Irritable Bowel Syndrome and Neuroimaging-based Biomarkers

Viola Helene Hansen

Department of Global Public Health and Primary Care &
Mohn Medical Imaging and Visualization Centre
University of Bergen

This thesis is submitted in partial fulfillment of the requirements for the degree of
Master of Science in Health Science, Study Track in Radiography

Spring 2020

Acknowledgement

First of all I would like to thank my main supervisor Professor Arvid Lundervold and my co-supervisor Dr. Birgitte Berentsen for giving me the oppurtunity to dive into such exciting scientific fields and also for the patience and support during my time as a master's student. I am indebted to Eivind A. Valestrand for the data collection used in this thesis and to Peder A.G. Lillebostad for the BIDS-structuring and general inspiration. I am thankful to the MMIV-team and the Brain-Gut team that I have learned a lot from, and I have greatly appreciated the meetings and social gatherings. Thanks to my family and friends for the encouragement and support through the writing of this thesis, especially Frida Brende Jenssen who firmly helped me through the process with patience and understanding. I would like to pay special regards to Professor Arvid Lundervold for his enthusiasm and invaluable assistance and guidance into the world of computational science and machine learning.

Abstract

Aims: To investigate the possible brain signatures in irritable bowel syndrome in terms of microstructural tissue changes assessed with fractional anisotropy (FA) measurements in targeted brain regions at group and subject levels.

Methods and material: Anatomical T1- and diffusion weighted brain MRI from 15 IBS patients (38.6 ± 12.4 years) and 15 HC (35.8 ± 12.4 years) were used in the analysis. Image segmentation and extraction of FA values from target regions insula, thalamus and pallidum for region wise analysis at a group level was conducted using state-of-the-art software. Different machine learning models were used for subject level classification based on median FA values from the target regions.

Main results: Median FA values were lower in IBS group compared to HC group for all three bilateral target regions, but statistical significance was found in bilateral insula and pallidum regions ($p<0.0001$). Highest accuracy, precision and recall score for subject classification obtained by machine learning was 64.6%, 64.8% and 69.5% respectively.

Conclusion: Results indicates a presence of white matter changes in target regions, and a small, but significant difference between IBS versus HC at group level in bilateral insula and pallidum regions. Machine learning models were moderately successful in subject level prediction of IBS using median FA values from the targeted brain regions.

Keywords: Irritable bowel syndrome, IBS, quantitative image biomarkers, magnetic resonance imaging, MRI, diffusion MRI, DTI, Artificial Intelligence, AI, machine learning

Sammendrag

Formål: Å undersøke mulige "hjerne-signaturer" ved irritabel tarm-syndrom (IBS) knyttet til mikrostrukturelle vevsendringer, uttrykt ved fraksjonell anisotropi (FA)-verdier, i spesifikke hjerneregioner på gruppe- og individnivå.

Metode og materiale: Anatomiske T1- og diffusjonsvektede MR-bilder av hjernen fra 15 IBS pasienter (38.6 ± 12.4 år) og 15 friske frivillige (HC) (35.8 ± 12.4 år) ble brukt i analysen. Bildesegmentering og ekstraksjon av FA-verdier fra målregionene på gruppenivå ble utført ved bruk av "state-of-the-art" programvare. Ulike maskinlæringsmodeller ble brukt for klassifisering på individnivå basert på median FA-verdier fra måleregionene.

Hovedresultat: Median FA-verdier var lavere hos IBS pasienter sammenlignet med HC i de tre spesifikke bilaterale målregionene, men statistisk signifikans ble funnet i bilaterale insula- og pallidumregioner ($p<0.0001$). Høyest nøyaktighetsscore (accuracy score), presisjonsscore (precision score) og tilbakekalling (recall score) i klassifikasjon på individnivå oppnådd av maskinlæringsmodellene var henholdsvis 64.6%, 64.8% og 69.5%.

Konklusjon: Indikasjoner på endringer i hvit substans i målområdene, og en liten, men signifikant forskjell mellom IBS og HC på gruppenivå i bilaterale insula- og pallidumregioner. Maskinlæringsmodellene klarte å klassifisere IBS ved bruk av median FA-verdier i målregionene med moderat suksess.

Nøkkelord: Irritabel tarm syndrom, IBS, bildebaserte biomarkører, magnetresonanstomografi, MR, diffusjonsavbildning, DTI, kunstig intelligens, KI, maskinlæring.

Table of contents

List of figures

List of tables

Selected abbreviations

ADC Apparent Diffusion Coefficient. 15

AI Artificial Intelligence. 27

ANN Artificial Neural Network. 28

BIDS Brain Imaging Data Structure. 18

CART Classification And Regression Tree. 49

DGBI Disorders of Gut Brain Interaction. 9

DIPY Diffusion Imaging in Python. 44

DTI Diffusion Tensor Imaging. 16

DWI Diffusion Weighted Imaging. 14

FA Fractional Anisotropy. 17

FGID Functional Gastrointestinal Disorder. 9

HPA Hypothalamic-Pituitary-Adrenal. 11

IBS Irritable Bowel Syndrome. 9

LUT Look-Up Table. 46

MD Mean Diffusivity. 17

ML Machine Learning. 27

MRI Magnetic Resonance Imaging. 13

QI Quantitative Imaging. 19

QIB Quantitative Imaging Biomarker. 19

QOL Quality of Life. 10

RF Random Forest. 13, 49

ROI Region Of Interest. 22

Part I

Introduction

Chapter 1

The interdisciplinary context of the work

With a focus on brain imaging this thesis addresses the brain-gut axis (Fig. 1.1), the irritable bowel syndrome (IBS) being a prevalent functional gastrointestinal disorder world wide, and the possible brain signatures of this disease in terms of microstructural tissue changes detectable by non-invasive magnetic resonance imaging techniques combined with advanced image analysis, tissue segmentation, and machine learning.

