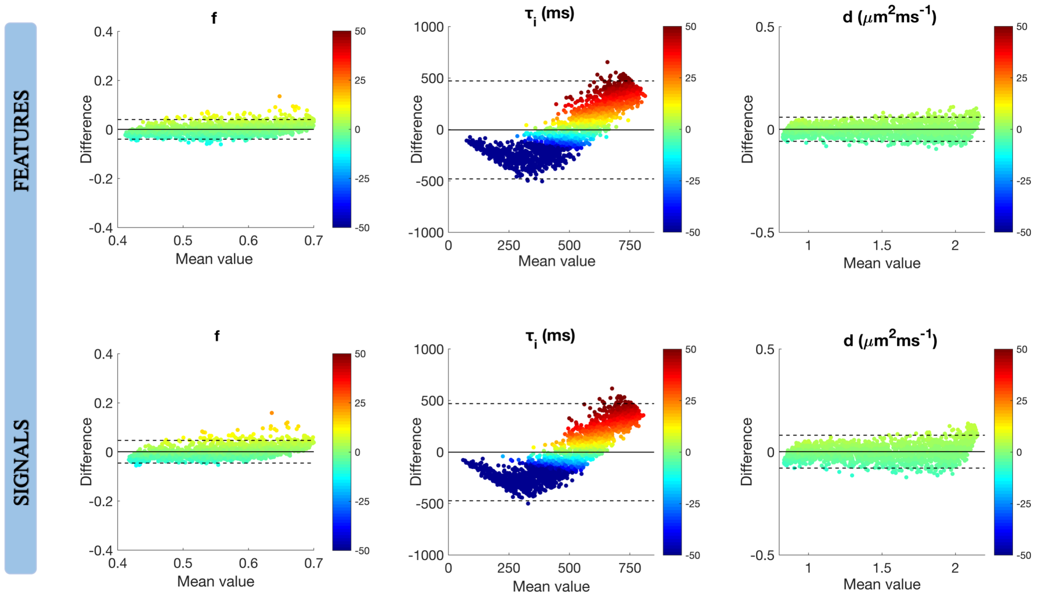

Figure 6.5: Bland-Altman plots for the random forest estimates of f, τ_i and d using the features (top row) and signals (bottom row) simulated database with SNR=40, matching the noise level in our in-vivo data. To aid visual interpretation, the plots are colour-coded with the percentage error capped at ±50%.

6.4 Discussion

Sensitivity analysis. Our sensitivity analysis shows that our imaging protocol has good sensitivity for exchange times in the range $\tau_i \varepsilon [0,400]$ ms for substrates with large axons ($\mu_D = 2$ µm) and in the range $\tau_i \varepsilon [0,250]$ ms for substrates with small axons ($\mu_D = 0.4$ µm), under noise conditions matching that of our in-vivo data (SNR=40). Generally speaking, the noise in the data affects the sensitivity differently, depending on the mean axon diameter in the substrate. For substrates with large axons ($\mu_D = 2$ µm), the sensitivity halves from $\tau_i \varepsilon [0,400]$ ms for SNR=40 to $\tau_i \varepsilon [0,200]$ ms for SNR=20. For substrates with smaller axons ($\mu_D = 0.4$ µm), decreasing the SNR from 40 to 20 has a smaller effect on the sensitivity range, reducing it by 44% from $\tau_i \varepsilon [0,250]$ ms to $\tau_i \varepsilon [0,140]$ ms. Furthermore, we found that the larger the axons in the substrate, the better the sensitivity range. Substrates with $\mu_D = 2$ µm have a sensitivity range wider by 60% for SNR=40 and by 43% for SNR=20 than substrates with $\mu_D = 0.4$ µm.

Shell selection. To optimise the performance of the machine learning model, we explore the wide range of parameters in our SDE protocol and select the best combination of 4 and 9 shells. We show that for our in-vivo data with SNR=40 the number of shells that maximises the performance of the model is 9, with the b-values [620, 775, 930, 1241, 1384, 2489, 4973, 6803, 7773] smm^{-2} and an R^2 score of 0.67. When analysing the best combinations of 4 and 9 shells, we observe that they sample every value of Δ in our sequence, resulting in a combination of low and high b-value shells. This finding is in accordance with the optimised STEAM protocol in Chapter 4, which contains two long Δ and two short Δ shells. Additionally, this result is also in line with the feature importance results from the same chapter, which show that both long and short Δ shells are informative for the parameter estimation. All in all, these results suggest that to maximise sensitivity to the intra-axonal exchange time, it is necessary to include a combination of short and long Δs.

We show that noise is an important factor for the performance of our model. We find that in the noise-free case, it is sufficient to use only 4 shells, as introducing more shells does not improve the model performance. However, in the presence of noise, we find that increasing the number of shells from 4 to 9 improves the R^2 score between the estimated and the ground truth τ_i. A potential explanation for this is that the addition of noise corrupts the information in each shell, and having more shells to corroborate information from helps the random forest model learn better. Our analysis also reveals that increasing the number of shells above 9 does not offer any additional benefits even in the presence of noise. Moreover, we show that noise has a stronger effect on the estimation of τ_i, for which the R^2 score drops from $\approx$0.85 in the noise-free case to $\approx$0.5 for SNR=20. The estimation of f and d is considerably more robust: $R^2_{noise-free}$=0.99 versus $R^2_{SNR=20}$=0.94 for f and no drop for d. This suggests that SNR plays an important role in a protocol's suitability for permeability estimation using our approach.

Synthetic data experiments. The random forest model estimates in the noise-free case have very strong correlations with the ground truth values, providing an excellent benchmark performance for our model and imaging protocol (f: R^2=0.99, τ_i:

R^2=0.84 d: R^2=0.99). We show that the addition of noise with SNR=40, matching our in-vivo data, does not affect much the estimation of f and d (f: R^2=0.97, d: R^2=0.99), however, it has a stronger effect on the estimation of τ_i. In line with our sensitivity results, for $\tau_i < 400$ ms the effect is present, however, the performance is still sufficiently good (R^2=0.68), while for $\tau_i > 400$ ms the performance of the model is severely affected (R^2=0.07).

In addition to this, we compare for the first time the signal and feature training approaches and show that there is no significant difference in the our model's performance according to which database is used for training. This is a significant result as it shows that when extracting the rotationally invariant features from the raw signals we do not lose information that is essential for training our model. Consequently, we can use the feature database without affecting the performance of our model. The advantage of a rotationally invariant feature approach is that it does not require the generation of a new library for every new acquisition protocol as long as the b-values and the TE of the protocols match. Nevertheless, as we will discuss in the next chapter, caution should be applied with this approach when the acquisition protocol uses high gradient strengths ($G \geq 300$ mT/m) and the SNR is low, such as conditions often found in the preclinical setting, and, then, using signals database might be the preferable choice. On the other hand, in the clinical setting, imaging protocols have much lower gradient strengths and sufficient SNR to fit the DTI and SH model parameters in the feature extraction approach, and consequently, we expect the rotationally invariant feature approach to be a better choice (as used in Chapters 4 and 5). Irrespective of the training approach, we expect our model's performance to be similar.

Limitations. In addition to the general limitations of our computational framework as discussed in Chapter 4.4, the specific limitation in this and the following two chapters is the sensitivity of our imaging protocol to τ_i in the presence of noise, which is not ideal, as shown in our simulation experiments. Although we performed some level of optimisation by choosing the most optimal shells in our large explorative protocol, the sensitivity might have been better if we optimised the protocol with respect to τ_i prior to imaging, as for the STEAM protocol in Chapters 4. Nevertheless, even with the protocol we use we can estimate values of $\tau_i \leq 400$ ms, which we expect to be sufficient for the in-vivo mouse application we used in this study. The machine learning model here can also easily be adapted to incorporate more specialised diffusion encoding sequences such as OGSE for more sensitivity to axon diameter [138] or STEAM for longer diffusion times [166]. An advantage of our computational framework is that it can easily incorporate changes in the imaging protocol.

6.5 Conclusions

In conclusion, this study looks at the sensitivity to the exchange time of the explorative preclinical SDE protocol used to acquire the in-vivo preclinical mouse data in order to ensure that there is sufficient information in the data. Using simulations closely matching our in-vivo data, we show that the imaging protocol has good sensitivity to $\tau_i \leq 400$ ms, which we expect to be sufficient for our mouse model, based on reported literature values [10, 9, 11]. Then, we optimise the performance of our random forest model by choosing the most informative shells with respect to τ_i and show that it is necessary to include a combination of short and long diffusion times to maximise the performance of our model, consistent with the optimised STEAM protocol in Chapters 4 and 5. Finally, we establish a benchmark performance for our model by testing it on noise-free simulations, where we find strong correlations between the estimated and ground truth parameters (f: $R^2 = 0.99$, τ_i: $R^2 = 0.84$, d: $R^2 = 0.99$). We find that, for levels of noise matching that of the in-vivo mouse data (SNR=40), the simulation performance is affected, however, the parameters are still well estimated within the sensitivity range (f: $R^2 = 0.97$, τ_i: $R^2 = 0.68$, d: $R^2 = 0.99$). In the next chapter, we apply our random forest model to a cohort of 16 mice, 8 healthy controls and 8 cuprizone treated mice with induced demyelination. Additionally, we seek further validation of our computational framework by directly comparing the estimates of our model with available histological data from the same mice.

Chapter 7

Random Forest based computational models with permeability: microstructure parameter estimation in preclinical mouse model of demyelination

In the previous chapter we have studied the sensitivity of the SDE protocol used for the in-vivo mouse data in this chapter and optimised the performance of the random forest model using simulations closely matching our in-vivo data.

In this chapter we investigate the in-vivo performance of the random forest model optimised in Chapter 6, using a cuprizone mouse model of demyelination with available histological data.

Some of the experiments in this chapter are performed with help from researchers at the *Institut du Cerveau et de la Moelle epiniere, ICM, Paris, France*, who carried out the scanning of the mice and the histological analysis. Parts of this work are featured in the recently submitted paper to *NeuroImage*: *Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination. I. Hill et al. [163].*

7.1 Motivation

The random forest approach for permeability estimation, covered in Chapter 4 and introduced for the first time in [132], is tested in-vivo only qualitatively and using just two healthy and two MS patients. As this is a novel approach to microstructure parameter estimation, more in-vivo validation is needed in order to investigate the clinical potential of the random forest model estimates, in particular τ_i, to act as a suitable biomarker for detecting and tracking changes in MS pathology. In-vivo validation of the exchange time is, however, extremely challenging, and accurate measurements of τ_i are not obtainable through histology due to tissue fixation altering the membrane permeability. Moreover, MS pathology is known to simultaneously cause other changes in the underlying tissue, such as axonal loss or axonal swelling, making it more difficult to measure the effect of τ_i in isolation. Another challenging aspect of the in-vivo estimation of the exchange time is the presence of dispersion, which has been shown to cause an underestimation of τ_i [1].

In this chapter, we address several of the limitations outlined above. For this, we use a highly controlled mouse cuprizone model of demyelination, commonly used as an animal model of MS [162]. Our in-vivo data consists of a cohort of 16 mice, 8 healthy wild-type (WT) mice and 8 cuprizone treated mice (CPZ) with available histological data. Using electron microscopy (EM) measurements of the mean axon diameter form the same mice, we investigate the potential confounding effects of axonal swelling and axonal loss. Furthermore, we look at the potential effect of dispersion by computing NODDI ODI and DTI maps. As accurate histological measurements of τ_i are not available, we analyse the correlation between the random forest estimates of τ_i and EM measurements of myelin thickness, which has recently been shown in a simulation study to correlate strongly with the exchange time [86]. We additionally compare the random forest estimates with EM measurements of the intra-axonal exchange time τ_i.

The aim of this chapter is to provide further validation of the random forest based computational model introduced in Chapter 4 by comparing the in-vivo random forest estimates from DW-MRI data with available histological data. An additional aim of this chapter is to investigate the potential confounding effects of dispersion and axonal swelling on the estimation of τ_i.

7.2 Methods

This section outlines the in-vivo rat data acquisition, as well as the sample preparation procedure for the electron microscopy (EM) analysis. In the experiments subsection, we introduce the in-vivo DTI, NODDI and random forest experiments, and we outline the procedure for calculating the histological measurements of myelin thickness, g_{ratio}, axon diameter and intra-axonal volume fraction.

7.2.1 Mouse data

7.2.1.1 In-vivo data acquisition

For the in-vivo data, sixteen 8-week old C57BL/6J female mice were imaged. All animal experiments were performed in accordance with the European Council Directive (88/609/EEC). Eight mice were fed 0.2% cuprizone for 6 weeks, which corresponds to a demyelination without recovery phase, and eight healthy age-matched wild-type (WT) mice of the same background were fed a normal chow diet and used as controls. All mice are scanned on a BrukerBioSpec 11.7T scanner (the data used in this study are available in the public domain and can be found at https://zenodo.org/record/996889#.WgH5E9vMx24 [167]) using the protocol described in Section 6.2.1 Table 6.1.

We post-processed the images by correcting for eddy currents using FSL-eddy [151]. No motion artefacts were observed. We restrict our analysis to white matter voxels within the CC as they match the assumption of non-abutting parallel cylinders in our learnt white matter model. To select the CC voxels, we compute maps of linearity (CL), planarity (CP) and sphericity (CS) [168] from the DTI fit to the lowest b-value shell with SNR 40. We created the CC maps by selecting the voxels with CL>0.3, CP<0.4, CS<0.5 and fractional anisotropy (FA)>0.3.

7.2.1.2 Histology samples

The WT (n=8) and CPZ (n=8) animals are sacrificed by deep anaesthesia and perfused intracardially with 1% paraformaldehyde and 2.5% glutaraldehyde in phosphate buffer 0.12 M, pH 7.4 at the end of the 6-week CPZ treatment. The extracted brains are then post-fixed overnight at 4°C in the same fixative and rinsed in phosphate buffer. Ten 100 µm-thick sagittal sections are cut with a vibratome (Thermo Scientific Microm HM 650 V Vibration microtome). The very first section closest to the brain midline is considered as #1 and sections #1, #4, #7, and #10 are selected. Sections are post-fixed with 1% osmium tetroxide in water for 1 hour at room temperature (RT°), rinsed 3x5 min with water and contrasted en bloc for 1 hour at RT° with 2% aqueous uranyl acetate. After rinsing, sections are progressively dehydrated with 50%, 70%, 90%, and 100% ethanol solutions for 2x5 min each. Final dehydration is achieved by immersing the sections twice in 100% acetone for 10 min. Embedding is performed in epoxy resin (Embed 812, EMS, Euromedex, France) overnight in 50% resin / 50% acetone at 4°C followed by 2x2 h in pure resin at RT°, and polymerization is achieved at 56°C for 48 h in a dry oven. Semi-thin sections (0.5 µm-thick) are collected with an ultramicrotome UC7 (Leica, Leica Microsystmes SAS, France) and stained with 1% toluidine blue in 1% borax buffer. Ultra-thin sections (70 nm-thick) are contrasted with Reynolds lead citrate (Reynold ES, 1963), and observed with a transmission electron microscope (HITACHI 120 kV HT 7700), operating at 70 kV. Images (2048x2048 pixels) are acquired with an AMT41B camera (pixel size: 7.4 µm x 7.4 µm).

7.2.2 Experiments

7.2.2.1 In-vivo imaging experiments

Before generating in-vivo parameter maps using our trained machine learning model, we first perform a data quality match to check that the in-vivo data is well represented by our synthetic training database. In addition to this, we investigate any potential bias in our in-vivo estimates of τ_i due to changes in the orientation dispersion by computing maps of the NODDI orientation dispersion index (ODI) [78] using the NODDI Matlab (The MathWorks, Inc, Natick, MA) Toolbox. Using the Camino toolbox, we additionally generate DTI maps at b=1500 s/mm^2 of axial diffusivity (AD), fractional anisotropy (FA) and radial diffusivity (RD) as measures of tissue properties that can be compared with already published works in cuprizone model [169, 170, 171].

Using the random forest trained on the noisy database, we generate parameter maps for the CCs of the 16 mice for three parameters of interest: τ_i, f and d. To investigate the differences between the two groups (CPZ and WT), we compute box-and-whisker plots of region-specific comparisons between WT (8 mice) and CPZ (8 mice) for the DTI and NODDI metrics as well as for the random forest estimates. Statistical significance is assessed by a two-tailed t-test, considering p-values< 0.05. We run these experiments using the signals database. The Camino feature extraction of the in-vivo data did not produce histologically plausible results for the shells with very high gradient strengths ($G > 300$ mT/m) in our protocol, and we therefore exclude this approach from the analysis in this section and discuss the potential explanations for this in the Discussion section of this chapter.

7.2.2.2 Correlation with post-mortem analysis

From the electron microscopy (EM) samples obtained as described Section 7.2.1.2, we estimate the mean and standard deviation of the g_{ratio}, myelin thickness, axonal diameter and the intra-axonal volume fraction of the WT and CPZ mice. The stereological analysis is performed in isolated regions of the CC (genu, body and splenium), where 4 random sections with uniform distance are quantified per animal (Figure 7.1B), with 30 randomly located images per region and per animal acquired at 62000X magnification. For volume fraction (VF) we proceed according to the Delesse principle [172]: volume fractions are calculated by dividing the total number of points hitting the structure ($P(Y)$) by the total number of points hitting the reference volume ($P(ref)$), following the equation:

$$VF(Y, ref) = \frac{\sum_{i=1}^{m} P(Y)_i}{\sum_{i=1}^{m} P(ref)_i}$$

A grid of 36 regularly spaced crosses (Figure 7.1E) is generated with Fiji, an open-source platform for biological image analysis [173]. Each of the 36 crosses represents an area of 0.5 μm^2. Axonal fibres are judged to be in transverse section by the orientation of axonal microtubules. Stereological analysis provides Myelin Volume Fractions (MVF), Axon Volume Fractions (AVF), and the total Axon Volume Fractions ($tAVF$), which includes both myelinated and unmyelinated axons. Total Axon Count (*tAxCount*) is manually quantified. The g_{ratio} of myelinated fibres is then calculated according to the following equation:

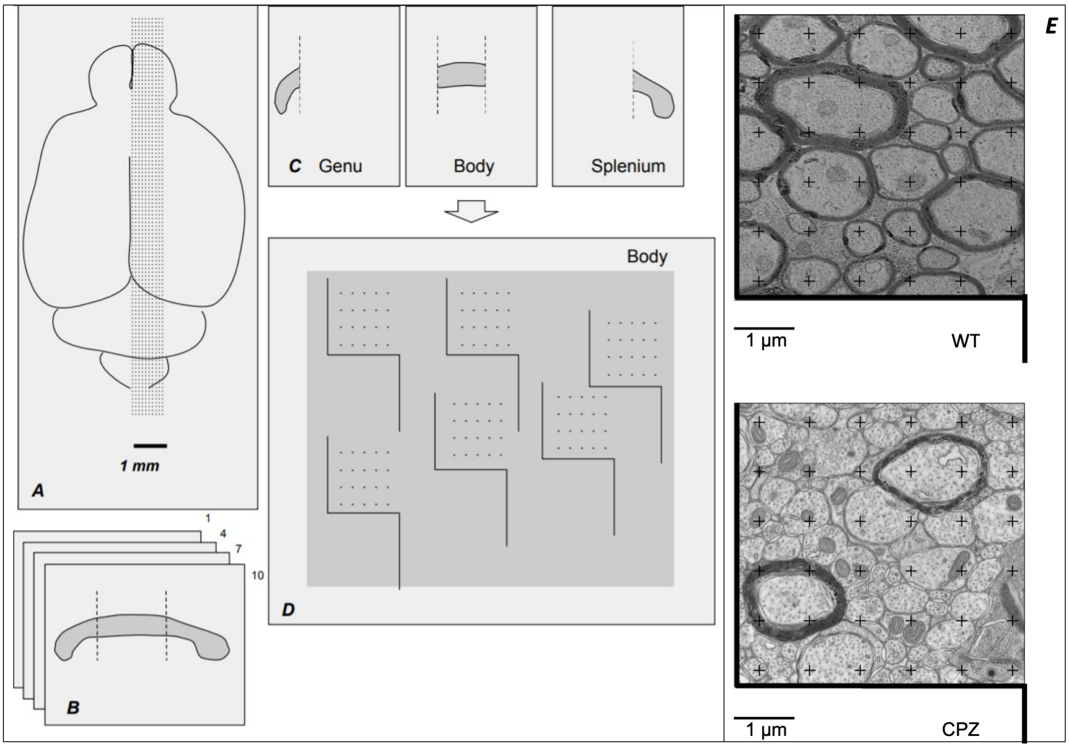

Figure 7.1: Schematic pipeline of the stereological analysis to compute g_{ratios} and axonal diameters in the CC of the mice. First, ten sagittal sections are cut from medial to lateral lines **(A)** and four sections are selected for analysis **(B)**. The genu, body and splenium are then extracted **(C)**, and for each of the three ROIs 30 random regions are selected for stereological analysis using point grids **(D)**. The point grids are then overlaid onto a tissue sample as illustrated in Figure **E**, where each of the 36 crosses represents an area of 0.5 µm^2. The point grids in *E* are used for quantification of the WT and CPZ mice.

$$g_{ratio} = \sqrt{\frac{AVF}{MVF + AVF}}$$

The mean axon diameters (DAX) are calculated as follows:

$$DAX = 2 * \sqrt{\frac{tAVF * \text{surface}}{\pi * tAxCount}}$$

The surface is the area of each cross (0.5 µm^2). The outliers induced by the non-perpendicular axons in the images are not taken into consideration. From the g_{ratio} and the DAX, myelin thickness is computed as:

$$\text{myelin thickness} = \frac{\text{DAX}}{2g_{ratio}}(1 - g_{ratio})$$

We compare the estimates of the random forest with the EM measurements by computing the group-wise mean in the CC ROIs of the myelin thickness and intra-axonal volume fraction (VF) and looking at the correlation between these and the random forest estimations for τ_i and f.

7.3 Results

7.3.1 In-vivo imaging experiments

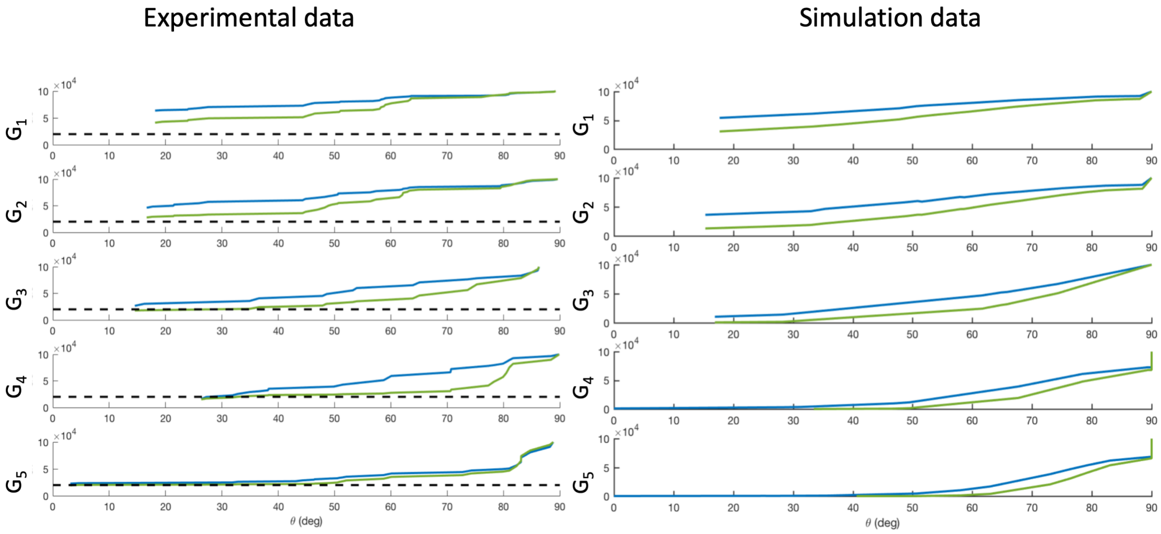

Figure 7.2: Comparison between the in-vivo *(left)* and simulated *(right)* signal intensity as a function of the angle between diffusion gradients and the cylindrical fibres axis θ (in degrees), for different diffusion gradient strengths ($G_{1-5} = 150-500$ mT/m) and two Δs: 10.8 ms *(blue lines)* and 20.0 ms *(green lines)*. The dashed black line in the experimental data represents the noise level.

To show that our in-vivo data is well represented by our synthetic training database, we perform a data quality match (Figure 7.2). We plot the signal intensity as a function of the angle between the diffusion gradients and the cylindrical fibres axis θ (in degrees), for different diffusion gradient strengths ($G_{1-5} = 150-500$ mT/m) and for $\Delta = \{10.8, 20.0\}$ ms. The experimental data in Figure 7.2 represents a randomly chosen voxel in the centre of the splenium of a WT mice, and the simulated data represents the signal in our synthetic database that minimises the residuals between the simulated and the in-vivo signal. We find a very good match between the simulated and in-vivo DW-MRI signals, demonstrating that our training data set is a good representation of the in-vivo mouse data.

Figure 7.3 shows examples of DW-MRI b=0 images for a WT (Fig. 7.3A) and for a CPZ (Fig. 7.3B) mouse. We can observe the appearance of the CC in the CPZ scan is different from the WT, showing the effect of demyelination.

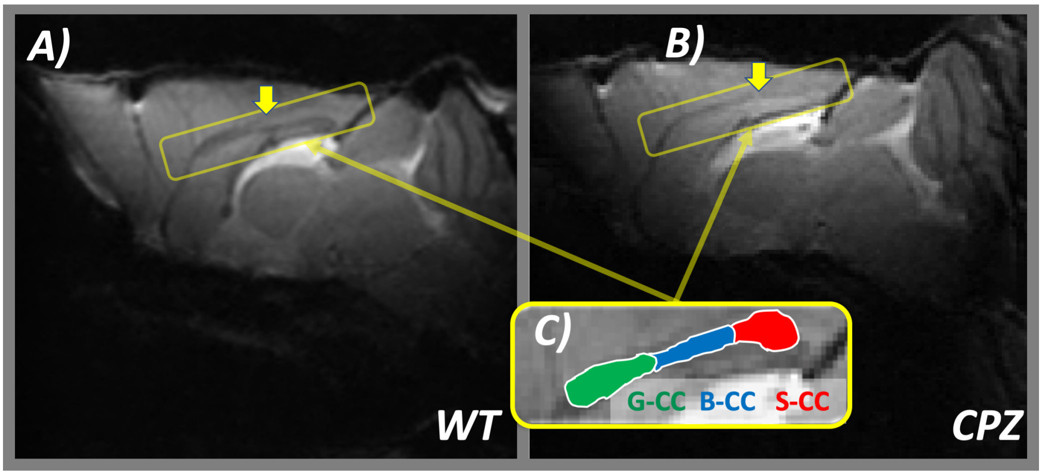

Figure 7.3: Representative DW-MRI b=0 images of: **A)** a WT mouse scan in our cohort and **B)** a CPZ mouse scan in our cohort. **C)** ROIs of the CC overlaid on the zoomed in b=0 image of the WT mouse scan. The three ROIs are genu (G-CC), body (B-CC) and splenium (S-CC). The yellow square indicates the region in which the CC is found.

Figure 7.3C) shows the three ROIs of the CC overlaid on the b=0 image of the WT scan. We manually define three ROIs on the CC masks of each mouse scan: splenium (S-CC), body (B-CC) and genu (G-CC). We then calculate the mean parameter estimates for NODDI (ODI), DTI (AD, RD, FA) and random forest (f, τ_i, d) in each ROI for every mouse, and study the differences between the WT and the CPZ groups. We present these results in the remaining of this section.

Figure 7.4A) shows CC maps for NODDI and DTI parameters for one exemplar healthy WT mouse (first column) and one exemplar CPZ mouse (second column). A visual inspection of the CC maps reveals no significant changes in ODI and AD, together with a significant increase in RD and decrease in FA.

We observe the same trends in the DTI and NODDI parameters at group level, as shown in Figure 7.4B). We illustrate the difference between the WT group $(N_{mice} = 8)$ and the CPZ group $(N_{mice} = 8)$ through box and whisker plots in the three ROIs of the CC: G-CC, B-CC and S-CC. We find the estimates of ODI in the two groups to be between 0.15 and 0.29, suggesting very low dispersion, in line with recently reported literature values [174]. Moreover, we find no statistically significant difference in NODDI ODI between the two groups in the three regions of

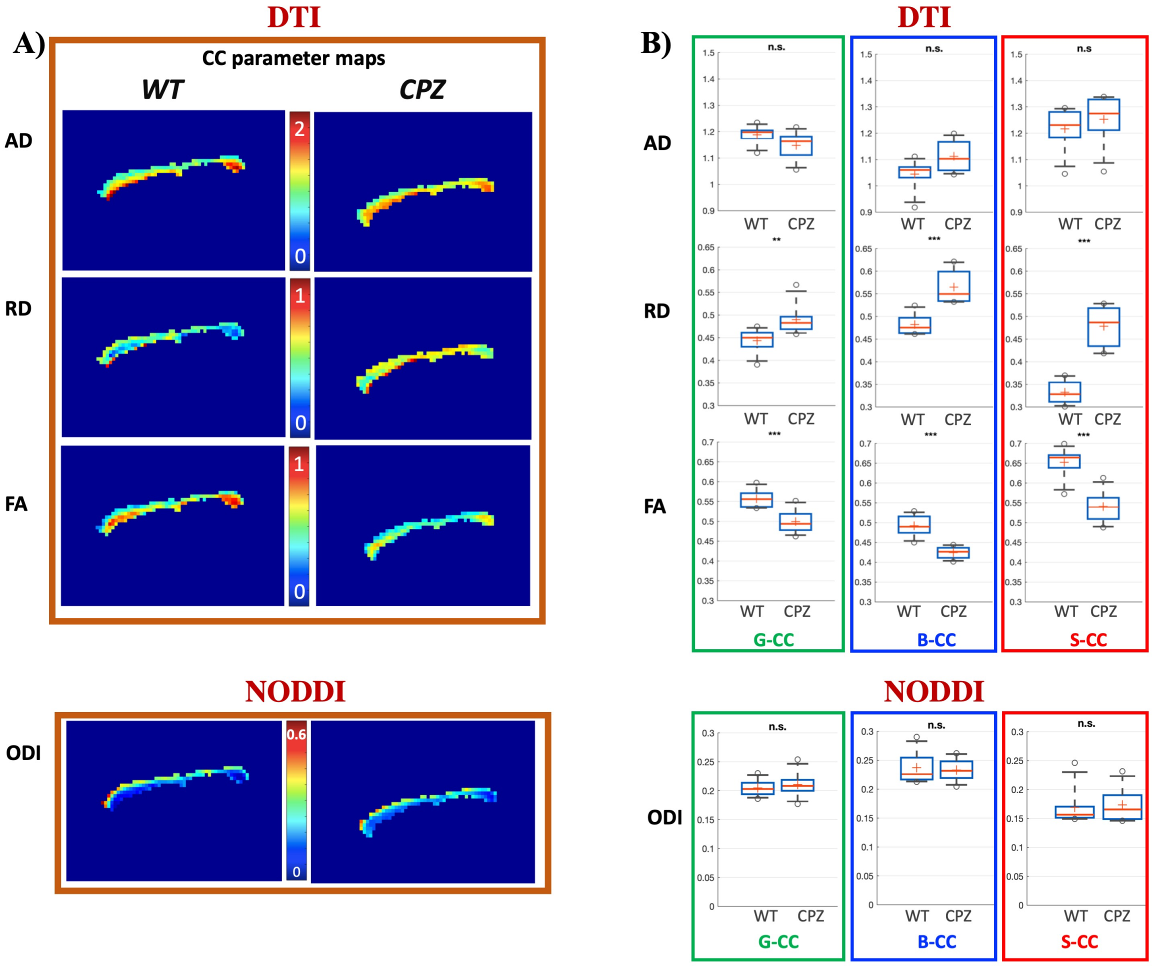

Figure 7.4: **A)** Parametric maps of the CC in a healthy WT mouse *(first column)* and a CPZ mouse *(second column)* obtained from conventional DTI at b=1500 s/mm^2 and from NODDI ODI. **B)** Box and whisker plots of region-specific comparison between WT (N_{mice}=8) and CPZ ($N_{mice} = 8$). DTI metrics (AD, RD, FA) are evaluated within the genu (G-CC), body (B-CC) and splenium (S-CC) of the CC. Statistical significance is assessed by using a 2-tailed t-test with equal variance and significance level: $^* = 0.01, ^{**} = 0.005, ^{***} = 0.001$. '*n.s.*' stands for *non-significant*.

the CC, a finding that is also in line with [174]. The DTI estimates show negligible changes in AD, a significant increase in RD and a significant decrease in FA.