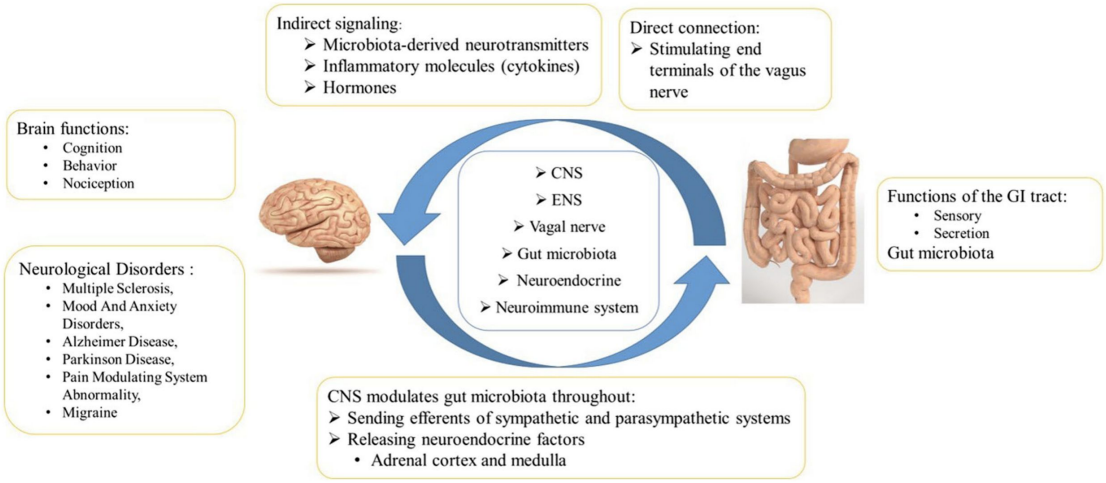

Fig. 1.1: The brain-gut axis, its many interconnected pathways, and its role in many disease processes including the irritable bowel syndrome (IBS). Adapted from Arzani et al. [2020] distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0)

The thesis, which from my perspective can be characterized as *integrative, quantitative radiography*, incorporates several disciplines and technologies. It is divided into a *Theory part* (Part II) and an *Experimental part* (Part III). In Part II the theoretical background for project is described. Initially, the irritable bowel syndrome (IBS) is presented (Chapter 2), including its clinical expression and the diagnostic criteria for the disease, its surprisingly high prevalence and impact on society, and possible disease mechanisms related to the microbial communities in the gut and the brain-gut axis, i.e. microbiota-brain-gut-axis. Next, I provide a short description of the principles of magnetic resonance imaging (Chapter 3) and the different MRI modalities of which diffusion MRI is the most important in my thesis. I also touch upon functional MRI as it is highly relevant to the IBS disease, and finally the increasingly popular Brain Imaging Data Structure (`BIDS`) used to represent all the imaging, demographic and questionnaire data in a compact and efficient way, suitable for comprehensive data analysis and machine learning. In Chapter 4 I introduce the concepts of *quantitative imaging* and *imaging biomarkers*, influencing the experimental part of the thesis. In relation to imaging biomarkers I also introduce *image segmentation* as an important component for obtaining region of interest (ROI) derived biomarkers. In my experimental work I have used *fractional anisotropy* (FA, a unit-less value between 0 and 1) derived from diffusion MRI as a quantitative biomarker, often interpreted as a proxy for "*tissue microstructure*". FA is explained in Section 3.2 and put into the context of imaging biomarker in Section 4.2. Based on literature and previous studies, IBS seems to be related to specific networks and regions in the brain, e.g. the salience network and the insula. In Section 4.3 the concepts of *brain connectivity* and *brain networks* are introduced. These are related to networks of physical (structural) connections between populations of neurons ("nodes") in the brain, revealed by diffusion MRI, on one side, and networks characterized by purely informational (functional) connectivity, revealed by functional MRI and pairwise inter-node similarity analysis of voxel time-courses.

A major methodological part of the thesis is related to *machine learning* (ML) and artificial intelligence - using ML models trained from data to predict IBS patients versus healthy controls based on their "brain signatures" (distribution of FA values in target regions). Chapter 5 provides a short introduction to AI and machine learning, including artificial neural networks (in contrast to biological neural networks previously explained) and deep learning that seems to introduce a new era in medical imaging and also in radiography. Finally, in Chapter 6, the brain networks and structures relevant in IBS are presented, and the *aims and hypotheses for the experimental part* are stated

in Chapter 7. This connects and motivates the wide range of topics described in the theory part of my thesis with the more focused research questions and study conducted in the experimental part.

A note on Radiographers and Artificial Intelligence

> *"Adoption of AI in medical imaging and radiation therapy requires radiographers and radiological technologists to adapt their imaging and treatment practices to ensure new technology is being implemented, used and regulated appropriately, based on high quality research evidence, maximising benefits to their patient".* [ISRRT and EFRS, 2020].

This quote from the joint statement of International Society of Radiographers and Radiological Technologists (ISRRT) and European Federation of Radiographer Societies (EFRS) shows the European and international attention towards artificial intelligence for radiographers in the field of medical imaging.

Some additions for curriculum development in the European Qualifications Framework (EQF) for radiographers are suggested by ISRRT and EFRS. It is suggested that radiographers should be involved in both piloting and research of algorithms prior to clinical implementation, and also understand how the algorithms arrive at the decisions. This involves to understand the probability errors within these decisions to be able to effectively communicate the finding to the patients. Radiographers should also *"embrace, adopt and adapt technology, ensuring that practice is evidence based and based on the patient"* [ISRRT and EFRS, 2020].

The conclusion from ISRRT and EFRS states that an active role in planning, development, implementation, use and validations of AI application must be assumed by radiographers. *"The optimal integration of AI into medical safety, clinical imaging and radiation therapy can only be achieved through appropriate education of the current and future workforce and the active engagement of radiographers and radiologic technologists in AI advancements going forwards"* [ISRRT and EFRS, 2020].

Most of the experimental work in this thesis has been conducted with the use of `Jupyter Notebooks`, an interactive environment for the development and testing of scripts in the Python programming language. All code developed in the project is stored and maintained on `GitHub` (`https://github.com/arvidl/viola-ibs-imaging`), being the major source code repository for software development and distribution world wide, supporting *open science* and *reproducible research*. In the experimental part of the thesis I have made use of this code and modified the notebooks according to my needs in collaboration with my main supervisor.

Part II

Theoretical Background

Chapter 2

Irritable bowel syndrome - IBS

2.1 Definitions and criteria

Irritable Bowel Syndrome (IBS) is a common medical disorder across the world and can be defined as "chronic functional disorder of the gastrointestinal system" [Canavan et al., 2014a]. IBS is thought to affect between 10-20% of the population globally, depending on the criteria used [Canavan et al., 2014b]. What has been termed "functional gastrointestinal disorders", commonly referred to as FGID are now also called *disorders of gut-brain interaction (DGBI)* [Mayer et al., 2019; Schmulson and Drossman, 2017]. As reported by Schmulson and Drossman [2017], the definition of DGBIs is "a group of disorders classified by GI symptoms related to any combination of motility disturbances, visceral hypersensitivity, altered mucosal and immune function, gut microbiota, and/or central nervous system processing". The symptoms of IBS are generally characterized by **abdominal pain**, a **change in bowel habits** with **constipation**, **diarrhea** or **both** [Canavan et al., 2014a]. According to Quigley [2018] it is considered to be the most common in adolescent and young adult females.