The in-vivo random forest estimates of f, τ_i and d obtained using the raw signal database are presented in Figure 7.5. The parametric CC maps shown in Figure 7.5A) correspond to the same WT mouse (first column) and CPZ mouse (second column) in Figure 7.4A). The CC maps show a statistically significant decrease in f (first row) and τ_i (second row), and no significant change in d (third row). The values estimated for all parameters are within plausible ranges for rat white matter

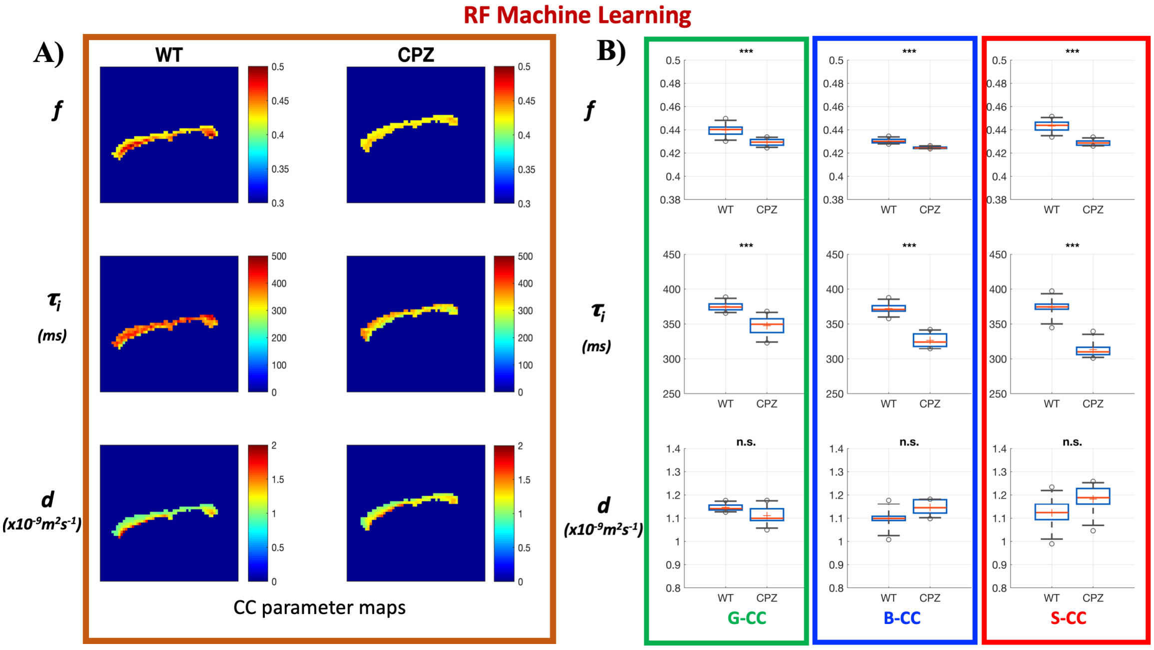

Figure 7.5: **A)** Parametric maps with the random forest estimates for f, τ_i and d in the CC of a healthy WT mouse (first column) and a CPZ mouse (second column). **B)** Box and whisker plots of region-specific comparison between WT ($N_{mice} = 8$) and CPZ ($N_{mice} = 8$). Random forest estimates for f, τ_i and d are computed independently for all voxels within the genu (G-CC), body (B-CC) and splenium (S-CC) of the CC. Statistical significance was assessed by using a 2-tailed t-test with equal variance and significance level: $^* = 0.01, ^{**} = 0.005, ^{***} = 0.001$. ’*n.s*’ stands for *non-significant*

.

tissue. For the control group, estimates of volume fraction f are in the range 0.42-0.46, estimates for τ_i are consistently in the range 330-400 ms, and the estimates for d lie in between 1-1.3 $x10^{-9}\text{m}^2\text{s}^{-1}$. To provide a more quantitative analysis, we plot the box and whisker plots of region-specific parameter comparisons between the WT and the CPZ group over the three CC ROIs (Figure 7.5B). The trends observed visually in Figure 7.5A) hold for the group-wise quantitative comparison (WT versus CPZ): we observe statistically significant decreases in f and τ_i and a negligible increase in d. These trends are consistent across all three regions of the CC. The mean and standard deviations of the random forest parameter estimates for each ROI are reported in Table 7.1.

7.3.2 Correlation with post-mortem analysis

The histological EM measurements in the splenium, body and genu of the CC over the cohort of WT (blue) and CPZ (black) mice are reported in the bar plots of Fig-

	f		**τ_i (ms)**		**d** $x10^{-9}(m^2s^{-1})$	
	WT	**CPZ**	**WT**	**CPZ**	**WT**	**CPZ**
G-CC	0.440 (0.006)	0.429(0.003)***	375 (7)	347 (15) ***	1.15 (0.017)	1.11 (0.042)
B-CC	0.431 (0.002)	0.425(0.0041)***	372 (9)	327 (10) ***	1.10 (0.046)	1.14 (0.032)
S-CC	0.444 (0.006)	0.429 (0.003)***	374 (14)	313 (12) ***	1.12 (0.071)	1.18 (0.065)

Table 7.1: Mean and standard deviation of the random forest estimates for f, τ_i and d in the three CC ROIs for the WT and CPZ group. CPZ regions that are statistically different from WT regions are marked with * for $p < 0.01$, ** for $p < 0.005$ and *** for $p < 0.001$

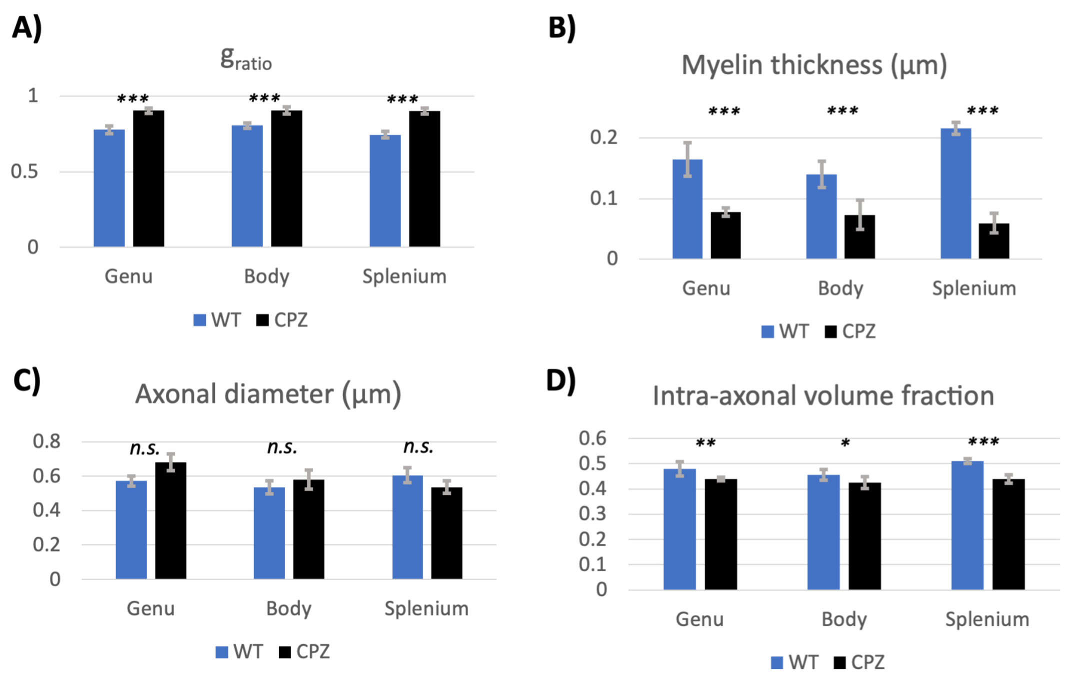

Figure 7.6: Histology results. The mean and the standard deviation of the EM measurements in the splenium, body and genu of the CC for the cohort of WT *(blue)* and CPZ *(black)* mice: the g_{ratio} **(A)**, myelin thickness **(B)**, mean axonal diameter **(C)** and intra-axonal volume fraction **(D)**.

ure 7.6. Our histological data shows no axonal size changes (Figure 7.6C) and no significant axonal loss (data not shown here) between the two cohorts. The axonal diameter measurements in Figure 7.6C) do not take into account the commonly accepted shrinkage factor of 30% [71, 175], after which the differences between the two groups continue to remain statistically non-significant. We also find a statistically significant decrease in myelin thickness (Figure 7.6B) correlated with an increase in the g_{ratio} (Figure 7.6A) and a decrease in the intra-axonal volume fraction

(Figure 7.6D).

Next, we study the correlation between these changes and the estimates of the random forest model and we show the results in Figure 7.7. We assess the statistical significance of the linear correlation between τ_i and myelin thickness from EM with a two-tailed t-test by looking at the mean and the standard deviation of each CC ROI of the WT *(blue squares)* and CPZ *(black circles)* group (Figure 7.7A). We find a pearson linear correlation coefficient ρ of 0.82 and a p-value < 0.05 for τ_i, showing a good correlation between the random forest estimates of the exchange time from DW-MRI *(y-axis)* and histological measurements of myelin thickness *(x-axis)*. As the p-value is less than the significance level of 0.05, it indicates the acceptance of the hypothesis that there is a statistically significant correlation between the EM and random forest estimates.

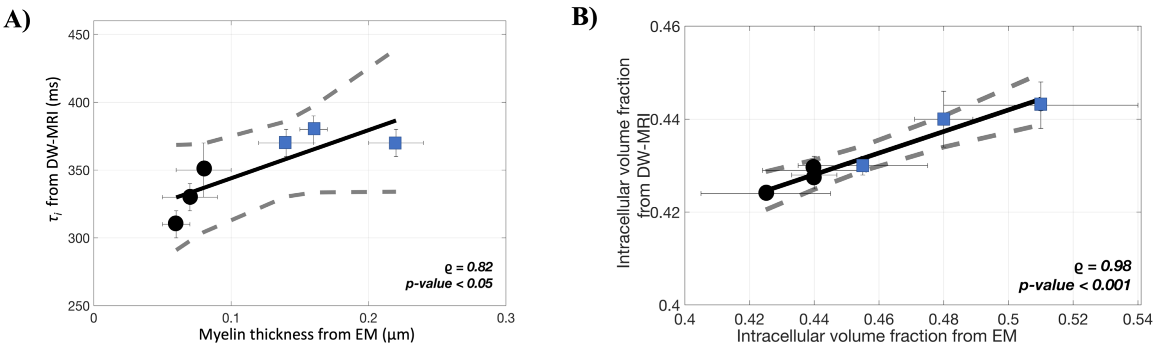

Figure 7.7: Linear correlations between: **A)** the intra-axonal exchange time from DW-MRI *(y-axis)* the myelin thickness from EM *(x-axis)* and **B)** the intra-axonal volume fraction from DW-MRI *(y-axis)* and EM *(x-axis)*. Each point represents the mean over one region of the CC for the WT *(blue squares)* and CPZ *(black circles)* group. Error bars indicate the standard deviation over the region.

Similarly, we investigate the statistical significance of the linear correlation between intra-axonal volume fraction f as estimated from DW-MRI *(y-axis)* and EM measurements *(x-axis)* (Figure 7.7B). We find a Pearson correlation coefficient ρ of 0.98 and a p-value< 0.001, showing a strong significant correlation between the random forest estimates and the histological measurements of the intra-axonal volume fraction.

7.4 Discussion

Feature extraction. When extracting the rotationally invariant features from our synthetic signals, we obtain meaningful values for all shells in the synthetic data. When we apply the same method to in-vivo data, the feature extraction becomes difficult and does not give meaningful results for shells with high gradient strength (above 300 mT/m) and high b-values. We believe that this difference is most likely due to the effect of fibre dispersion, present in the in-vivo data but not included in our simulations. As the gradient strength increases, the dispersed fibres would cause larger drops in the signal, as can also be seen in (Figure 7.2), where we notice that the drop in the signal intensity relative to the gradient direction is less prominent in the synthetic signals than in the in-vivo data.

In-vivo experiments. Our data quality match shows that our synthetic training data is a good representation of the in-vivo data. Our DTI results show an increase in RD and a decrease in FA between the two groups. This could be explained by the breakdown of the myelin layer which allows water to diffuse more in the radial direction, leaving AD unchanged and having the overall effect of reducing FA. These changes in DTI metrics are in agreement with those reported in several literature studies of the CPZ mouse model of demyelination [169, 170, 171]. Nevertheless, the DTI metrics provide only indirect measures of the underlying microstructural changes in the CPZ model.

Our random forest estimates of τ_i, on the other hand, provide a more direct and specific measure of permeability. We find that our estimations of τ_i in the healthy mice compare well with literature values. Our random forest estimates of τ_i in the control group are consistently in the range 350-400 ms, lower than in the human data in Chapter 4, where most τ_i estimates lie in the range 400-500 ms. This could be explained by the size of the axons in the human versus the mouse white matter. As the axon diameter range in the human white matter is larger than that of rodents [176], this can intuitively result in longer exchange times as the water molecules have to travel longer before leaving the intra-axonal space. Another potential explanation is the lower sensitivity of the SDE protocol in this chapter compared to the STEAM protocol in Chapter 4, which may lead to lower estimates of the intra-axonal exchange in the SDE estimates. Studies on sphingomyelin membranes found in axonal membranes suggest values between 300 ms and 600 ms for axons with radii between 0.5 and 1 µm [9]. Contrast agent and relaxometry studies in the rat brain estimate the intracellular water exchange lifetime in the rat brain to be between 200 ms [10] and 550 ms [11]. As accurate histology measurements of τ_i are not available, we compare our estimates of τ_i with EM measurements of myelin thickness. We compute myelin thickness from myelinated axons only, and it includes both the effect of demyelination induced by CPZ and some remyelination that happens spontaneously in the CPZ model [177]. We find a strong correlation between the random forest estimates of τ_i and myelin thickness ($\rho_{\tau_i} = 0.82$). This is in very good agreement with a recently published simulation work investigating the link between exchange time and myelin thickness [86]. For d, our random forest estimates lie in the range 1-1.3 x10^{-9}m^{2}s^{-1}, an expected range for the mouse CC [165], and our estimates of f correlate very strongly with the EM intra-axonal volume fraction measurements ($\rho_f = 0.98$).

When comparing the two groups, we observe the following general trends: a statistically significant decrease in the intra-axonal volume fraction f and in the intra-axonal exchange time τ_i, together with a negligible increase in the intrinsic diffusivity d. We expect f to be lower in the CPZ group as there is an increase in the extracellular space due to the breakdown of myelin. Demyelination is also thought to cause a decrease in the intra-axonal exchange time as the water molecules encounter less barriers when moving from the intracellular to the extracellular space. In line with this, the random forest estimations of τ_i in the CPZ group are significantly lower than in the WT group.

While estimating myelin thickness would allow a direct comparison with histology, it would only apply to pathologies and stages of disease characterised by a change in the thickness of the myelin sheath. Estimating τ_i lends itself to a wider range of diseases in which changes in permeability are driven by other factors than myelin thickness. For example, τ_i in the NAWM of MS patients may change despite the normal appearance of the tissue and the myelin sheath. By measuring τ_i in NAWM, we can potentially gain an earlier window into the disease, prior to the onset of myelin breakdown. Furthermore, in other diseases such as cancer, shown to affect permeability, the cells may be not be covered in a myelin sheath if the tumour is not in the brain.

The strong correlation between myelin thickness and the estimated τ_i suggests that demyelination could be one of the main factors behind our measured decrease in τ_i. To strengthen this hypothesis, we analyse the potential confounding effect of other underlying processes. Our AD measurements from the DTI fit suggest that, if undulation or beading effects are present, they have a negligible effect [178, 159, 179]. We additionally rule out the effect of dispersion by measuring NODDI ODI and showing that the differences between the two groups are not statistically significant. Previous work investigating the impact of dispersion on axonal permeability estimation shows that the presence of orientation dispersion would probably result in an underestimated values of τ_i [1]. Here, we estimate dispersion in the CC of our in-vivo data using NODDI ODI, and our results (Figure 7.4) show that there is no significant change in this parameter between the WT and CPZ groups. This, together with the non-significant change in AD from DTI, suggests that although dispersion may be present in the CC, there is no significant change between the WT and CPZ groups. If dispersion would change significantly between the two groups, we would expect a significant change in ODI and AD. Based on this argument, we are confident that the significant changes in τ_i that we measure (Figure 7.5) are not driven by dispersion. We also rule out the potential confounding effect of axonal swelling by looking at the statistically non-significant changes in axonal diameter as measured by EM. This, together with the measured changes in RD, FA and the random forest estimations of τ_i suggest that demyelination is the main process underpinning our DW-MRI contrast. In particular, our histological data strongly supports τ_i as a biomarker directly related to the thickness of the myelin sheath, which suffers degeneration in demyelinating diseases such as MS.

7.5 Conclusions

In this chapter, we apply our random forest model on in-vivo data consisting of a cohort of 8 CPZ mice and 8 healthy age-matched WT mice. In the control group, we find that the model estimates sensible microstructure parameters matching values found in literature for healthy rat brain white matter. In the demyelinated group, the random forest estimates a statistically significant decrease in τ_i in all three regions of the CC (splenium/body/genu) of the CPZ group ($\mu_{\tau_i} = 313\text{ms}/327\text{ms}/347\text{ms}$) compared to the WT group ($\mu_{\tau_i} = 374\text{ms}/372\text{ms}/375\text{ms}$), in line with our expectations that τ_i is lower in regions where the myelin sheath is damaged, as axonal membranes become more permeable. Furthermore, we use NODDI ODI and EM measurements of axon diameter to rule out the confounding effect of dispersion and axonal swelling on the estimated difference in τ_i between the two groups. When comparing the model estimates with histology data from the same mice, we find a strong correlation between the in-vivo random forest estimates of τ_i and the EM measurements of myelin thickness ($\rho_{\tau_i} = 0.82$), in line with a recently published study [86], and between the random forest estimates and EM measurements of intra-axonal volume fraction ($\rho_f = 0.98$).

Overall, these results, together with the findings in Chapter 6, demonstrate, in simulations and in vivo, the feasibility of random forest based compartment models with permeability as a potential biomarker for demyelinating pathologies such as MS.

Chapter 8

Neural network based computational models with permeability: microstructure parameter estimation in preclinical mouse model of demyelination

In the previous two chapters we presented an experimental study investigating the performance of a random forest based compartment model with permeability through direct comparison with histology from a mouse model of demyelination.

This chapter presents an experimental study of the neural network model introduced in Chapter 5 on the same preclinical mouse model. Parts of this work were selected for an oral presentation at the high-profile Combined Educational-Scientific Session 'Machine Learning for Magnetic Resonance in Medicine' at the International Society for Magnetic Resonance in Medicine (ISMRM) Annual Meeting 2018 in Paris, France. The abstract was one of the five abstracts selected to illustrate the most novel applications of machine learning to MRI data.

8.1 Motivation

The neural network approach for microstructure parameter estimation introduced in Chapter 5 shows promising results, with in-vivo estimates of f, τ_i and d within plausible ranges for human white matter and trends in line with the expected MS pathology. However, as this is a new approach, more quantitative validation of this technique is necessary to assess the suitability of neural network based models for microstructure imaging.

In this chapter, we aim to experimentally investigate the feasibility of neural network based computational models with permeability for microstructure imaging by comparing the in-vivo estimates with histological data from a cuprizone mouse model of demyelination. In line with the previous chapter, we compare our estimates of τ_i with EM measurements of myelin thickness, which has been shown to correlate strongly with the intra-axonal exchange time [86]. This is intuitive as the myelin layer surrounding the axons acts as a barrier to the water molecules leaving the intra-axonal space, hence its thickness is likely to influence directly the average time it takes the water molecules to leave the intra-axonal space, as measured by τ_i. For the intra-axonal volume fraction f, we compare the neural network estimates with direct electronmicroscopy (EM) measurements of f from the same mice.

8.2 Methods

8.2.1 Imaging protocol and data

For both the synthetic and the in-vivo data acquisition, we use the SDE imaging protocol (see Table 6.1) and databases from Chapter 6. Our synthetic data is formed of 12,500 unique Monte Carlo simulations of the DW-MRI signal and their associated ground truth microstructure parameters. The synthetic substrates used to generate the DW-MRI signals mimic the in-vivo mouse data. More details about the dataset are provided in Section 6.2.2. We generate two synthetic databases, one for the noise-free scenario and another by adding Rician noise with a standard deviation σ corresponding to an SNR of 40, matching the level of noise present in our in-vivo data. Our in-vivo dataset is formed of DW-MRI scans of sixteen 8-week old C57BL/6J female mice. Eight of the mice were fed 0.2% cuprizone for 6 weeks (CPZ), which corresponds to a demyelination without recovery phase, and eight healthy age-matched wild-type (WT) mice of the same background were fed a normal chow diet and used as controls. Using electronmicroscopy (EM) data from the same mice, we estimate the mean and standard deviation of the myelin thickness, axonal diameter and intra-axonal volume fraction. For more details regarding the acquisition, preprocessing and histology, see Section 7.2.1.

8.2.2 Neural network

We implement the neural network using the open source Keras toolkit [160] with Tensorflow backend [161]. We minimise the loss using ADAM [122] for 100 epochs with learning rate 0.001. We use a 3 hidden layer architecture with 300, 150 and 75 neurons with 'ReLU' activations and dropout 0.4. We select the best performing architecture as the one that minimises the mean squared error on a validation set formed of 2,000 DW-MRI synthetic signals.

We train the neural network to estimate three microstructure parameters of interest: the intra-axonal volume fraction f, the intra-axonal exchange time τ_i and the intrinsic diffusivity d. For this, we use our synthetic database of 12,500 DW-MRI signals and their associated ground truth tissue parameters acting as labels during the supervised learning task. We use 10,000 synthetic signals for training and 2,500 for testing. We train two neural networks, one using the noise-free database and another using the SNR=40 database. The noise-free trained neural network provides a benchmark performance of our model under ideal noise conditions, while the neural network trained on the SNR=40 database provides an indication of the performance we might expect to see under realistic noise conditions. For the in-vivo estimation, we train the neural network on the SNR=40 database as the one the most closely resembles the in-vivo data.

8.2.3 Experiments

For the experiments in this chapter, we train the neural network using the DW-MRI signal databases for the synthetic and in-vivo data since the features extraction of the high G shells ($G > 300$ mT/m) did not produce reasonable results. However, we do not expect to see any significant differences between the two training approaches according to the results in Sections 6.3.3.

8.2.3.1 Synthetic experiments

To assess the quality of the neural network estimates after training is completed, we compute the Pearson correlation coefficient R^2 between the ground truth values and the parameter estimates in the previously unseen test set. To evaluate any potential bias in the estimates, we use Bland-Altman plots showing the mean of the ground truth and estimated parameter values against their difference. We first analyse the performance of the model on the noise-free synthetic databases to establish a benchmark given our data and imaging protocol. Next, we apply our machine learning model to the SNR=40 database for a more accurate approximation of the performance we expect given the noise present in our in-vivo data.

8.2.3.2 In-vivo experiments

Using the neural network trained on the noisy database, we generate parameter maps for the CCs of the 16 mice for three parameters of interest: τ_i, f and d. To investigate the difference between the two groups (CPZ and WT), we compute box-and-whisker plots of region-specific comparisons between WT (8 mice) and CPZ (8 mice) for the DTI and NODDI metrics as well as for the neural network estimates. Statistical significance is assessed by a two-tailed t-test, considering p-values< 0.05. We run these experiments using the signals database. We compare the mean ROI τ_i and f estimates of the neural network with EM measurements of myelin thickness and intra-axonal volume fraction from the same mice. For more details on how the EM measurements of myelin thickness and the intra-axonal volume fraction are computed, see Section 7.2.2.2 in the previous chapter.

8.3 Results

8.3.1 Synthetic experiments

Figure 8.1 shows how the training and the mean squared errors of the training (blue curve) and the validation set (orange curve) change after each epoch during training. The training error (blue curve) drops significantly during the first 20 epochs, after which it decreases in smaller amounts, suggesting that the neural network is still learning. The change in the validation error indicates whether the network is generalising well to new data or whether it is overfitting. If the validation error

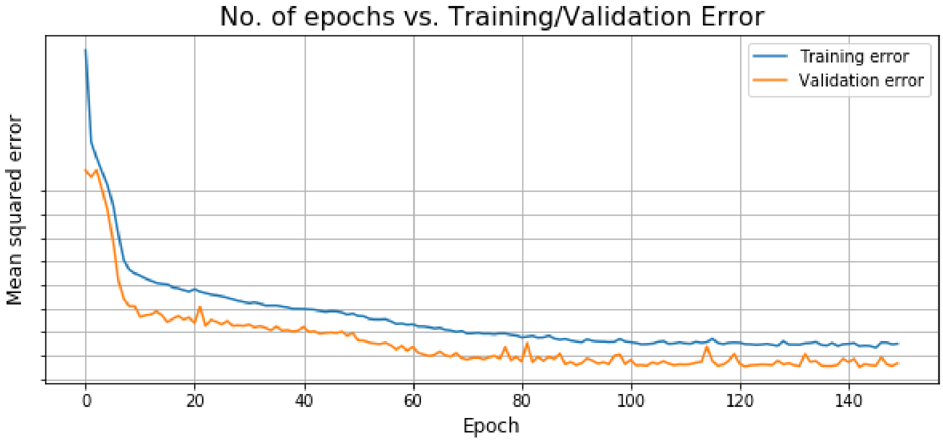

Figure 8.1: Figure shows how the training and the validation error change with each epoch of training.

decreases we can say that the model is learning to generalise better to new data. Conversely, an increase in the validation error suggests that the network is learning the training data too closely (or overfitting) and does not generalise well to new data. The validation error curve (orange) in Fig 8.1 drops significantly during the first 20 epochs, after which it decreases mildly and finally flattens out around epoch 100. This suggests that the first 100 training epochs improve the performance of the neural network, after which the performance stays constant. Consequently, we set the number of training epochs to 100.

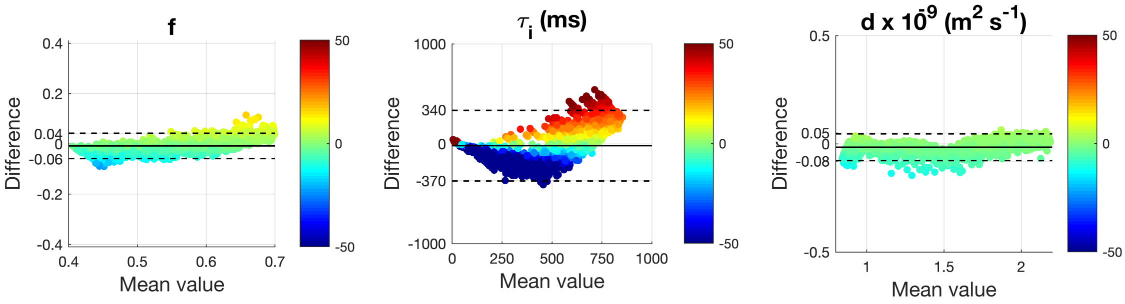

Figure 8.2: Bland-Altman plots for the neural network estimates of f, τ_i and d using the noise-free simulated database. To aid visual interpretation, the plots are colour-coded with the percentage error capped at ±50%.

Figure 8.2 shows the neural network results obtained using the noise-free synthetic database. To assess the quality of our fit, we display the results using Bland-Altman plots and colour each data point according to how close the estimates are to the ground truth values. To aid visual interpretation, we cap the percentage error at ±50%. The mean difference between the ground truth and the estimated values is shown by the black line and the 95% upper and lower limits of agreement by the dashed lines. For all three parameters of interest, we observe no overall estimation bias as the estimates are spread equally around the zero-difference black line. However, for τ_i, short exchange times ($\tau_i < 500$ ms) are overestimated and large exchange times are underestimated ($\tau_i > 500$ ms). This results in wider 95% upper and lower limits of agreement (dotted lines) than for f and d. The recovery of this parameter is not perfect, nevertheless, the correlation coefficient is still strong $R^2_{\tau_i} = 0.83$. Exchange times up to 700 ms are estimated well, after which τ_i is consistently underestimated. For f and d, we find a very strong correlation between the ground truth and the estimates of the neural network ($R^2_f = 0.98$, $R^2_d = 0.99$) with no significant estimation bias.

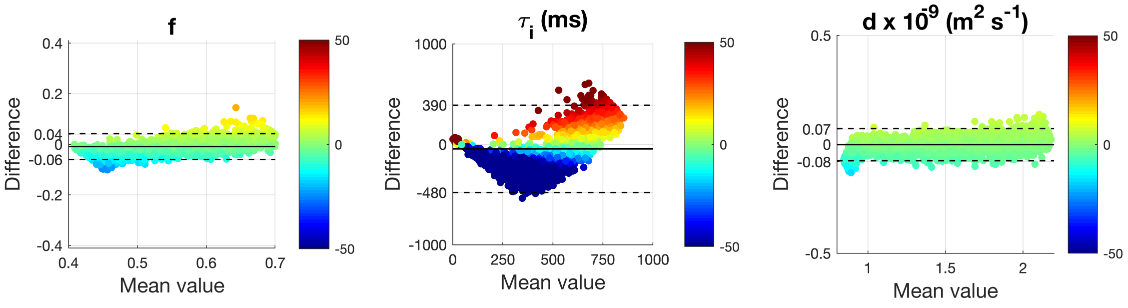

Figure 8.3: Bland-Altman plots for the neural network estimates of f, τ_i and d using the SNR=40 simulated database. To aid visual interpretation, the plots are colour-coded with the percentage error capped at ±50%.

Figure 8.3 shows the equivalent results for SNR=40. The presence of noise results in wider limits of agreement for τ_i and d and affects differently the prediction of each parameter. The mean difference lines for all three parameters remain at zero, showing no general bias in the estimates. Intra-axonal volume fraction and diffusivity continue to be very well estimated and their correlation coefficients are only very mildly affected by the presence of noise: $R^2_f = 0.96$ and $R^2_d = 0.99$. The limits of agreement for f remain unchanged from the noise-free case, while for d only the upper limit changes very slightly, from 0.05 to 0.07. This suggests that despite the presence of noise, the neural network continues to estimate well f and

d. In contrast to this, the presence of noise has a stronger effect on the estimation of τ_i, resulting in a lower overall R^2 score (0.63 versus 0.83 in the noise-free case), together with a slightly more pronounced overestimation/underestimation bias for small and large values respectively. Despite this, we find that the neural network works well within the sensitivity range computed in Section 6.3.1, with a very good correlation coefficient between the models estimates and ground truth for $\tau_i \leq 400$ ms (R^2=0.75). Outside this indicative sensitivity range, the correlation coefficient is very weak: R^2= 0.13 for $\tau_i \geq 400$ ms.

	SNR=40			SNR=∞		
	f	τ_i	d	f	τ_i	d
NN	0.96	0.75	0.99	0.98	0.83	0.99
RF	0.97	0.68	0.99	0.99	0.84	0.99

Table 8.1: Neural network (NN) and random forest (RF) correlation coefficients (R^2 scores) for the synthetic databases (noise free and SNR=40).

Table 8.1 summarises the performance of the neural network and random forest (from Chapter 6) on the noise-free and SNR=40 databases. In the noise-free scenario (SNR=∞), the neural network and the random forest have the very similar performances. Under noise conditions matching our in-vivo data (SNR=40), the neural network outperforms the random forest in the estimation of τ_i, with a correlation coefficient higher by 0.07 than that of the random forest.

8.3.2 In-vivo experiments

The in-vivo neural network estimates of f, τ_i and d obtained using the raw signal database are presented in Figure 8.4. In Figure 8.4A we show parametric maps of the CC across a representative sagittal slice for an examplar healthy WT mouse (first column) and one exemplar CPZ mouse (second column). A simple visual inspection of the CC maps reveals a decrease in f and τ_i, together with a very mild increase in d between the WT and CPZ mice. The values estimated for all parameters are within plausible ranges for rat white matter tissue. For the control group, estimates of volume fraction f are in the range 0.4-0.46, estimates for τ_i are consistently in the range 330-400 ms, and the estimates for d lie in between 1-1.3 $x10^{-9}\text{m}^2\text{s}^{-1}$. To provide a more quantitative analysis, we plot the box and

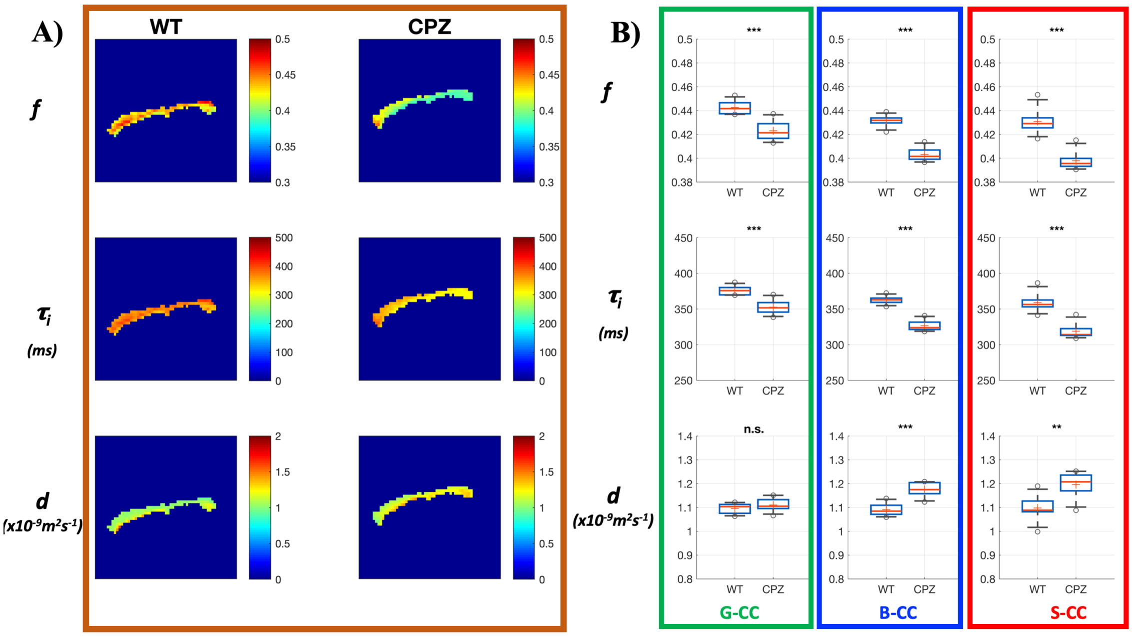

Figure 8.4: **A)** Parametric maps with the neural network estimates for f, τ_i and d in the CC of a healthy WT mouse (first column) and a CPZ mouse (second column). **B)** Box and whisker plots of region-specific comparison between WT ($N_{mice} = 8$) and CPZ ($N_{mice} = 8$). Parameter estimates for f, τ_i and d are computed independently for all voxels within the genu (G-CC), body (B-CC) and splenium (S-CC) of the CC. Statistical significance was assessed by using a 2-tailed t-test with equal variance and significance level: $^* = 0.01, ^{**} = 0.005, ^{***} = 0.001$. '*n.s*' stands for *non-significant*.

whisker plots of region-specific parameter comparisons between the WT and the CPZ group over the three CC ROIs (Figure 8.4B). The trends observed visually in Figure 8.4A) hold for the group-wise quantitative comparison (WT versus CPZ). For f and τ_i, we observe statistically significant decreases in all three regions of the CC (p-val < 0.001). For d, the changes between WT and CPZ are smaller than for the other two parameters and depend on the ROI. In the G-CC, we find a statistically insignificant change, while in the B-CC and S-CC, the increase in d is statistically significant (p-val$_{G-CC} < 0.005$ and p-val$_{S-CC} < 0.001$). The mean and standard deviations of the neural network parameter estimates for each ROI are reported in Table 8.2.

Next, we study the correlation between the changes in f and τ_i as estimated by the neural network and the EM estimates of the intra-axonal exchange time and myelin thickness respectively. We assess the statistical significance of the linear correlation between τ_i and myelin thickness from EM with a two-tailed t-test by

	f		τ_i (ms)		**d** $x10^{-9}(\mathrm{m^2s^{-1}})$	
	WT	**CPZ**	**WT**	**CPZ**	**WT**	**CPZ**
G-CC	0.443 (0.006)	0.423 (0.009)***	376 (7)	353 (11) ***	1.10 (0.022)	1.11 (0.029)
B-CC	0.432 (0.005)	0.404 (0.006)***	363 (6)	327 (8) ***	1.10 (0.027)	1.17 (0.031)***
S-CC	0.431 (0.011)	0.400 (0.008)***	359 (13)	320 (11) ***	1.10 (0.055)	1.19 (0.055)**

Table 8.2: Mean and standard deviation of neural network estimates for f, τ_i and d in the three CC ROIs for the WT and CPZ group. CPZ regions that are statistically different from WT regions are marked with * for $p < 0.01$, ** for $p < 0.005$ and *** for $p < 0.001$

looking at the mean and the standard deviation of each CC ROI of the WT *(blue squares)* and CPZ *(black circles)* group (Figure 8.5A). For τ_i and myelin thickness, we find a linear correlation coefficient ρ of 0.75 and a p-value of 0.08. Despite the good correlation coefficient between the neural network estimates of the exchange time from DW-MRI *(y-axis)* and histological measurements of myelin thickness *(x-axis)*, the p-value above the significance level 0.05 indicates that the correlation between the two set of measurements is not statistically significant.