As a disorder IBS is not new, but the acceptance of IBS as a "real disorder" that can be potentially incapacitating is new. Previously, IBS patients could be dismissed as "neurotic" and was encouraged to "get their act together". The new recognition of the impact of IBS on a personal and societal level, research in this field have been progressed both scientifically and clinically [Quigley, 2018]. Traditionally, the diagnosis of IBS was set when no obvious structural or biochemical abnormalities were found. In recent years, however, researchers have suggested that there are distinct pathophysiological

disturbances that may account for the symptoms. Evidence that IBS is not likely to be one single disease or a psychiatric/somatosensory disorder is also emerging [Ford et al., 2017].

Patients with IBS are commonly subtyped by their predominant bowel habits; Diarrhea-predominant (**IBS-D**), constipation-predominant (**IBS-C**) and a mixture of both (**IBS-M**) [Weaver et al., 2016].

There has been different diagnosis criteria through the years, the two most recently being Rome III (2006-2016) and Rome IV (from May 2016). [Schmulson and Drossman, 2017]. The criteria from both Rome III and Rome IV are shown in the table below.

Rome III	**Rome IV**
Recurrent abdominal pain or discomfort *at least 3 days/month* in the last 3 months associated with two or more of the following criteria:	Recurrent abdominal pain on average *at least 1 day/week* in the last 3 months associated with two or more of the following criteria:
1. Improvement with defecation **2.** Onset associated with a change in frequency of stool **3.** Onset with a change in form (appearance) of stool	**1.** Related to defecation **2.** Associated with a change in frequency of stool **3.** Associated with a change in form (appearance) of stool

Table 2.1: Comparison of Rome III and Rome IV criteria. [Aziz et al., 2018]

2.2 Socioeconomic impact of IBS

Approximately 30% of people with IBS symptoms will consult physicians for their symptoms. Although there has not been observed increased mortality in patients with IBS, the quality of life (QOL) is greatly reduced. It is suggested that patients with IBS account for approximately 20% of the outpatient clinic time in gastroenterology globally [Canavan et al., 2014b]. A cross-sectional population-based survey conducted in Norway in 2001 shows that 51% of 30-year olds and 79% of 75- year olds with IBS had sought medical help for their symptoms. The same study also suggests that different comorbid symptoms and disorders also need to be considered when measuring

the impact of IBS on people's lives and on community costs. Some of these symptoms or disorders are fibromyalgia, chronic fatigue, musculoskeletal problems and mood disorders and were found to be two to three times higher in patients with IBS than people without IBS [Vandvik et al., 2006]. A substantial economic cost, both for patients and society can be related to IBS. This can include sick leave from work, hospitalization, medical consultations, pharmacological medications and diagnostic tests [Tack et al., 2019]. In the study by Drossman et al. [2009] with 1966 respondents reporting impaired health status and poor health-related quality of life (HRQoL) it is shown that the IBS patients are willing to give up 25% of their remaining life, averaging 15 years, to become symptom free.

2.3 Microbiota-brain-gut axis in IBS

Since IBS is a *disorder of gut-brain interaction* - ***DGBI***, there is an increasing understanding that the brain-gut axis plays an essential role in IBS. The brain-gut axis can be defined as "**a bidirectional channel of communication between the "big brain" (i.e. the central nervous system - CNS) in the cranium and the "little brain" (i.e. the enteric nervous system - ENS)**" [Quigley, 2018] (see also Fig. 1.1). The role of the brain-gut axis is to both monitor and integrate gut functions as well as to link cognitive and emotional centers of the brain with peripheral intestinal functions and mechanisms. These functions and mechanisms can be immune activation, intestinal permeability, enteric reflexes, and enteroendocrine signaling [Carabotti et al., 2015]. The afferent visceral signals travels from the gut to the brain through spinal and vagal pathways, while the efferent signals are conveyed mainly through the autonomic nervous system and hypothalamic-pituitary-adrenal (HPA) axis. [Kano et al., 2018] The concept of the brain-gut axis was first introduced to describe the central and peripheral effects of gut-brain peptides such as cholecystokinin and bombesin which was implicated in disorders like anorexia nervosa. With further description of the function and morphology of the enteric nervous system, it became clear that it shares many features other than the certain peptides with the central nervous system (CNS) [Quigley, 2018].

According to Mukhtar et al. [2019], the bidirectional communication occurs through a number of different neuronal pathways. The communication is modified by different

anatomical and environmental factors such as the endocrine system, the autonomic nervous system (ANS), the limbic system, and the HPA axis.

The **gut microbiota** consists of "*a diverse population of mainly prokaryotes that are expected to have a symbiotic relationship with the human host*" [Marques et al., 2016]. The main body of these prokaryotes are anaerobic and they belong to a few abundant **bacterial phyla**: *Firmicutes*, *Actinobacteria*, *Proteobacteria*, *Bacteroidetes* and *Verrucomicrobiota*. Over 90% of the human gut microbiota is constituted by the *Firmicutes* and *Bacteriodetes* phyla that accomodate the most abundant species. The gut is almost sterile at birth, but during the first days of life the gut of a human being is rapidly colonized and this is affected by different factors such as antibiotic therapy, mode of delivery and type of feeding. The **microbiome** of the gut is relatively stable in an adult human, even though it develops rapidly in both complexity and diversity the few years of life. Despite the relatively stable microbiome, it still may be modified by different factors such as pharmaceutical treatments that target the activity, stability and composition of the microbiota, food components and major dietary changes [Marques et al., 2016].

Recently, the gut microbiota has gained increasing recognition in influencing key physiological processes which also includes the gastrointestinal function [Mungovan and Ratcliffe, 2016]. Therefore the microbiota is considered a key player in the brain-gut axis, i.e **the microbiota-brain-gut axis** [Mukhtar et al., 2019]. According to Quigley [2018], microbiota was first suggested to be a factor in IBS by the observation of IBS-development anew after an acute enteric viral, bacterial or parasitic infection.

As stated in Marques et al. [2016] it is now generally accepted that *a dysregulation along the microbiota-brain-gut axis is present in IBS*, which is shown by the high prevalence of psychological comorbidities and an increased visceral hypersensitivity in IBS patients. These psychological comorbidities may be anxiety and depression among others [Marques et al., 2016].

Chapter 3

Magnetic Resonance Imaging - MRI

Magnetic Resonance Imaging (MRI) [1] is a non-invasive medical imaging technique that uses powerful magnets to produce a strong magnetic field, usually between 1.5 and 7 Tesla (T). MRI exploits the abundance of hydrogen atoms in the human body and their nuclear spins. The hydrogen atoms are primarily found in water and lipids which the body contains a lot of. The external static magnetic field, called B_0, forces the hydrogen protons in the body to align with the magnetic field. The intrinsic angular momentum, commonly called spin, of the hydrogen nucleus precesses at the *Larmor frequency* when in a magnetic field. The Larmor equation states that the precession frequency of the nuclear magnetic moment is directly proportional to the product of the gyromagnetic ratio and the strength of the magnetic field which is shown by the following equation:

$$\omega_0 = \gamma B_0$$

The gyromagnetic ratio γ is a constant which is 42.58 MHz/T for the hydrogen nucleus. A short radio frequency (RF) pulse at the Larmor frequency is applied to excite the hydrogen nuclei, which in turn results in a rotation of the net magnetization away from B_0; the excited protons will absorb the energy from the RF pulse and resonate. The magnetic field B_0 is in the longitudinal direction of the MRI, the z-direction, and the magnetization vectors corresponds to the planes of the MRI. The longitudinal

[1] A detailed description of MRI, especially the physics behind it, is beyond the scope of this work and therefore this chapter will only deal with the matter in the most superficial way.

direction has the magnetization vector M_z and the transverse direction's magnetization vector M_{xy}. After the excitement of the hydrogen nuclei by the RF pulse, the protons return to equilibrium with a rate determined by the time constant, T1. This is also called longitudinal relaxation or spin-lattice relaxation. While returning to equilibrium, the protons release the energy received from the RF pulse. The relaxation process gives an electromagnetic signal which can be read by a receiver coil. There is also a transverse relaxation, T2 relaxation, that happens simultaneously as T1 and describes the dephasing of protons that occur in the transverse M_{xy} plane [Brown et al., 2014].

3.1 Multimodal Magnetic Resonance Imaging

Multimodal imaging is a methodological approach where several acquisition methods and physical principles are applied together to provide complimentary information on both the structure and the function of tissue or organ of interest, in our case the human brain [Mayer et al., 2019]. The term **multimodal MRI** is commonly used when combining structural or anatomical MRI (aMRI), functional MRI (fMRI), diffusion MRI (dMRI) and in some cases perfusion imaging (pMRI) or magnetic resonance spectroscopy (MRS), or even spectroscopic MR imaging (MRSI). By incorporating multiple informative MRI modalities and perform joint analysis (e.g. by co-registration), one gets more information to utilize than by a single MRI modality alone, or by analysing the various acquisitions separately Hao et al. [2013].

3.2 Diffusion MRI - dMRI

It has long been known that MR has the capability to both image and measure molecular (water) diffusion. This principle has a wide range of applications, from cancer and stroke to the assessment of white matter microstructure and reconstruction of fiber tracts in the living brain.

3.2.1 Diffusion weighted image acquisition

The image contrast in diffusion weighted imaging (DWI, or dMRI) of the brain is based on the differences in the magnitude and directional preferences of water diffusion

between voxels or tissue regions. Diffusion represents Brownian motion, the random thermal movement (displacement) of molecules [Douek et al., 1991; Huisman, 2010]. Diffusion MRI measures the dephasing of proton spins in the presence of a spatially-varying magnetic field gradients defined by the diffusion weighted pulse sequence [Winston, 2012]. The first description of DWI sequences was done as early as 1965 by Stejskal and Tanner [Stejskal and Tanner, 1965], but was not in use in the clinical routines until the mid-1980s because of the hardware and software requirements [Huisman, 2010].

In a DW image of the brain, areas such as the ventricles have higher diffusivity because of the presence of cerebrospinal fluid (CSF) and this results in increased dephasing, signal loss, and a darker region in the image. The ventricles are also examples of *isotropic* diffusion areas. Isotropic areas have similar diffusion properties in every direction, and the recorded signal is not dependent on the applied direction of the gradient used for the diffusion sensitization. Areas where the structure of the tissue favors water movement in a particular direction (e.g. tissues consisting of elongated fiber structures) will be characterised by different diffusion coefficients in different directions. The signal attenuation is thus a reflection of the diffusion properties and local tissue organization in the direction of the applied diffusion sensitization gradient. Such tissue regions where water diffusion are direction-dependent are called *anisotropic* areas [Grusso and Wheeler-Kingshott, 2018]. Because of the anisotropy and the restriction of water diffusion in complex systems like living tissue, the term "*apparent diffusion coefficient*" (ADC) is more commonly used, because the value of the diffusion coefficient is very dependent on the choice of diffusion weighting, ultimately the *b-factor*, and the diffusion direction [Grusso and Wheeler-Kingshott, 2018]. To characterize (Gaussian) diffusion where displacements of water molecules per unit time are different in different directions, the diffusion tensor (DT) model is typically used [Jones, 2009] and is voxel-wise estimated (fitted) from the diffusion measurements in different directions (where at least 6 directions are necessary, see below).

3.2.2 Diffusion tensor reconstruction and derived metrics

The 3×3 matrix (D) shown below characterizes the *diffusion tensor* and water displacement along six intrinsic directions:
$D_{xx}, D_{yy}, D_{zz}, D_{xy}(= D_{yx}), D_{xz}(= D_{zx}), D_{yz}(= D_{zy})$, being the components in the symmetric D matrix.

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

Diffusion tensor imaging has been said to have revolutionized the field of white matter mapping, by taking advantage of macroscopic geometrical arrangement of the white matter fiber bundles in the brain [Assaf and Pasternak, 2008].