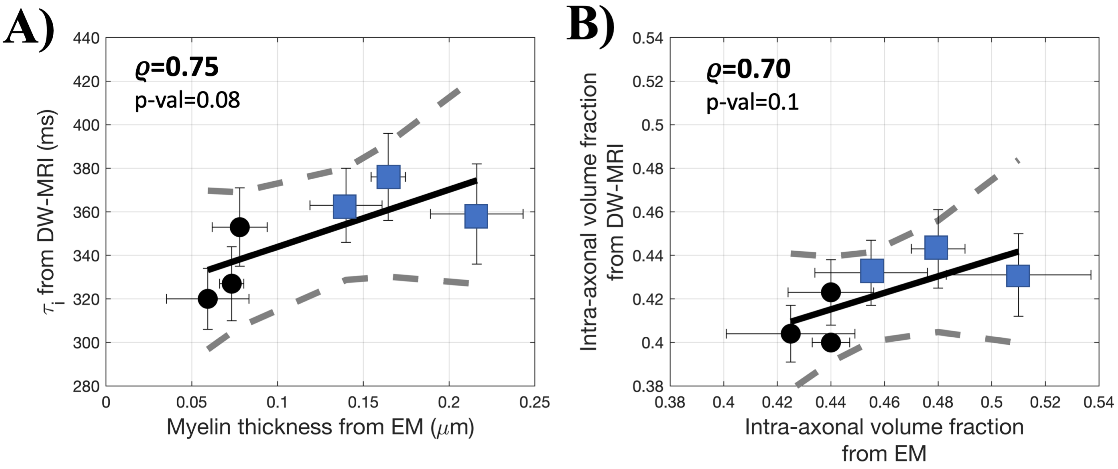

Figure 8.5: Linear correlations between: **A)** the exchange time from DW-MRI *(y-axis)* and myelin thickness from EM *(x-axis)* and **B)** the intra-axonal volume fraction from DW-MRI *(y-axis)* and EM *(x-axis)*. Each point represents the mean over one region of the CC for the WT *(blue squares)* and CPZ *(black circles)* group. Error bars indicate the standard deviation over the region

Similarly, we investigate the statistical significance of the linear correlation between intra-axonal volume fraction f as estimated from DW-MRI*(y-axis)* and EM measurements *(x-axis)* (Figure 8.5B). We find a Pearson correlation coefficient ρ of 0.70 and a p-value of 0.1, showing a good correlation between the neural network estimates and the histological measurements of the intra-axonal volume fraction. Despite this, the p-value of this correlation is above the significance level of 0.05.

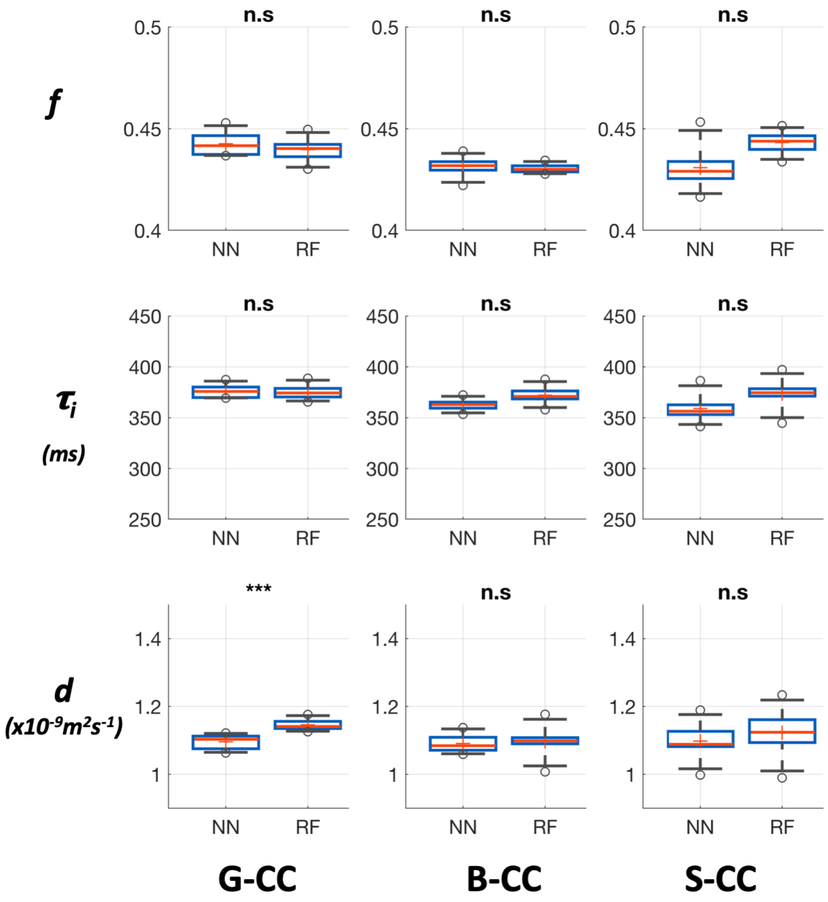

Figure 8.6: Box and whisker plots of region-specific comparison between the WT estimates of the neural network in this chapter and the random forest in Chapter 7. Statistical significance was assessed by using a 2-tailed t-test with equal variance and significance level: $^* = 0.01$,$^{**} = 0.005$,$^{***} = 0.001$. '*n.s*' stands for *non-significant*.

Finally, we compare the estimates of the neural network with those of the random forest from Chapter 7. For this, we show box and whisker plots of region-specific parameter comparisons between the neural network (NN) and random forest (RF) for the healthy control WT group. We show the results in Figure 8.6. We assess statistical significance by using a two-tailed t-test, and we indicate p-values < 0.001 with '***', p-values < 0.005 with '**' and p-values < 0.01 with '**' and p-values > 0.01 indicating non-statistically significant differences with 'n.s.'. The figure reveals a very good agreement between the two machine learning approaches, with no significant differences for f and τ_i for any ROIs in the CC. For d, we observe no significant differences in B-CC and S-CC, and a small difference in G-CC. Looking closer at estimated values in this region, we find that the mean random forest estimate is higher by only 2.7% percent than the neural network estimate (1.14 versus 1.1). Unlike the other regions, the estimates in the G-CC are very closely clustered around the mean, which results in a p-value below the significance level, despite the very small difference in the mean.

8.4 Discussion

Synthetic experiments. The neural network model estimates in the noise-free case have very strong correlations with the ground truth values, providing an excellent benchmark performance for our model and imaging protocol (f: R^2=0.98, τ_i: R^2=0.83 d: R^2=0.99). We show that the addition of noise with SNR=40, matching our in-vivo data, does not affect much the estimation of f and d (f: R^2=0.96, d: R^2=0.99), however, it has a stronger effect on the estimation of τ_i. In line with our sensitivity results, for $\tau_i < 400$ ms the effect is present, however, the performance is still good (R^2=0.75), while for $\tau_i > 400$ ms the performance of the model is severely affected (R^2=0.13).

In-vivo experiments. As discussed in the previous chapter, the absolute values of the estimates are very difficult to validate, especially for τ_i. Despite this, we find that the neural network estimations of τ_i in the healthy mice compare well with the values reported in previous studies. Our estimates of τ_i for the WT group are between 330-400 ms, a range which lies within the 300-600 ms reported estimates of the exchange time in several literature studies discussed in the previous chapters. We find a good correlation between the neural network estimates of τ_i and myelin thickness ($\rho_{\tau_i} = 0.75$). Despite this, the p-value of the correlation is slightly above the 0.05 significance level (p-val = 0.08), indicating a weaker correlation than in the case of the random forest (p-val < 0.05). For f, the availability of EM measurements of intra-axonal volume fraction allow for a direct comparison. We find a good correlation between the estimates of the neural network and the histological data ($\rho_f = 0.70$). Similarly to τ_i, the p-value of the correlation is above the significance level (p-val = 0.1), in contrast with the random forest estimates of f which correlate very well with histological data ($\rho_f = 0.98$, p-val < 0.001). For d, the neural network estimates lie in the range 1-1.3 $\mu\text{m}^2\text{ms}^{-1}$, an expected range for the mouse CC [165].

When comparing the two groups, we observe the following general trends: a statistically significant decrease in the intra-axonal volume fraction f and in the intra-axonal exchange time τ_i over all CC ROIs, together with an increase in the intrinsic diffusivity d, negligible in the genu and significant in the splenium and body. Our neural network model estimates a decrease in f and τ_i between the WT and the CPZ group, in line with expected trends in demyelinated tissue. In the case of diffusivity, we notice a slight increase across the body and the genu. This could be explained by an increase in radial diffusivity as the water moves more freely along the radial direction due to demyelination. Indeed, our radial diffusivity measurements from the DTI fit in Figure 7.4 from Chapter 7 show that there is an increase in the diffusivity along the radial direction.

Random forest (RF) versus neural networks (NN) for microstructure imaging. When comparing the performance of the two machine learning approaches, we find that for simulations mimicking our in-vivo data (SNR=40), the neural network outperforms the random forest in the estimation of τ_i (neural network: $R^2 = 0.75$ vs. random forest: $R^2 = 0.68$). For f and d, the differences between the two algorithms are negligible (neural network/random forest: $R^2_f = 0.96/0.97$ and $R^2_d = 0.99/0.99$). Moreover, we find a very good agreement between the two approaches for f and τ_i (see Fig. 8.6), with all differences above the statistically significant level. Nevertheless, when comparing the neural network and the random forest estimates of f and τ_i with electronmicroscopy data from the same mice, we find that the random forest estimates correlate better with histological data.

It is generally accepted that the performance of neural networks is extremely challenging to explain due to their black-box nature [180]. Recent research [123, 181, 182] shows that neural networks can sometimes learn perfectly the training data due to their high effective capacity [123]. In other words, due to their high number of parameters, neural networks act as powerful regressor that learn the training data very well [7, 181], sometimes including idiosyncrasies that do not apply to slightly different test data [183].

While the high learning capacity of neural networks can be an advantage when the test data is very similar to the training data, the model performance is affected when there are differences between the train and the test datasets, resulting in arbitrary changes in the predictions [182]. Indeed, our case corresponds to the former scenario as our training data is formed of synthetic simulations of the DW-MRI signal, while the test data is formed of real in-vivo DW-MRI signals. Even though our simulations closely mimic the in-vivo data, we expect some of the underlying tissue properties not accounted for in our simulations (e.g. undulation, dispersion) to result in subtle differences between the synthetic training signals and in-vivo test signals. Such subtle, imperceptible differences in the input data have been shown to lead to arbitrary and profound changes in a neural network's output [182, 123], as the model learns aspects of the training data that do not generalise to the test data. Therefore, the expected differences between synthetic and real DW-MRI signals may cause arbitrary alterations of our neural network's in-vivo estimations, which could impact negatively the correlation of the network's estimates with histological measurements.

Indeed, our results show that when the test data is very similar to the training data, as it is the case with the synthetic SNR=40 database, the neural network generalises very well, exceeding the performance of the random forest. On the other hand, when the test data has small perturbations not accounted for in the training data, as it is the case with our in-vivo data (e.g. effect of dispersion), the neural network estimates correlate less well with histological data than the random forest estimations. These results are in line with other literature studies which show that sometimes more complex models such as neural networks perform less well on real datasets than simpler models [184, 183]. A potential explanation for this is that complex models with a high-number of parameters learn the training data too closely, including aspects which do not generalise well outside the training set, leading to a type of overfitting [185].

One way to address this limitation and improve the generalisation capacity of a neural network is through ensembling, where a collection of neural networks are trained and their predictions combined, similarly to the decision trees in the random forest. Ensembling has been shown to improve the generalisation capacity of a single machine learning model [186] and is used extensively to achieve state-of-the-art performance in many domains [187]. Formed of a collection of decision trees, random forests are ensemble methods and, therefore, expected to generalise better than single models [108]. This constitutes a potential explanation for the better correlation of our random forest results with histological data. In addition to ensembling techniques, there are several other neural network architectures (e.g. generative adversarial networks) which have the potential to address the limitations of our multilayer perceptron. These are subject to ongoing research [188, 181, 182] and form part of future work, as discussed in the next chapter.

8.5 Conclusions

To conclude, this chapter assesses the performance of a neural network based computational model with permeability using an in-vivo mouse model of demyelination. Our results show that, in simulations, the neural network outperforms the random forest when estimating τ_i (neural network: $R^2 = 0.75$, random forest: $R^2 = 0.68$), in line with the findings in Chapter 5. Next, we apply our model to the in-vivo data, consisting of a cohort of 8 CPZ mice and 8 healthy age-matched WT mice, and we find that the model estimates sensible microstructure parameters for rat white matter and detects trends consistent with expectations from the pathology of demyelination (significant decreases in f and τ_i and a mild increase in d). We compare the model's estimates of τ_i and f with EM measurements from the same mice and find a good correlation between the in-vivo estimates of τ_i and myelin thickness ($\rho_{\tau_i} = 0.75$), and between the neural network estimations and EM measurements of f ($\rho_f = 0.70$). However, we find that the p-value of the correlations between the estimations and the histological data is above the significance level of 0.05 ($\text{p-val}_{\tau_i} = 0.08$ and $\text{p-val}_{\tau_i} = 0.1$), in contrast to the random forest estimates in Chapter 7, which show a strong, statistically significant correlation with histology.

Overall, the results in this the last three chapters suggest that, in simulations, the neural network model provides marginally better estimates of τ_i than the random forest approach. In-vivo, both machine learning approaches provide plausible parameter estimates for white matter tissue and detect the same expected trends between the control and demyelinated cuprizone groups: a significant decrease in f and τ_i and a small increase in d. Despite this, the neural network estimates do not correlate as well with histology measurements as the random forest estimates, which show strong correlations with electronmicroscopy measurements of f and myelin thickness. In addition to this, we previously show (in Chapter 5) that the random forest estimates of f and τ_i also have better reproducibility than the neural network estimates. All this suggests that random forest based computational models are a more suitable choice for machine learning based models with permeability for white matter microstructure imaging than simple multilayer perceptrons neural network models. Nevertheless, the field of neural networks is subject to continuous research and new architectures might provide additional improvements over other machine learning algorithms such as random forests, as we will discuss in the next chapter.

Chapter 9

Conclusions and future work

This chapter discusses the contributions of this thesis from a general perspective and details possible future extensions based on the current work.

9.1 Summary

The aim of this thesis was to develop and test the feasibility of machine learning based computational models with permeability for white matter microstructure imaging. The motivation of this research was to contribute towards the development of imaging biomarkers for MS and other demyelinating pathologies.

Current mathematical models with permeability, such as the Karger model [85] or AXR [95], are based on assumptions that do no hold in the brain white matter tissue, while mathematical expressions that accurately incorporate permeability remain intractable. Computational models bypass the need for mathematical expressions by using simulations with available ground truth microstructure parameters to learn how permeability affects the DW-MRI signals. Machine learning algorithms are particularly suited for this task as they have been shown to be very good at discovering intricate structures in high-dimensional data, improving on the state-of-the-art in a multitude of domains including medical imaging [7]. To date, the performance of machine learning methods for microstructure parameter estimation such as intra-axonal exchange time τ_i has not been investigated. The experiments in this thesis were therefore designed with the following specific objectives:

1. To build a machine learning based computational framework that learns a mapping between synthetic DW-MRI signals and microstructure parameters,

including τ_i, using a random forest and a neural network approach.

2. To assess the performance of the machine learning computational models on noise-free and synthetic data to obtain indicative benchmark performances.

3. To asses the in-vivo performance of the machine learning computational framework and study its feasibility as an imaging biomarker for demyelinating pathologies, with a focus on MS.

4. To experimentally investigate the performance of the random forest and the neural network based computational models when compared to histology.

The conclusions of the investigations in this thesis are discussed below.

9.1.1 Evaluating the performance of the random forest and the neural network models on synthetic data

Synthetic DW-MRI signals with available ground truth microstructure parameters are necessary to asses the performance of our machine learning computational models under idealised and realistic noise conditions.

We use noise-free simulations to establish a benchmark performance for our models given the data available and under ideal noise conditions (SNR=∞). We show that both the random forest and the neural network approaches have an excellent benchmark performance, with strong correlation coefficients ($\geq$0.83) and no general estimation bias for all parameters of interest.

We use the noisy databases to obtain an indication of the model performance under realistic noise conditions, matching those of our in-vivo data. As expected, we show that the model performance is affected by the presence of noise, especially for parameters such as τ_i , which only weakly influence the DW-MRI signals under clinical scanner settings [95]. This results in wider 95% limits of agreement between the ground truth and the estimated values and lower correlation coefficients than in the noise-free case, especially for τ_i. Moreover, we show in Chapters 5 and 8, that the neural network provides a marginal improvement over the random forest when estimating τ_i: the R^2 correlation coefficient is higher by 0.04 for the STEAM protocol in Chapter 4 and 0.07 for the SDE protocol in Chapter 6 than the R^2 score of the random forest. This suggests that the neural network learns the relationship between the synthetic simulations and τ_i better than the random forest in the presence of noise. This is most likely explained by the high effective learning capacity of neural networks due to their extremely high number of parameters, which enables them to learn all idiosyncrasies of a training set [123]. Despite the small differences between the two, the performance of both approaches is still sufficiently good, with correlation coefficients above 0.68 for all parameters of interest across both models.