Diffusion tensor imaging (DTI) enables the analysis of the geometric three-dimensional shape of the diffusion tensor, determined by the components of D. In tissues like white matter in the brain, "*the apparent (scalar) diffusivity of water depends on the angle between the fiber-tract axis and the applied magnetic field gradient*" [Basser et al., 1994]. When the diffusion-sensitizing gradient is parallel to the direction of the fibers the apparent diffusivity is largest, and the apparent diffusivity is smallest when the diffusion-sensitizing gradient is perpendicular to the fibers [Basser et al., 1994; Douek et al., 1991]. Therefore, along the axon in the cytoskeleton there is a relatively large ADC, and perpendicular to the length of an axon the bi-lipid cell membranes limit diffusion, meaning the ADC is smaller [Winston, 2012]. In a simplified view, the ADC is the scalar of isotropic diffusion, and fractional anisotropy (FA) is the scalar of anisotropic diffusion.

The diffusion tensor of white- or gray matter should be considered mathematically and physically as a three-dimensional structure with three principal diffusivities, the *eigenvalues*, λ_1, λ_2 and λ_3, associated with three mutually perpendicular principal directions, the *eigenvectors*, denoted as $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$ and $\boldsymbol{\varepsilon}_3$ [Huisman, 2010]. These eigenvalues are sorted according to their size, i.e. $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and the components of the eigenvector corresponding to the largest eigenvalue, i.e. $\boldsymbol{\varepsilon}_1 = (\varepsilon_1^x, \varepsilon_1^y, \varepsilon_1^z)$ define the so-called *principal diffusion direction* in that particular voxel. The shape of the diffusion probabilities in the voxel, i.e. any configuration from fully isotropic to fully anisotropic diffusion, is determined by the size relationships between λ_1, λ_2 and λ_3.

Axial diffusivity (AD) describes the diffusing waters' mean diffusion coefficient parallel to the tract in the voxel of interest, or rate of diffusion along the diffusion main axis. Mathematically it is expressed as:

$$AD = \lambda_1$$

Radial diffusivity (RD) describes the diffusion rate in the transverse axis [Soares et al., 2013; Winklewski et al., 2018].

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

The average value of the diffusion coefficients over a voxel is the **mean diffusivity** (MD) and can be expressed as:

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

The **fractional anisotropy** (FA) value is diffusion characteristic often used in analysis of diffusion images [Alexander et al., 2007; Baliyan et al., 2016]. Mathematically it is expressed as:

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

One can describe FA ($0 \leq$FA≤ 1) as "*a normalised, dimensionless index that measures the properties of anisotropy of the diffusion tensor*" [Grusso and Wheeler-Kingshott, 2018], and we see that if $\lambda_1 = \lambda_2 = \lambda_3$, we have FA$= 0$ (isotropic diffusion). FA gives information on how noticeable the directional dependence of the diffusion process is. If the FA values are low, it implies that the diffusion along all directions are similar. High FA values implies a directional dependence such that the diffusion occurs preferentially along one dominant direction, i.e. $\boldsymbol{\varepsilon}_1$ [Grusso and Wheeler-Kingshott, 2018]. FA has become the most widely used index or metric in DTI brain research, because of its sensitivity to both the presence and integrity of white matter fibers [Assaf and Pasternak, 2008]. A high FA value in white matter is often interpreted as a more intact fiber organization, and FA measurements taken to reflect changes in tissue microstructure due to diabetes and gastrointestinal symptoms has been studied and discussed in Frøkjær et al. [2013]. In addition to FA, MD can also be used to characterize microstructural properties and the integrity of different brain structures, especially in white matter [Giannelli et al., 2010].

3.3 Functional Magnetic Resonance Imaging - fMRI

Even though this thesis does not focus on functional MRI, this technique has been used to investigate various DGBIs including IBS.

When neurons in the brain are activated, an increase of blood flow is delivered through capillaries close by these neurons. This makes an increase in the level of oxygenated blood in the area where the neurons are activated, resulting in a change in terms of the relative levels of oxyhemoglobin and deoxyhemoglobin. This change can be detected in MRI by exploiting the differential magnetic susceptibility [Han et al., 2018]. This is called the **Blood Oxygenation Level Dependent** (BOLD) contrast and is often used to provide *in vivo* real-time maps of blood oxygenation in the brain [2] [Ogawa et al., 1990]. Both task-based fMRI and resting state fMRI (rs-fMRI) have been extensively used for functional connectivity research by exploring the level of co-activation of rs-fMRI time-series between spatially distributed, but functionally linked brain regions [van den Heuvel and Hulshoff Pol, 2010].

3.4 The Brain Imaging Data Structure - BIDS

Open sharing of research data has been proposed as a way to address issues of reproducibility in neuroimaging studies [Borghi and Van Gulick, 2018; Koslow, 2000]. For sharing and reusing data, it initially requires a consensus on how to organize the data, so as to avoid misunderstandings and possible causes of errors [Gorgolewski et al., 2016]. The **Brain Imaging Data Structure** (BIDS) was therefore developed as a standard for the organization of neuroimaging and behavioral data, so researchers can ensure the data can be navigated, assessed and reproduced by collaborators and others [Borghi and Van Gulick, 2018; Gorgolewski et al., 2016]. The image files in BIDS are converted from the Digital Imaging and COmmunication in Medicine (**DICOM**) file format to Neuroimaging Informatics Technology Initiative (**NIfTI**) file format, because it is the largest commonality across different neuroimaging software. The metadata found in the DICOM files is stored in JavaScript Object Notation (**JSON**) file in BIDS, which has the same filename as the NIfTI file. Other metadata can be found in tab-separated value (**TSV**) files and comma-separated value (**CSV**) files.

[2] This simplified explanation serves first and foremost as a reminder.

Chapter 4

Quantitative Imaging and Image Biomarkers

Quantitative imaging (QI) is a term used to refer to both the extraction and the use of statistical/numerical features from medical images [Abramson et al., 2015]. The research field of QI includes the development, optimization, standardization and application of imaging acquisition protocols (anatomical, functional and molecular acquisition protocols). It also includes data analyses, reporting structures and display methods, and in addition it includes the validation of QI results against the relevant clinical and biological data [Abramson et al., 2015]. The Radiological Society of North America organizes the Quantitative Imaging Biomarkers Alliance (QIBA) which formally has defined what a **QI biomarker** is, and the definition is as follows:

"An objective characteristic derived from an in vivo image measured on a ratio or interval scale as indicators of normal biological processes, pathogenic processes, or a response to a therapeutic intervention" [Abramson et al., 2015; Kessler et al., 2014; Sullivan et al., 2015].