In addition to the effect of noise, the performance of the model also depends on the sensitivity of the DW-MRI signal to the parameters of interest. Indeed, we see this being reflected in the estimation of τ_i, where the correlation coefficients over the [0, 1000] ms range are higher when using the optimised STEAM protocol than when using the non-optimised SDE protocol. This finding is also in line with the our sensitivity study in Chapter 3, where we show that long Δs, as in the optimised STEAM protocol, play a crucial role in the sensitivity to the exchange time.

9.1.2 Evaluating the in-vivo performance of the random forest and the neural network models on healthy and demyelinated tissue

In order to investigate the in-vivo performance of our machine learning based computational framework, we analyse the parameter estimates from in-vivo data from clinical human scans (Chapters 4 and 5) and preclinical mouse scans (Chapters 6-8).

In the healthy in-vivo data, our machine learning based computational model provides estimates of f, τ_i and d within plausible ranges for white matter tissue, and that compare well with literature values. In addition to this, we show that our computational framework estimates expected trends across the white matter, such as the high-low-high trend in f across the mid-sagittal CC. Using scan-rescan data from two healthy subjects, we show in Chapter 5, that both the random forest and the neural network models have good reproducibility.

We use in-vivo data from diseased tissue (MS lesions and cuprizone-induced demyelination) to study the feasibility of our computational model as an imaging biomarker for demyelinating pathologies. We find that our machine learning based computational framework estimates trends consistent with expectations from the pathology of demyelinated tissue, demonstrating the clinical potential of this new technique. Both the random forest and the neural network based models estimate a significant decrease in f and τ_i between healthy and diseased tissue, in both human and mouse data. These results agree with expectations from the pathology of MS lesions, where the breakdown of myelin would make axons thinner and, thus, more permeable, as well as reduce the intra-axonal volume fraction. We seek further validation of the trends estimated in f and τ_i through a direct comparison with histological data, as detailed in the following subsection.

When comparing the two machine learning approaches, we find that there is a very good agreement between the in-vivo control estimates of the random forest and the neural network, which we show in Chapters 5 and 8. Additionally, both approaches estimate the same trends across the white matter, such as the high-low-high trend in f. Nevertheless, the random forest scan-scan rescan estimates show a better reproducibility than the neural network for f and τ_i, while for d the methods have equivalent reproducibility.

9.1.3 Evaluating the performance of the random forest and the neural network models when compared to histology

In Chapters 7 and 8, we perform a direct comparison between the random forest and the neural network in-vivo estimates of f and τ_i with histology measurements of f and myelin thickness from a cuprizone mouse model. While estimates of f are directly obtainable from histology, accurate estimates of τ_i are not accessible via histology. Instead, we compare our estimates of τ_i with myelin thickness, which has recently been shown to correlate well with τ_i in simulations [86], in line with previous literature studies hypothesising the link between the two [1, 2, 3].

We show in Chapter 7 that the random forest estimates of f and τ_i are in strong agreement ($\rho_{\tau_i} = 0.82$, $\rho_f = 0.98$) with our EM histology measurements of myelin thickness and intra-axonal volume fraction. In addition to this, we show that potentially confounding factors such as axonal swelling and dispersion have a negligible effect on the estimated differences between the control and the demyelinated group. In Chapter 8, we find that there is a good correlation between the neural network estimates of f and τ_i and histology ($\rho_{\tau_i} = 0.75$, $\rho_f = 070$), however, the correlations are less strong than for the random forest and above the statistically significant level. A potential explanation for this is that the neural network learns the simulated data better than the random forest, as supported by our synthetic data results in section 9.1.1, but does not generalise as well when there are slight differences between the training (i.e. the synthetic data) and the test set (i.e. the in-vivo data), consistent with findings in several other literature studies [182, 181, 185]. On the other hand, ensemble methods, such as random forests, have been shown to improve the generalisation ability of single models, such as neural networks [186] or a single decision trees [108], which further explains the better in-vivo performance of our random forest model. Nevertheless, due to recent advances in the field of transfer learning, which focuses on improving the performance of neural networks when there are differences between the train and the test set, new network architectures that aim to address this issue have become available [189, 187] and form part of future work.

The results in this thesis show for the first time, quantitatively and in-vivo, the clinical potential of machine learning based computational models with permeability as a suitable biomarkers for the detection and tracking of changes in demyelinating pathologies such as MS. Additional improvements in the tissue model, the diffusion sequence and the machine learning approach could further increase the microstructure parameter estimation accuracy and form part of future work, as detailed in the next and final section of this thesis.

9.2 Future work

There are several improvements and general future directions that could further improve the accuracy of microstructure parameters estimations using machine learning based computational models of white matter with permeability.

9.2.1 Tissue models

One limitation of the present work is the simplifying set of assumptions of the white matter synthetic substrates used in the Monte Carlo simulations. Due to current limitations in our simulation system, we make several assumptions about the geometry of the tissue such as representing axons as non-abutting parallel cylinders that mimic axon bundles without accounting for the effects of myelin water, dispersion, crossing fibres or the undulation, which are usually present in the in-vivo tissue. Previous work investigating the impact of dispersion on axonal permeability estimation shows that the presence of orientation dispersion could result in underestimated values of τ_i [1], and, therefore, could affect our estimates. Although in Chapter 7 we conclude that we can rule out the effect of dispersion on our estimation of the difference in τ_i, the absolute values of the estimations might still be affected. Consequently, including the effect of dispersion in the simulations may be essential for improving the accuracy of the intra-axonal exchange time. More generally, from a machine learning perspective, using more realistic simulations will also narrow down the gap between the synthetic and in-vivo data and make the generalisation between the two databases more accurate.

Another limitation that stems from using one type of tissue model is that we test the machine learning approaches on the same type of substrates that it was trained on. We note, however, that despite using the same tissue model, the model parameters between the training and the test data are different so the results demonstrate generalisation to some extent. Developing a different tissue model would require substantial further development of our simulation system and is outside the scope of this work. Future work should aim to train and test the machine learning model on more diverse realistic simulations, which are subject to ongoing research and account for different effects such as myelin water [148, 86], axonal undulation [159], dispersion [146, 147], neurons and glial cells [179]. Such effects, once included in the simulations, can easily be incorporated in the machine learning framework used in this paper and may improve the performance of our computational framework.

9.2.2 Diffusion sequences

The simulation study in Chapter 3 covers several clinical and preclinical combinations of DW-MRI sequence parameters without fully exploring the parameter space within the physical constrains of a scanner. In addition to this, we do not vary all degrees of freedom at the same time. Future work can produce a more detailed analysis, similar to the work in [138], by fully exploring the space of each parameter and varying multiple parameters at the same time.

The success of our computational framework depends on the sensitivity of the DW-MRI protocol to our parameters of interest. The sensitivity of the imaging protocol in Chapter 6 to the exchange time in the presence of noise is not ideal. Even though the sensitivity range covers a range of exchange time values that we expect to find in the mouse white matter, the accuracy of the estimates could potentially be improved by optimising our protocol with respect to τ_i as done in [132] or by increasing the maximum diffusion time, in line with our findings from Chapter 3. One of the main advantages of our computational framework is that it can easily be adapted to incorporate not only more complex simulations, but also more specialised diffusion encoding sequences such as AXR [95] or OGSE for more sensitivity to axon diameter [138], or STEAM sequences for longer diffusion times, as shown in Chapter 4.

9.2.3 Validation

An important step in the development of new tissue models, which was not performed for the clinical studies in Chapters 4 and 5, is experimental validation. The gold standard for validating in-vivo microstructure parameter estimates is to compare them with histology. However, within the clinical setting, histological samples are extremely invasive and painful, and sometimes impossible to obtain from certain areas of the brain.

In an attempt to overcome this limitation, we use a preclinical mouse model with available histological data. While we obtain histology measurements of f, accurate estimates of the intra-axonal exchange time τ_i are not obtainable via histology as tissue fixation is known to alter membrane permeability. Here, we try to bypass this limitation and compare our estimates of τ_i with histological measurements of myelin thickness, which has been shown to correlate well with τ_i in simulations [86]. Nevertheless, our validation is not ideal and we propose to adress this limitation in the future by using permeable microstructural phantoms, which have recently become available [190, 191]. While microstructural phantoms offer some degree of validation, they cannot capture the entire microstructural complexity of in-vivo tissue, and validating estimates of τ_i remains an intrinsically difficult task.

9.2.4 Machine learning approaches

Throughout the thesis, one secondary theme has been to compare the performance of random forests and multilayer perceptron neural networks when estimating microstrucuture parameters including the intra-axonal exchange time. We find that while the neural network learns the relationship between noisy simulations and τ_i better than the random forest, the in-vivo estimations of the neural network are less reproducible (Chapter 5) and correlate less well with histology than those of the random forest (Chapter 8). There are several approaches that may improve the in-vivo performance of the neural network and which form part of future work.

One technique which has been shown to improve the generalisation performance of single neural networks is ensembling [186, 187], which trains independently a collection of neural networks and combines their predictions for the final output of the ensemble. Neural networks ensembles are a promising approach, which have been shown to improve the generalisation performance of single neural network models [186] and have been successfully applied to a wide range of domains including medical imaging and diagnosis [187, 192, 193]. Another set of techniques which may improve the performance of our neural network based model belongs to the transfer learning field, which aims to improve a network's ability to generalise to conditions that are different from the ones encountered during training, such as when the network is trained on simulations but is used to estimate on real data. Two of the most popular approaches are domain confusion and adversarial training, which use both the synthetic and real data during training in order to improve the robustness of the neural network to modifications in the test data. This field is currently at the forefront of machine learning research and a list of the most successful techniques so far can be found in [189].

Each of the two machine learning approaches used throughout this thesis has its own advantages and disadvantages. The main advantage of neural networks in the context of this thesis is that their architecture can be adapted and extended to address one of the main limitations on the current work: the gap between the synthetic and the in-vivo data through transfer learning techniques. This can be done by adopting newly introduced architectures such as generative adversarial networks, convolutional neural networks or domain adaptation semi-supervised architectures. As transfer learning is undergoing continuous development, future work implementing these approaches has great potential to improve our current results. However, neural networks are known for being prone to overfitting and difficult to tune [126], while random forests are known for being very robust, easy to tune and interpretable, making them one of the most widely used machine learning algorithms in medical imaging [110]. Therefore, the choice of the machine learning model needs to be carefully assessed according to each scenario.

Another important future direction, applicable to both machine learning approaches, is the inclusion of uncertainty measures, as in [194], to highlight areas of the brain where the model estimates are less reliable due to unfamiliar signals. This is especially important for the potential clinical applications of our computational framework as it ensures informed treatment and diagnosis decisions for white matter pathologies.

Our machine learning based computational framework can easily be adapted to any of the machine learning approaches outlined above, opening the pathway to a plethora of new emerging machine learning techniques.

9.2.5 Applications

Developing imaging techniques which provide information about microstructural features has many potential applications, both in biomedical research for a better understanding of brain pathologies, as well as in the clinical setting for their diagnosing and monitoring. The machine learning based computational framework we use here opens the doors to estimating a wider range of microstructure parameters for which mathematical models remain intractable such as undulation or properties of the extracellular space. The estimations of these otherwise intractable parameters offers not only clinically useful information, but accounting for their effects may also improve the estimation of other microstructure parameters such as the axon diameter, which is usually overestimated by compartment models [73].

The extension of our white matter model with permeability to clinical systems can potentially be important for numerous white matter pathologies of the human nervous system. In this thesis, we preliminary show that our approach may be suitable for clinical and biomedical research applications through a qualitative study on two healthy and two MS patients, where we show that our approach provides parameter estimates consistent with expectations from the pathology of MS lesions. We further consolidate this hypothesis by demonstrating our framework's clinical potential as a biomarker for MS in a cuprizone mouse model of demyelination, which is extensively used in the MS literature due to its close similarity to the demyelination and remyelination processes occurring in MS lesions [162].

The applicability of our approach extends to other myelin damaging pathologies such as spinal cord injury or leukodystrophies due to the hypothesised correlation between τ_i and the condition of the myelin sheath [1, 2, 3]. In addition to this, the clinical applications of our framework can also be extended to other pathologies such as Parkinson's disease or cancer, which have been shown to cause a change in the permeability of the tissue [4, 5]. The current key limitation is the reduced sensitivity to the intra-axonal exchange time of clinically available imaging protocols and the simplifying assumptions of the tissue model . Nevertheless, with the continually improving SNR in clinical scanners and the availability of more realistic simulations, the number of potential clinical and biomedical research applications of machine learning based computational models is likely to increase in the future.

Bibliography

[1] Markus Nilsson, Danielle van Westen, Freddy Ståhlberg, Pia C Sundgren, and Jimmy Lätt. The role of tissue microstructure and water exchange in biophysical modelling of diffusion in white matter. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 26(4):345–370, 2013.

[2] J Chetley Ford and David B Hackney. Numerical model for calculation of apparent diffusion coefficients (ADC) in permeable cylinderscomparison with measured ADC in spinal cord white matter. *Magnetic resonance in medicine*, 37(3):387–394, 1997.

[3] Scott N Hwang, Chih-Liang Chin, Felix W Wehrli, and David B Hackney. An image-based finite difference model for simulating restricted diffusion. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 50(2):373–382, 2003.

[4] Michael J Volles, Seung-Jae Lee, Jean-Christophe Rochet, Mark D Shtilerman, Tomas T Ding, Jeffrey C Kessler, and Peter T Lansbury. Vesicle permeabilization by protofibrillar α-synuclein: implications for the pathogenesis and treatment of Parkinson's disease. *Biochemistry*, 40(26):7812–7819, 2001.

[5] Jie Hu, AS Verkman, Jie Hu, and AS Verkman. Increased migration and metastatic potential of tumor cells expressing aquaporin water channels. *The FASEB Journal*, 20(11):1892–1894, 2006.

[6] Markus Nilsson, Erik Alerstam, Ronnie Wirestam, F Sta, Sara Brockstedt, Jimmy Lätt, et al. Evaluating the accuracy and precision of a two-

compartment Kärger model using Monte Carlo simulations. *Journal of Magnetic Resonance*, 206(1):59–67, 2010.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[8] Antonio Criminisi and Jamie Shotton. *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.

[9] Alan Finkelstein. Water and nonelectrolyte permeability of lipid bilayer membranes. *The Journal of general physiology*, 68(2):127–135, 1976.

[10] Andrew Michael Prantner. *Re-evaluation of transmembrane water exchange in the rat brain*. Washington University in St. Louis, 2008.

[11] James D Quirk, G Larry Bretthorst, Timothy Q Duong, Avi Z Snyder, Charles S Springer, Joseph JH Ackerman, and Jeffrey J Neil. Equilibrium water exchange between the intra-and extracellular spaces of mammalian brain. *Magnetic resonance in medicine*, 50(3):493–499, 2003.

[12] Markus Nilsson, Jimmy Lätt, Emil Nordh, Ronnie Wirestam, Freddy Ståhlberg, and Sara Brockstedt. On the effects of a varied diffusion time in vivo: is the diffusion in white matter restricted? *Magnetic resonance imaging*, 27(2):176–187, 2009.

[13] Julia M Edgar and I Griffiths. White Matter Structure: A Microscopists View. *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy*, 75, 2009.

[14] Eva-Maria Krämer-Albers, Katja Gehrig-Burger, Christoph Thiele, Jacqueline Trotter, and Klaus-Armin Nave. Perturbed interactions of mutant proteolipid protein/DM20 with cholesterol and lipid rafts in oligodendroglia: implications for dysmyelination in spastic paraplegia. *Journal of Neuroscience*, 26(45):11743–11752, 2006.

[15] Sabine Hofer and Jens Frahm. Topography of the human corpus callosum revisitedcomprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage*, 32(3):989–994, 2006.

[16] Francisco Aboitiz, Arnold B Scheibel, Robin S Fisher, and Eran Zaidel. Fiber composition of the human corpus callosum. *Brain research*, 598(1):143–153, 1992.

[17] Ricardo Olivares, Juan Montiel, and Francisco Aboitiz. Species differences and similarities in the fine structure of the mammalian corpus callosum. *Brain, behavior and evolution*, 57(2):98–105, 2001.

[18] A-S LaMantia and P Rakic. Cytological and quantitative characteristics of four cerebral commissures in the rhesus monkey. *Journal of Comparative Neurology*, 291(4):520–537, 1990.

[19] A Compston and A Coles. Multiple sclerosis. 372(9648):1502–1517, 2008.

[20] Thomas Zeis, Ursula Graumann, Richard Reynolds, and Nicole Schaeren-Wiemers. Normal-appearing white matter in multiple sclerosis is in a subtle balance between inflammation and neuroprotection. *Brain*, 131(1):288–303, 2007.

[21] Ranjan Dutta and Bruce D Trapp. Relapsing and progressive forms of multiple sclerosis–insights from pathology. *Current opinion in neurology*, 27(3):271, 2014.

[22] Genaro Gabriel Ortiz, Fermín Paul Pacheco-Moisés, Miguel Ángel Macías-Islas, Luis Javier Flores-Alvarado, Mario A Mireles-Ramírez, Erika Daniela González-Renovato, Vanessa Elizabeth Hernández-Navarro, Angélica Lizeth Sánchez-López, and Moisés Alejandro Alatorre-Jiménez. Role of the blood–brain barrier in multiple sclerosis. *Archives of medical research*, 45(8):687–697, 2014.