Quantitative image biomarkers (QIB) are based on mathematical definitions, and the calculation of QIB can be automated to enable high-throughput analyses. Image biomarkers characterise the contents of an image, often in regions of an image. The contents might be volume, mean intesity or cortical thickness. Often, the term *feature* is used instead of image biomarker, because of the historically close relationship to the field of computer vision [Zwanenburg et al., 2019]. To derive features from images, a sequence of *image processing* operations is required to be performed. Image processing

can be performed in a number of different ways, using a wide variety of processing schemes. However, the Image Biomarker Standardisation Initiative (IBSI) designed a general processing scheme for feature calculation that describes the different steps in image processing [Zwanenburg et al., 2019].

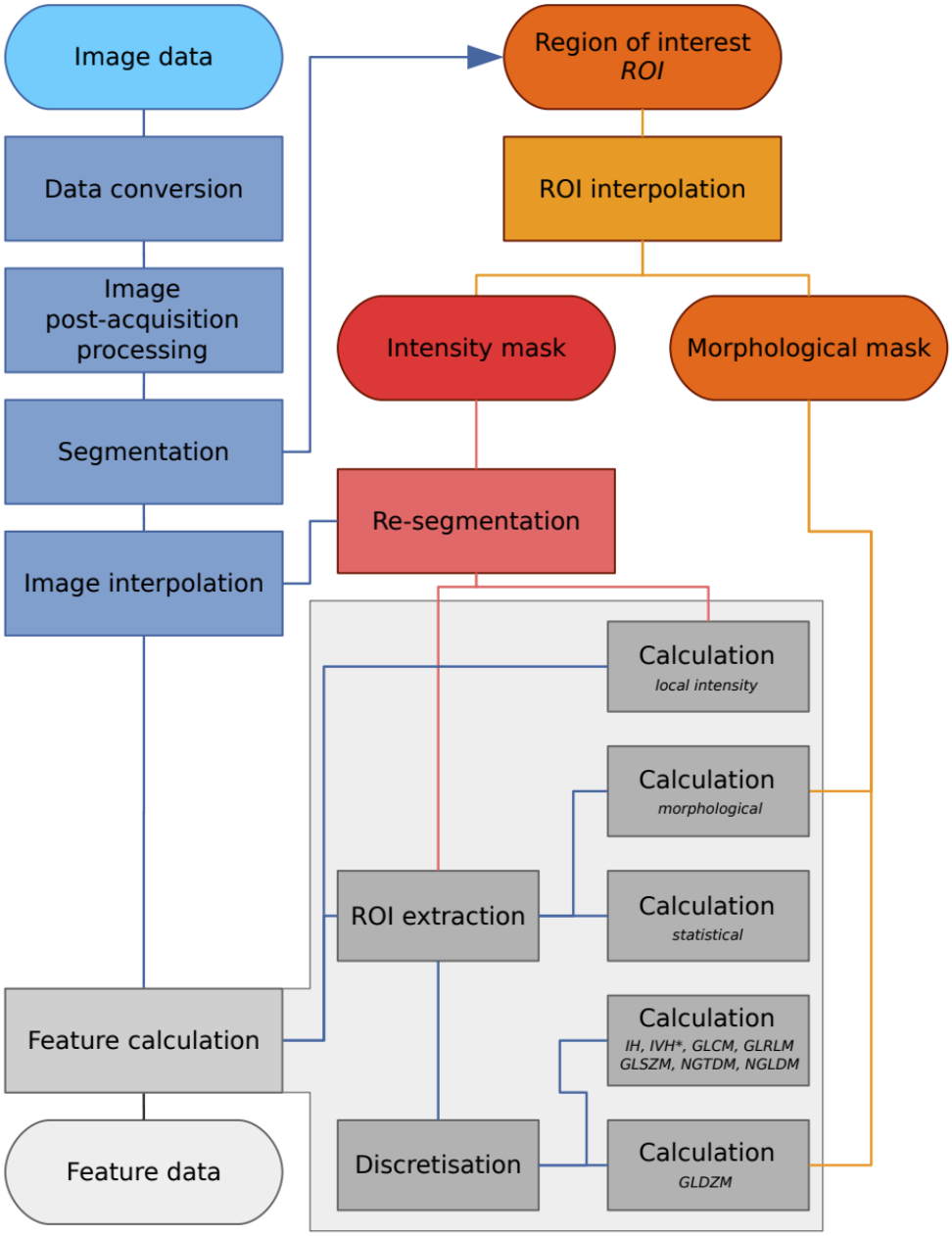

Fig. 4.1: General image processing scheme designed by IBSI [Zwanenburg et al., 2019] (Fig. 2.1). Creative Commons (CC BY 4.0) licensed.

For a quantitative imaging biomarker to be considered as such, it is recommended that the measurand must be on a ratio or interval scale. A **ratio variable** is a variable for which (i) there is a clear definition of zero and (ii) the ratio of two values can be meaningfully interpreted [Sullivan et al., 2015]. A "clear definition of zero" means in this case that zero indicates that no signal intensity is present from the feature being measured. **Interval variables** are measures (i) for which the difference between two values is meaningful, but the *ratio* of two values are not, and (ii) for which the scales do not have a meaningful zero. The centigrade Celsius and the Fahrenheit temperature scales are examples of scales that do not have a "meaningful zero", because a temperature measurement of zero does not mean that the entity has no heat energy [Sullivan et al., 2015]. In imaging, the Hounsfield scale used in context of X-ray computed tomography (CT) has no meaningful zero, because by definition the zero in Hounsfield scale is the density of water and therefore a substance that measure zero Hounsfield units have some density.

According to Abramson et al. [2015], there is potential for both increased precision and standardization in the interpretation of images by using QI, not only in research but also in clinical settings. Increased diagnostic accuracy and increased automation of data reporting are potential gains from the growth of QI, as well as decreased subjectivity and variability of the image analysis. A more robust association of imaging findings with other clinical and biological parameters and the large-scale attempts of linking phenotypic imaging patterns with the genomic profile are also a potential gain from the use of QI.

There are also challenges for developing QIBs and the key challenges according to Abramson et al. [2015] are:

- Analytical validation
- Qualification
- Utilization

Analytical validation involves "*demonstration of the accuracy, precision, and feasibility of biomarker measurement*". The analytical validation process includes the generation of data on limits of quantification, limits of detection and reference normal values, and also the inclusion of assessing reproducibility and repeatability. **Reproducibility** can be defined as "*the agreement between successive measurements*

made with varying conditions, such as location or operator" and the **repeatability** can be defined as "*the agreement between successive measurements made under the same conditions*". These latter two needs to be specified by the appropriate statistical parameters which can include the intraclass correlation coefficient (ICC) , kappa or weighted kappa, the confidence interval (CI) of the mean, or the more general Krippendorff's alpha.

Qualification involves "*demonstrating that a biomarker is associated with a clinical endpoint*". The goal of this process is to establish the ability of the QIB to serve as a measurable indicator of a pathologic process, biological process or response to an intervention, in accordance with the definition of a QIB.

Finally, **utilization** involves "*the assessment of biomarker performance in the specific context of its proposed use*". Depending on the clinical- or research setting, the requirements and performance threshold for a QI biomarker may be distinctly different. The utilization challenge also includes the practical issues of incorporating the biomarker into routine workflows, where the point is that the biomarker has to be extracted and reported both efficiently and at a reasonable cost for biomarker adoption and translation into standard clinical practice. (The previous definitions are stated according to [Abramson et al., 2015]).

4.1 The role of image segmentation

Segmentation of medical images is an important step in the search for image-based biomarkers and in quantitative image analysis in general. The goal of image segmentation is to *partition an image into a set of regions of similar attributes*. The regions are homogeneous and not overlapping each other, and examples of the attributes might be intensity in the image, texture, or shape [Despotović et al., 2015; Lundervold and Lundervold, 2019]. Image analysis in the feature-based paradigm relies on the definition of region of interest (ROI), because the ROI is used to define the region where the features are calculated. Depending on the objective of a study, the ROI can be e.g. a tumor volume or an anatomical region. An ROI can be defined manually by an expert or (semi-) automatically using algorithms [Zwanenburg et al., 2019].

4.2 FA as a quantitative imaging biomarker

Fractional anisotropy (FA), derived from dMRI measurements and defined previously, is a measure that is sensitive to many different tissue properties such as myelination, axonal density and ordering. It is, however, not very specific to any of these properties [Jones et al., 2013; Szczepankiewicz et al., 2015]. FA has been used as a marker in DTI studies of brain development, aging, neurological and neurodegenerative diseases (e.g. multiple sclerosis (MS) and amyotrophic lateral sclerosis (ALS)). FA has been found to be reduced in different neurodegenerative diseases and the reduction has been linked to axonal degeneration, breakdown of myelin and decreased white matter integrity [Timmers et al., 2016]. In several studies of Parkinson's disease (PD), the FA parameter has been used to distinguish between patients with PD and HC, where patients with PD exhibited consistently lower FA values in different anatomical brain regions [Andica et al., 2019]. In studies of patients with ALS and patients with Huntington's disease, FA was also able to distinguish between the patients and HCs [Li et al., 2012; Liu et al., 2016]. Flores-Alvarez et al. [2019] used FA as a biomarker regarding overall survival for patients with glioblastoma and concluded that in the immediate zone of peritumoral edema, FA depicted a significant association with the overall survival for the glioblastoma patients. Relevant to our study, changes in regional FA values due to diabetes and gastrointestinal symptoms has been reported and discussed in Frøkjær et al. [2013], as well as in patients with chronic pancreatitis [Frøkjær et al., 2011].

4.3 Brain connectivity - brain networks

The research field of *brain connectivity* has grown rapidly over the last decades. According to Pawela and Biswal [2011], the research of brain connectivity started with the man considered to be the founder of modern neuroscience; Santiago Ramón y Cajal and his detailed illustrations of the connections at the cellular level. Connectivity research nowadays is different from the previous static brain mapping, and is now concerned with the anatomical pathways, functional communication and interactions between the units of the CNS. With the use of MRI, the units focused on can not be on the micro scale (single neuron scale), because of the limited spatial and temporal resolution of current MRI technology. However, MRI and whole brain examinations

has demonstrated to be very useful at the macro (region) scale to reveal large-scale structural and functional connectivity in the living brain, both in health and disease.

The neurons in the brain has been considered to form complicated networks for a long time. An increasing number of studies, both empirical and theoretical, approach the human brain and its function from a network perspective [Sporns, 2013]. Technological advancements and developments, especially in noninvasive neuroimaging like MRI have given new opportunities for studying the structure and function of the human brain. Network measures have the potential to become objective diagnostic markers, as well as markers of effectiveness of either psychological or pharmaceutical therapies [Honey and Bullmore, 2004; Sporns, 2011].

4.3.1 Structural connectivity

To understand a complex system, like the brain, we need knowledge about the elementary components of the system and knowledge about how these components interact with each other is needed as well. Sporns [2013] defines **structural connectivity** as descriptions of "*anatomical connections linking a set of neural elements*". At the scale of the human brain, these connections generally refer to "*white matter projections linking cortical and subcortical regions.*" In human brains, this is usually measured as sets of undirected links. The structural networks of the human brain are also called the **human connectome**, and there have been several attempts to map these networks at the scale of brain regions. This have been done using diffusion MRI, more specifically diffusion tensor imaging [Bullmore and Sporns, 2009]. The measures from structural MRI (aMRI) can also reveal differences and variations in both volume and surface area of specific structures in the brain, and it can be used to infer structural connectivity. According to Sporns [2011], correlations in the volume or thickness of gray matter between two cortical brain areas have been shown to be associated with the presence of a fiber tract linking these areas. This is usually measured on brain data sets from multiple participants [Sporns, 2011].

4.3.2 Functional connectivity

Functional connectivity describes patterns of *statistical dependence* among neural elements generally derived from time series observations [Sporns, 2013]. In the world of

MRI, the time series data is derived with the technique of functional MRI (fMRI). The presence of a statistical relationship between neural elements is often viewed as a sign of functional coupling, but it is important to remember that this does not imply a causal coupling. Because of the high time dependency of functional connectivity, changes in a functional network can happen in a matter of tens or hundreds of milliseconds based on the continual modulation of sensory stimuli and task context, and the low sampling rate of fMRI (1-2 s per time frame) can not show this fast dynamics [Sporns, 2013]. However, slow dynamics representing e.g. resting state networks in the brain, can be captured by fMRI and with whole-brain coverage. Such (resting state) fMRI examinations of the IBS brain versus HCs performed before, during and after acute or long-term interventions (e.g. visceral pain provocation, psychotherapy, or low FODMAP diet) is an important methodology to study disorders of the gut-brain interaction such as IBS, but is outside the scope of the experimental work conducted and reported in this thesis due to time and resource constraints.