[23] Alberto Ascherio and Kassandra L Munger. Environmental risk factors for multiple sclerosis. Part i: the role of infection. *Annals of neurology*, 61(4):288–299, 2007.

[24] Jin Nakahara, Michiko Maeda, Sadakazu Aiso, and Norihiro Suzuki. Current concepts in multiple sclerosis: autoimmunity versus oligodendrogliopathy. *Clinical reviews in allergy & immunology*, 42(1):26–34, 2012.

[25] M Trojano and D Paolicelli. The differential diagnosis of multiple sclerosis: classification and clinical features of relapsing and progressive neurological syndromes. *Neurological Sciences*, 22(2):S98–S102, 2001.

[26] W Ian McDonald, Alistair Compston, Gilles Edan, Donald Goodkin, Hans-Peter Hartung, Fred D Lublin, Henry F McFarland, Donald W Paty, Chris H Polman, Stephen C Reingold, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 50(1):121–127, 2001.

[27] Achilles Ntranos and Fred Lublin. Diagnostic criteria, classification and treatment goals in multiple sclerosis: the chronicles of time and space. *Current neurology and neuroscience reports*, 16(10):90, 2016.

[28] Frederik Barkhof, Massimo Filippi, David H Miller, Philip Scheltens, Adriana Campi, Chris H Polman, Giancarlo Comi, Herman J Ader, Nick Losseff, and Jacob Valk. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain: a journal of neurology*, 120(11):2059–2069, 1997.

[29] Alan J Thompson, Brenda L Banwell, Frederik Barkhof, William M Carroll, Timothy Coetzee, Giancarlo Comi, Jorge Correale, Franz Fazekas, Massimo Filippi, Mark S Freedman, et al. Diagnosis of multiple sclerosis: 2017 revisions of the mcdonald criteria. *The Lancet Neurology*, 17(2):162–173, 2018.

[30] Andrew Webb and George C Kagadis. Introduction to biomedical imaging. *Medical Physics*, 30(8):2267–2267, 2003.

[31] Massimo Filippi, Maria A Rocca, Nicola De Stefano, Christian Enzinger, Elizabeth Fisher, Mark A Horsfield, Matilde Inglese, Daniel Pelletier, and Giancarlo Comi. Magnetic resonance techniques in multiple sclerosis: the present and the future. *Archives of neurology*, 68(12):1514–1520, 2011.

[32] A Gass, GJ Barker, D Kidd, JW Thorpe, D MacManus, A Brennan, PS Tofts, AJ Thompson, WI McDonald, and DH Miller. Correlation of magnetization transfer ration with clinical disability in multiple sclerosis. *Annals of neurology*, 36(1):62–67, 1994.

[33] Marco Rovaris, Federica Agosta, Elisabetta Pagani, and Massimo Filippi. Diffusion tensor MR imaging. *Neuroimaging Clinics of North America*, 19(1):37–43, 2009.

[34] David Paling, Bhavana S Solanky, Frank Riemer, Daniel J Tozer, Claudia AM Wheeler-Kingshott, Raju Kapoor, Xavier Golay, and David H Miller. Sodium accumulation is associated with disability and a progressive course in multiple sclerosis. *Brain*, 136(7):2305–2317, 2013.

[35] C Laule, E Leung, D KB Li, AL Traboulsee, DW Paty, AL MacKay, and G RW Moore. Myelin water imaging in multiple sclerosis: quantitative correlations with histopathology. *Multiple Sclerosis Journal*, 12(6):747–753, 2006.

[36] Cynthia Wisnieff, Sriram Ramanan, John Olesik, Susan Gauthier, Yi Wang, and David Pitt. Quantitative susceptibility mapping (QSM) of white matter multiple sclerosis lesions: interpreting positive susceptibility and the presence of iron. *Magnetic resonance in medicine*, 74(2):564–570, 2015.

[37] Marvin M Goldenberg. Multiple sclerosis review. *Pharmacy and Therapeutics*, 37(3):175, 2012.

[38] E Mark Haacke, Robert W Brown, Michael R Thompson, Ramesh Venkatesan, et al. *Magnetic resonance imaging: physical principles and sequence design*, volume 82. Wiley-Liss New York:, 1999.

[39] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[40] Greg J Stanisz, Ewa E Odrobina, Joseph Pun, Michael Escaravage, Simon J Graham, Michael J Bronskill, and R Mark Henkelman. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(3):507–512, 2005.

[41] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging: a signal processing perspective*. SPIE Optical Engineering Press, 2000.

[42] John Bertrand Johnson. Thermal agitation of electricity in conductors. *Physical review*, 32(1):97, 1928.

[43] Harry Nyquist. Thermal agitation of electric charge in conductors. *Physical review*, 32(1):110, 1928.

[44] Francesco Grussu. *Microstructural imaging of the human spinal cord with advanced diffusion MRI*. PhD thesis, UCL (University College London), 2016.

[45] Paul Tofts. *Quantitative MRI of the brain: measuring changes caused by disease*. John Wiley & Sons, 2005.

[46] Heidi Johansen-Berg and Timothy EJ Behrens. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2013.

[47] C Andersen. In vivo estimation of water content in cerebral white matter of brain tumour patients and normal individuals: towards a quantitative brain oedema definition. *Acta neurochirurgica*, 139(3):249–256, 1997.

[48] Ludovico Minati and Wladyslaw P Weglarz. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. *Concepts in Magnetic Resonance Part A*, 30(5):278–307, 2007.

[49] E Hahn. L 1950 Spin echoes Phys. *Rev*, 80:580–94.

[50] Edward O Stejskal and John E Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288–292, 1965.

[51] Herman Y Carr and Edward M Purcell. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical review*, 94(3):630, 1954.

[52] Noam Shemesh, Sune N Jespersen, Daniel C Alexander, Yoram Cohen, Ivana Drobnjak, Tim B Dyrby, Jurgen Finsterbusch, Martin A Koch, Tristan Kuder, Fredrik Laun, et al. Conventions and nomenclature for double diffusion encoding NMR and MRI. *Magnetic resonance in medicine*, 75(1):82–87, 2016.

[53] William S Price. Pulsed-field gradient nuclear magnetic resonance as a tool for studying translational diffusion: Part 1. Basic theory. *Concepts in Magnetic Resonance Part A*, 9(5):299–336, 1997.

[54] Henrik Lundell, Daniel C Alexander, and Tim B Dyrby. High angular resolution diffusion imaging with stimulated echoes: compensation and correction in experiment design and analysis. *NMR in biomedicine*, 27(8):918–925, 2014.

[55] Oscar Brihuega-Moreno, Frank P Heese, and Laurance D Hall. Optimization of diffusion measurements using Cramer-Rao lower bound theory and its application to articular cartilage. *Magnetic resonance in medicine*, 50(5):1069–1076, 2003.

[56] Tim B Dyrby, Matt G Hall, Maurice Ptito, Daniel Alexander, et al. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. *Magnetic resonance in medicine*, 70(3):711–721, 2013.

[57] Peter J Basser, James Mattiello, and Denis LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.

[58] Timothy EJ Behrens and Heidi Johansen-Berg. *Diffusion MRI: from quantitative measurement to in-vivo neuroanatomy*. Academic Press, 2009.

[59] Peter J Basser and E Ozarslan. Anisotropic diffusion: from the apparent diffusion coefficient to the apparent diffusion tensor. *Diffusion MRI: Theory, Methods, and Applications*, pages 79–91, 2010.

[60] Derek K Jones and Peter J Basser. squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. *Magnetic Resonance in Medicine*, 52(5):979–993, 2004.

[61] Cheng Guan Koay. Least squares approaches to diffusion tensor estimation. *Diffusion MRI: Theory, Methods and Applications*, pages 272–284, 2010.

[62] C Beaulieu. What makes diffusion anisotropic in the nervous system, 2011.

[63] Derek K Jones. *Diffusion MRI*. Oxford University Press, 2010.

[64] G Schwarz. Estimating the dimension of a model The Annals of S tatistics 6 (2), 461–464. *URL: http://dx. doi. org/10.1214/aos/1176344136*, 1978.

[65] Eleftheria Panagiotaki, Torben Schneider, Bernard Siow, Matt G Hall, Mark F Lythgoe, and Daniel C Alexander. Compartment models of the diffusion MR signal in brain white matter: a taxonomy and comparison. *Neuroimage*, 59(3):2241–2254, 2012.

[66] Eleftheria Panagiotaki, Hubert Fonteijn, Bernard Siow, Matt G Hall, Anthony Price, Mark F Lythgoe, and Daniel C Alexander. Two-compartment models of the diffusion MR signal in brain white matter. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 329–336. Springer, 2009.

[67] Greg J Stanisz, Graham A Wright, R Mark Henkelman, and Aaron Szafer. An analytical model of restricted diffusion in bovine optic nerve. *Magnetic Resonance in Medicine*, 37(1):103–111, 1997.

[68] Timothy EJ Behrens, Mark W Woolrich, Mark Jenkinson, Heidi Johansen-Berg, Rita G Nunes, Stuart Clare, Paul M Matthews, J Michael Brady, and Stephen M Smith. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic resonance in medicine*, 50(5):1077–1088, 2003.

[69] Yaniv Assaf, Raisa Z Freidlin, Gustavo K Rohde, and Peter J Basser. New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 52(5):965–978, 2004.

[70] Yaniv Assaf, Tamar Blumenfeld-Katzir, Yossi Yovel, and Peter J Basser. AxCaliber: a method for measuring axon diameter distribution from diffusion MRI. *Magnetic resonance in medicine*, 59(6):1347–1354, 2008.

[71] Daniel Barazany, Peter J Basser, and Yaniv Assaf. In vivo measurement of axon diameter distribution in the corpus callosum of rat brain. *Brain*, 132(5):1210–1220, 2009.

[72] Daniel C Alexander. A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. *Magnetic Resonance in Medicine*, 60(2):439–448, 2008.

[73] Daniel C Alexander, Penny L Hubbard, Matt G Hall, Elizabeth A Moore, Maurice Ptito, Geoff JM Parker, and Tim B Dyrby. Orientationally invariant indices of axon diameter and density from diffusion MRI. *Neuroimage*, 52(4):1374–1389, 2010.

[74] Uli Bürgel, Katrin Amunts, Lars Hoemke, Hartmut Mohlberg, Joachim M Gilsbach, and Karl Zilles. White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage*, 29(4):1092–1105, 2006.

[75] Francisco de Assis Aquino Gondim and FP Thomas. Spinal cord, topographical and functional anatomy. *Journal of Electromyography and Kinesiology*, 8(2):125–132, 1998.

[76] Daniel C Alexander and Kiran K Seunarine. Mathematics of crossing fibers. *Diffusion MRI: theory, methods, and applications. Oxford University Press, New York*, pages 451–464, 2010.

[77] Sune N Jespersen, Christopher D Kroenke, Leif Østergaard, Joseph JH Ackerman, and Dmitriy A Yablonskiy. Modeling dendrite density from magnetic resonance diffusion measurements. *Neuroimage*, 34(4):1473–1486, 2007.

[78] Hui Zhang, Torben Schneider, Claudia A Wheeler-Kingshott, and Daniel C Alexander. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*, 61(4):1000–1016, 2012.

[79] Nicolas Kunz, Hui Zhang, Lana Vasung, Kieran R O'brien, Yaniv Assaf, François Lazeyras, Daniel C Alexander, and Petra S Hüppi. Assessing white matter microstructure of the newborn with multi-shell diffusion MRI and biophysical compartment models. *Neuroimage*, 96:288–299, 2014.

[80] Inge Timmers, Hui Zhang, Matteo Bastiani, Bernadette M Jansma, Alard Roebroeck, and M Estela Rubio-Gozalbo. White matter microstructure pathology in classic galactosemia revealed by neurite orientation dispersion and density imaging. *Journal of inherited metabolic disease*, 38(2):295–304, 2015.

[81] Zach Eaton-Rosen, Andrew Melbourne, Eliza Orasanu, M Jorge Cardoso, Marc Modat, Alan Bainbridge, Giles S Kendall, Nicola J Robertson, Neil Marlow, and Sebastien Ourselin. Longitudinal measurement of the developing grey matter in preterm subjects using multi-modal MRI. *NeuroImage*, 111:580–589, 2015.

[82] Gavin P Winston, Caroline Micallef, Mark R Symms, Daniel C Alexander, John S Duncan, and Hui Zhang. Advanced diffusion imaging sequences could aid assessing patients with focal cortical dysplasia and epilepsy. *Epilepsy research*, 108(2):336–339, 2014.

[83] Enrico Kaden, Thomas R Knösche, and Alfred Anwander. Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. *NeuroImage*, 37(2):474–488, 2007.

[84] Els Fieremans, Dmitry S Novikov, Jens H Jensen, and Joseph A Helpern. Monte Carlo study of a two-compartment exchange model of diffusion. *NMR in Biomedicine*, 23(7):711–724, 2010.

[85] J Kärger. Principles and applications of self-diffusion measurements by nuclear magnetic resonance. *Adv Magn Reson*, 12:1–89, 1988.

[86] Lorenza Brusini, Gloria Menegaz, and Markus Nilsson. Monte Carlo simulations of water exchange through myelin wraps: Implications for diffusion MRI. *IEEE transactions on medical imaging*, 38(6):1438–1445, 2019.

[87] Sune Nørhøj Jespersen, Michael Pedersen, and Hans Stødkilde-Jørgensen. The influence of a cellular size distribution on NMR diffusion measurements. *European Biophysics Journal*, 34(7):890–898, 2005.

[88] Ingrid Åslund, Agnieszka Nowacka, Markus Nilsson, and Daniel Topgaard. Filter-exchange PGSE NMR determination of cell membrane permeability. *Journal of Magnetic Resonance*, 200(2):291–295, 2009.

[89] Sarah L Codd and Paul T Callaghan. Spin echo analysis of restricted diffusion under generalized gradient waveforms: planar, cylindrical, and spherical pores with wall relaxivity. *Journal of Magnetic Resonance*, 137(2):358–372, 1999.

[90] Denis S Grebenkov. Exploring diffusion across permeable barriers at high gradients. II. Localization regime. *Journal of Magnetic Resonance*, 248:164–176, 2014.

[91] Samo Lasič, Markus Nilsson, Jimmy Lätt, Freddy Ståhlberg, and Daniel Topgaard. Apparent exchange rate mapping with diffusion MRI. *Magnetic resonance in medicine*, 66(2):356–365, 2011.

[92] Jimmy Lätt, Markus Nilsson, Danielle van Westen, Ronnie Wirestam, Freddy Ståhlberg, and Sara Brockstedt. Diffusion-weighted MRI measurements on stroke patients reveal water-exchange mechanisms in sub-acute ischaemic lesions. *NMR in biomedicine*, 22(6):619–628, 2009.

[93] Josef Pfeuffer, Ulrich Flögel, Wolfgang Dreher, and Dieter Leibfritz. Restricted diffusion and exchange of intracellular water: theoretical modelling and diffusion time dependence of 1H NMR measurements on perfused glial cells. *NMR in Biomedicine*, 11(1):19–31, 1998.

[94] Olaf Dietrich, Alexander Hubert, and Sabine Heiland. Imaging cell size and permeability in biological tissue using the diffusion-time dependence of the apparent diffusion coefficient. *Physics in medicine and biology*, 59(12):3081, 2014.

[95] Markus Nilsson, Jimmy Lätt, Danielle van Westen, Sara Brockstedt, Samo Lasič, Freddy Ståhlberg, and Daniel Topgaard. Noninvasive mapping of water diffusional exchange in the human brain using filter-exchange imaging. *Magnetic resonance in medicine*, 69(6):1572–1580, 2013.

[96] Dmitry S Novikov and Valerij G Kiselev. Effective medium theory of a diffusion-weighted signal. *NMR in Biomedicine*, 23(7):682–697, 2010.

[97] Samo Lasic, Sune Nørhøj Jespersen, Hans Magnus Henrik Lundell, Markus Nilsson, Tim B Dyrby, and Daniel Topgaard. Apparent exchange rate in multi-compartment anisotropic tissue. In *24th Annual meeting for International Soceity for Magnetic Resonance in Medicine*, 2016.

[98] Matt G Hall and Daniel C Alexander. Convergence and parameter choice for Monte-Carlo simulations of diffusion MRI. *IEEE transactions on medical imaging*, 28(9):1354–1364, 2009.

[99] Aurélien Géron. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. *Sebastopol, CA: O*, 2017.

[100] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[101] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[102] Tara N Sainath, Abdel-rahman Mohamed, Brian Kingsbury, and Bhuvana Ramabhadran. Deep convolutional neural networks for LVCSR. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 8614–8618. IEEE, 2013.

[103] Hui Y Xiong, Babak Alipanahi, Leo J Lee, Hannes Bretschneider, Daniele Merico, Ryan KC Yuen, Yimin Hua, Serge Gueroussov, Hamed S Najafabadi, Timothy R Hughes, et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218):1254806, 2015.

[104] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv preprint arXiv:1711.05225*, 2017.

[105] Stephan Dreiseitl, Lucila Ohno-Machado, Harald Kittler, Staal Vinterbo, Holger Billhardt, and Michael Binder. A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of biomedical informatics*, 34(1):28–36, 2001.

[106] Miles N Wernick, Yongyi Yang, Jovan G Brankov, Grigori Yourganov, and Stephen C Strother. Machine learning in medical imaging. *IEEE signal processing magazine*, 27(4):25–38, 2010.

[107] C Bishop. Pattern recognition and machine learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn. *Springer, New York*, 2007.

[108] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[109] Leo Breiman. *Classification and regression trees*. Routledge, 2017.

[110] Antonio Criminisi, Jamie Shotton, and Ender Konukoglu. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. *Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114*, 5(6), 2011.

[111] Antonio Criminisi, Jamie Shotton, Duncan P Robertson, and Ender Konukoglu. Regression forests for efficient anatomy detection and localization in CT studies. *MCV*, 2010:106–117, 2010.

[112] Ezequiel Geremia, Olivier Clatz, Bjoern H Menze, Ender Konukoglu, Antonio Criminisi, and Nicholas Ayache. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*, 57(2):378–390, 2011.

[113] Victor Lempitsky, Michael Verhoek, J Alison Noble, and Andrew Blake. Random forest classification for automatic delineation of myocardium in real-time 3d echocardiography. In *International Conference on Functional Imaging and Modeling of the Heart*, pages 447–456. Springer, 2009.

[114] Albert Montillo, Jamie Shotton, John Winn, Juan Eugenio Iglesias, Dimitri Metaxas, and Antonio Criminisi. Entangled decision forests and their application for semantic segmentation of CT images. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 184–196. Springer, 2011.

[115] Alexander Statnikov, Lily Wang, and Constantin F Aliferis. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC bioinformatics*, 9(1):319, 2008.

[116] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[117] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[118] Eduardo Valle, Michel Fornaciali, Afonso Menegola, Julia Tavares, Flávia Vasques Bittencourt, Lin Tzy Li, and Sandra Avila. Data, depth, and design: learning reliable models for melanoma screening. *arXiv preprint arXiv:1711.00441*, 2017.

[119] Geoffrey E Hinton et al. Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the cognitive science society*, volume 1, page 12. Amherst, MA, 1986.

[120] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[121] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[122] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[123] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. *arXiv preprint arXiv:1611.03530*, 2016.

[124] Jan Kukačka, Vladimir Golkov, and Daniel Cremers. Regularization for deep learning: A taxonomy. *arXiv preprint arXiv:1710.10686*, 2017.

[125] I Goodfellow, Y Bengio, and A Courville. Deep Learning http://www. deeplearn ingbook. org. *MIT Press, Cambridge, MA*, 2016.

[126] Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958, 2014.

[127] Zhanlong Yang, Geng Chen, Dinggang Shen, and Pew-Thian Yap. Robust fusion of diffusion MRI data for template construction. *Scientific reports*, 7(1):12950, 2017.

[128] Daniel C Alexander, Darko Zikic, Jiaying Zhang, Hui Zhang, and Antonio Criminisi. Image quality transfer via random forest regression: applications in diffusion MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 225–232. Springer, 2014.

[129] Ryutaro Tanno, Daniel E Worrall, Aurobrata Ghosh, Enrico Kaden, Stamatios N Sotiropoulos, Antonio Criminisi, and Daniel C Alexander. Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 611–619. Springer, 2017.

[130] Maxime Descoteaux, Rachid Deriche, Thomas R Knosche, and Alfred Anwander. Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE transactions on medical imaging*, 28(2):269–286, 2008.

[131] T Ganepola, Z Nagy, D Alexander, and M Sereno. An unsupervised group average cortical parcellation using HARDI data. In *An unsupervised group average cortical parcellation using HARDI data*, volume 2015, page 221. Organization for Human Brain Mapping, 2015.

[132] Gemma L Nedjati-Gilani, Torben Schneider, Matt G Hall, Niamh Cawley, Ioana Hill, Olga Ciccarelli, Ivana Drobnjak, Claudia AM Gandini Wheeler-Kingshott, and Daniel C Alexander. Machine learning based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150:119–135, 2017.

[133] Rutger HJ Fick, Neda Sepasian, Marco Pizzolato, Andrada Ianus, and Rachid Deriche. Assessing the feasibility of estimating axon diameter using diffusion models and machine learning. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 766–769. IEEE, 2017.

[134] Gaëtan Rensonnet, Benoît Scherrer, Gabriel Girard, Aleksandar Jankovski, Simon K Warfield, Benoît Macq, Jean-Philippe Thiran, and Maxime Taquet. Towards microstructure fingerprinting: Estimation of tissue properties from a dictionary of Monte Carlo diffusion MRI simulations. *NeuroImage*, 184:964–980, 2019.

[135] Thomas A Roberts, Ben Hipwell, Giulia Agliardi, Angela d'Esposito, Valerie Taylor, Mark F Lythgoe, and Simon Walker-Samuel. Deep learning diffusion fingerprinting to detect brain tumour response to chemotherapy. *bioRxiv*, page 193730, 2017.

[136] Marco Reisert, Elias Kellner, Bibek Dhital, Juergen Hennig, and Valerij G Kiselev. Disentangling micro from mesostructure by diffusion MRI: a Bayesian approach. *Neuroimage*, 147:964–975, 2017.

[137] Daniele Ravi, Nooshin Ghavami, Daniel C Alexander, and Andrada Ianus. Current applications and future promises of machine learning in diffusion MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 105–121. Springer, 2018.

[138] Ivana Drobnjak, Hui Zhang, Andrada Ianuş, Enrico Kaden, and Daniel C Alexander. PGSE, OGSE, and sensitivity to axon diameter in diffusion MRI: insight from a simulation study. *Magnetic resonance in medicine*, 75(2):688–700, 2016.

[139] Ivana Drobnjak, Bernard Siow, and Daniel C Alexander. Optimizing gradient waveforms for microstructure sensitivity in diffusion-weighted MR. *Journal of Magnetic Resonance*, 206(1):41–51, 2010.

[140] PA Cook, Y Bai, S Nedjati-Gilani, KK Seunarine, MG Hall, GJ Parker, and DC Alexander. Camino: open-source diffusion-MRI reconstruction and pro-

cessing. In *14th scientific meeting of the international society for magnetic resonance in medicine*, volume 2759. Seattle WA, USA, 2006.

[141] David G Regan and Philip W Kuchel. Mean residence time of molecules diffusing in a cell bounded by a semi-permeable membrane: Monte Carlo simulations and an expression relating membrane transition probability to permeability. *European Biophysics Journal*, 29(3):221–227, 2000.

[142] MG Hall, G Nedjati-Gilani, and DC Alexander. Monte-Carlo simulation of diffusion MRI with realistic voxel sizes. In *Proc Intl Soc Magn Reson Med*, volume 2463, 2014.

[143] G Haggqvist. Analyse der faserveteilung in einem ruckenmarkquerschnitt (TH3). *Z. mikrosk. anat. Forsch.*, 39:1–34, 1936.

[144] Mitsunori Makino, Kentaro Mimatsu, Haruhiko Saito, Nobuo Konishi, and Yoshio Hashizume. Morphometric study of myelinated fibers in human cervical spinal cord white matter. *Spine*, 21(9):1010–1016, 1996.

[145] Tanguy Duval, Ariane Saliani, Harris Nami, Antonio Nanci, Nikola Stikov, Hugues Leblond, and Julien Cohen-Adad. Axons morphometry in the human spinal cord. *NeuroImage*, 185:119–128, 2019.

[146] Kévin Ginsburger, Felix Matuschke, Fabrice Poupon, Jean-François Mangin, Markus Axer, and Cyril Poupon. MEDUSA: A GPU-based tool to create realistic phantoms of the brain microstructure using tiny spheres. *NeuroImage*, 193:10–24, 2019.

[147] Ross Callaghan, Daniel C Alexander, Hui Zhang, and Marco Palombo. Contextual fibre growth to generate realistic axonal packing for diffusion MRI simulation. In *International Conference on Information Processing in Medical Imaging*, pages 429–440. Springer, 2019.

[148] Kevin D Harkins and Mark D Does. Simulations on the influence of myelin water in diffusion-weighted imaging. *Physics in Medicine & Biology*, 61(13):4729, 2016.

[149] Paul T Callaghan. A simple matrix formalism for spin echo analysis of restricted diffusion under generalized gradient waveforms. *Journal of Magnetic Resonance*, 129(1):74–84, 1997.

[150] P Vangelderen, D DesPres, PCM Vanzijl, and CTW Moonen. Evaluation of restricted diffusion in cylinders. Phosphocreatine in rabbit leg muscle. *Journal of Magnetic Resonance, Series B*, 103(3):255–260, 1994.

[151] Stephen M Smith, Mark Jenkinson, Mark W Woolrich, Christian F Beckmann, Timothy EJ Behrens, Heidi Johansen-Berg, Peter R Bannister, Marilena De Luca, Ivana Drobnjak, David E Flitney, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23:S208–S219, 2004.

[152] Daniel C Alexander, Darko Zikic, Aurobrata Ghosh, Ryutaro Tanno, Viktor Wottschel, Jiaying Zhang, Enrico Kaden, Tim B Dyrby, Stamatios N Sotiropoulos, Hui Zhang, et al. Image quality transfer and applications in diffusion MRI. *Neuroimage*, 152:283–298, 2017.

[153] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[154] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

[155] GC DeLuca, GC Ebers, and MM Esiri. Axonal loss in multiple sclerosis: a pathological survey of the corticospinal and sensory tracts. *Brain*, 127(5):1009–1018, 2004.

[156] GC DeLuca, K Williams, N Evangelou, GC Ebers, and MM Esiri. The contribution of demyelination to axonal loss in multiple sclerosis. *Brain*, 129(6):1507–1516, 2006.

[157] Susie Y Huang, Sean M Tobyne, Aapo Nummenmaa, Thomas Witzel, Lawrence L Wald, Jennifer A McNab, and Eric C Klawiter. Characterization of axonal disease in patients with multiple sclerosis using high-gradient-diffusion MR imaging. *Radiology*, 280(1):244–251, 2016.

[158] Massimo Filippi, Wolfgang Brück, Declan Chard, Franz Fazekas, Jeroen JG Geurts, Christian Enzinger, Simon Hametner, Tanja Kuhlmann, Paolo Preziosa, Àlex Rovira, et al. Association between pathological and MRI findings in multiple sclerosis. *The Lancet Neurology*, 18(2):198–210, 2019.

[159] Markus Nilsson, Jimmy Lätt, Freddy Ståhlberg, Danielle van Westen, and Håkan Hagslätt. The importance of axonal undulation in diffusion MR measurements: a Monte Carlo simulation study. *NMR in Biomedicine*, 25(5):795–805, 2012.

[160] François Chollet et al. Keras. `https://keras.io`, 2015.

[161] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[162] Richard M Ransohoff. Animal models of multiple sclerosis: the good, the bad and the bottom line. *Nature neuroscience*, 15(8):1074, 2012.

[163] Ioana Hill, Marco Palombo, Mathieu Santin, Francesca Branzoli, Anne-Charlotte Philippe, Demian Wassermann, Marie-Stephane Aigrot, Bruno

Stankoff, Anne Baron-Van Evercooren, Mehdi Felfi, et al. Machine learning based white matter models with permeability: An experimental study in cuprizone treated in-vivo mouse model of axonal demyelination. *arXiv preprint arXiv:1907.02324*, 2019.

[164] Patryk Filipiak, Rutger Fick, Alexandra Petiet, Mathieu Santin, Anne-Charlotte Philippe, Stephane Lehericy, Philippe Ciuciu, Rachid Deriche, and Demian Wassermann. Reducing the number of samples in spatiotemporal dMRI acquisition design. *Magnetic resonance in medicine*, 81(5):3218–3233, 2019.

[165] Qi-Zhu Wu, Qing Yang, Holly S Cate, Dennis Kemper, Michele Binder, Hong-Xin Wang, Ke Fang, Melissa J Quick, Mark Marriott, Trevor J Kilpatrick, et al. MRI identification of the rostral-caudal pattern of pathology within the corpus callosum in the cuprizone mouse model. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(3):446–453, 2008.

[166] Els Fieremans, Lauren M Burcaw, Hong-Hsi Lee, Gregory Lemberskiy, Jelle Veraart, and Dmitry S Novikov. In vivo observation and biophysical interpretation of time-dependent diffusion in human white matter. *NeuroImage*, 129:414–427, 2016.

[167] Demian Wassermann, Mathieu Santin, Anne-Charlotte Philippe, Rutger Fick, Rachid Deriche, Stephane Lehericy, and Alexandra Petiet. Test-Retest qt-dMRI datasets for ”Non-Parametric GraphNet-Regularized Representation of dMRI in Space and Time”, September 2017.

[168] C-F Westin, Stephan E Maier, Hatsuho Mamata, Arya Nabavi, Ferenc A Jolesz, and Ron Kikinis. Processing and visualization for diffusion tensor MRI. *Medical image analysis*, 6(2):93–108, 2002.

[169] Susann Boretius, Angelika Escher, Tobias Dallenga, Claudia Wrzos, Roland Tammer, Wolfgang Brück, Stefan Nessler, Jens Frahm, and Christine Stadel-

mann. Assessment of lesion pathology in a new animal model of MS by multiparametric MRI and DTI. *Neuroimage*, 59(3):2678–2688, 2012.

[170] Sheng-Kwei Song, Jun Yoshino, Tuan Q Le, Shiow-Jiuan Lin, Shu-Wei Sun, Anne H Cross, and Regina C Armstrong. Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage*, 26(1):132–140, 2005.

[171] Jiangyang Zhang, Melina V Jones, Michael T McMahon, Susumu Mori, and Peter A Calabresi. In vivo and ex vivo diffusion tensor imaging of cuprizone-induced demyelination in the mouse corpus callosum. *Magnetic resonance in medicine*, 67(3):750–759, 2012.

[172] Peter R Mouton. *Principles and practices of unbiased stereology: an introduction for bioscientists*. JHU Press, 2002.

[173] Johannes Schindelin, Ignacio Arganda-Carreras, Erwin Frise, Verena Kaynig, Mark Longair, Tobias Pietzsch, Stephan Preibisch, Curtis Rueden, Stephan Saalfeld, Benjamin Schmid, et al. Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676, 2012.

[174] Nian Wang, Jieying Zhang, Gary Cofer, Yi Qi, Robert J Anderson, Leonard E White, and G Allan Johnson. Neurite orientation dispersion and density imaging of mouse brain microstructure. *Brain Structure and Function*, pages 1–17, 2019.

[175] Giorgio M Innocenti, Roberto Caminiti, and F Aboitiz. Comments on the paper by Horowitz et al.(2014). *Brain structure and function*, 220(3):1789–1790, 2015.

[176] Ariane Saliani, Blanche Perraud, Tanguy Duval, Nikola Stikov, Serge Rossignol, and Julien Cohen-Adad. Axon and myelin morphology in animal and human spinal cord. *Frontiers in neuroanatomy*, 11:129, 2017.

[177] Glenn K Matsushima and Pierre Morell. The neurotoxicant, cuprizone, as a model to study demyelination and remyelination in the central nervous system. *Brain pathology*, 11(1):107–116, 2001.

[178] Matthew D Budde and Joseph A Frank. Neurite beading is sufficient to decrease the apparent diffusion coefficient after ischemic stroke. *Proceedings of the National Academy of Sciences*, 107(32):14472–14477, 2010.

[179] Marco Palombo, Clemence Ligneul, Edwin Hernandez-Garzon, and Julien Valette. Can we detect the effect of spines and leaflets on the diffusion of brain intracellular metabolites? *NeuroImage*, 182:283–293, 2018.

[180] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.

[181] Tuan Anh Le, Atilim Giineş Baydin, Robert Zinkov, and Frank Wood. Using synthetic data to train neural networks is model-based reasoning. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3514–3521. IEEE, 2017.

[182] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[183] Andreas Weigend. On overfitting and the effective number of hidden units. In *Proceedings of the 1993 connectionist models summer school*, volume 1, pages 335–342, 1994.

[184] J Quinlan and R Cameron-Jones. Oversearching and layered search in empirical learning. *breast cancer*, 286:2–7, 1995.

[185] Tom Dietterich. Overfitting and undercomputing in machine learning. *ACM computing surveys*, 27(3):326–327, 1995.

[186] Lars Kai Hansen and Peter Salamon. Neural network ensembles. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (10):993–1001, 1990.

[187] Zhi-Hua Zhou, Jianxin Wu, and Wei Tang. Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1-2):239–263, 2002.

[188] Devansh Arpit, Stanisław Jastrzebski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, et al. A closer look at memorization in deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 233–242. JMLR. org, 2017.

[189] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[190] Antonios Papaioannou, Dmitry S Novikov, Els Fieremans, and Gregory S Boutis. Observation of structural universality in disordered systems using bulk diffusion measurement. *Physical Review E*, 96(6):061101, 2017.

[191] Els Fieremans and Hong-Hsi Lee. Physical and numerical phantoms for the validation of brain microstructural MRI: A cookbook. *Neuroimage*, 182:39–61, 2018.

[192] Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, and Shi-Fu Chen. Lung cancer cell identification based on artificial neural network ensembles. *Artificial Intelligence in Medicine*, 24(1):25–36, 2002.

[193] PáDraig Cunningham, John Carney, and Saji Jacob. Stability problems with artificial neural networks and the ensemble solution. *Artificial Intelligence in medicine*, 20(3):217–225, 2000.

[194] Ryutaro Tanno, Aurobrata Ghosh, Francesco Grussu, Enrico Kaden, Antonio Criminisi, and Daniel C Alexander. Bayesian image quality transfer. In *Inter-*

national Conference on Medical Image Computing and Computer-Assisted Intervention, pages 265–273. Springer, 2016.