Chapter 5

Artificial Intelligence and Machine Learning

Artificial intelligence (AI) as a term is used when a device, typically a computer, mimics cognitive functions. These functions may be learning and problem solving [Pesapane et al., 2018]. Another definition is that "AI refers to a field of computer science dedicated to the creation of systems performing tasks that usually require human intelligence, branching off into different techniques" [Pesapane et al., 2018].

5.1 Machine learning

(ML) is a subfield of AI and includes approaches that allow computers to learn from experience. The experience in this case is are provided by training data, where the goal is to train the machine learning model make accurate predictions and from there be able to make generalized predictions on new and unseen data [Lundervold and Lundervold, 2019]. Machine learning is used in many different aspects in the modern society, amongst these are transcription of speech into text, predict activity of potential drug molecules, identify objects in images and reconstructing brain circuits [LeCun et al., 2015]. Lee et al. [2017] defines machine learning as "a set of methods that automatically detect patterns in data, and then utilize the uncovered patterns to predict future data or enable decision making under uncertain conditions" [Lee et al., 2017]. There are different forms of machine learning, often categorized according to how the models uses the input data in the training phase. The three most common described forms

are *reinforcement learning*, *unsupervised learning* and *supervised learning*, with the latter being the most common form of machine learning systems [LeCun et al., 2015; Lundervold and Lundervold, 2019; Raschka, 2015]. In *supervised learning* the computer is given a set of annotated/labeled data with the goal of producing correct labels on new and unseen sets of data based on "rules" discovered in the labeled data set [Lundervold and Lundervold, 2019]. A supervised learning task with discrete class labels is called a classification task. The goal of classification is to predict the categorical class labels for new, unseen data based on the past observations. A binary classification task is when a machine learning algorithm learns sets of rules to classify the new data into one class of only two possible classes. Distinguishing between spam and non-spam e-mail is an example of a binary classification task [Raschka, 2015].

5.2 Artificial Neural Networks

Artificial neural networks (ANN) was introduced already in the 1950s and is one of the most famous machine learning models [Lundervold and Lundervold, 2019]. Guresen and Kayakutlu [2011] points to the common character of ANN definitions in literature which is the comparison of ANNs with biological neural networks.

Biological Neural Networks	Artificial Neural Networks
Stimulus	Input
Receptors	Input Layer
Neural Net	Processing Layer
Neuron	Processing Element
Effectors	Output Layer
Response	Output and an entry

Table 5.1: Similarities between biological neural networks and ANNs. From Guresen and Kayakutlu [2011]

Neural networks are composed by a number of *neurons* - connected computational units, that are arranged in layers. The input layer is where the data enters the network, and this layer is followed by one or more hidden layers where the data is transformed before it ends up at the output layer where the neural network's predictions are produced [Lundervold and Lundervold, 2019]. The basic form of artificial neural networks are called *feedforward neural networks* and is described in detail in Goodfellow

et al. [2016], Chapt. 6. ANNs are often pictured as a network of nodes, as seen in the Fig. 5.1.

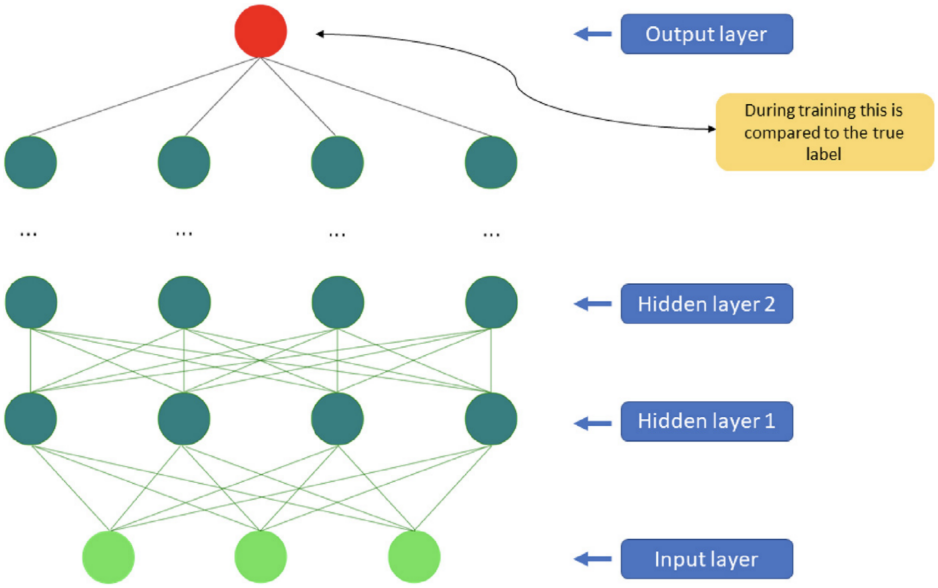

Fig. 5.1: Artificial neural network. Figure from Lundervold and Lundervold [2019] (Fig.1). Creative Commons (CC BY 4.0) licensed.

5.2.1 Deep Learning

Deep learning (DL) methods are *representation-learning methods*, which means a set of methods that allows a machine to use the raw data and automatically discover the representations needed for either detection or classification while bypassing the step of manually designing features from the raw data [LeCun et al., 2015; Lundervold and Lundervold, 2019]. Further, these representation learning methods have multiple levels of representation which are obtained by making simple, non-linear modules that alters the representation at one level, the first being the raw input, into another representation at a higher, more abstract level [LeCun et al., 2015]. The automatically learning of representations is called *feature learning* and it is the most common characteristic of deep learning methods, and also the main difference between deep learning and other "classic" machine learning approaches [Lundervold and Lundervold, 2019].

According to LeCun et al. [2015] deep learning is solving problems that has been major challenges for the AI community for years, and has proven especially good at

discovering structures of intricacy in high-dimensional data which therefore is used for not only science or research purposes, but also in domains of business and government.

DL in Medical Imaging - Convolutional Neural Networks

What triggered the interest of using deep learning in medical imaging was *convolutional neural networks* called "CNNs" or "ConvNets" [LeCun et al., 2015; Lundervold and Lundervold, 2019]. As stated in LeCun et al. [2015] CNNs "are designed to process data that come in the form of multiple arrays [...]. Many data modalities are in the form of multiple arrays: 1D for signals and sequences, including language; 2D for images or audio spectrograms; and 3D for video or volumetric images" [LeCun et al., 2015]. The CNN is aimed at preserving the spatial relationships in the data [Lundervold and Lundervold, 2019], which is important when CNN is used in medical imaging. According to LeCun et al. [2015] there are four key concepts behind CNNs that take advantage of the natural signals' properties: use of many layers, pooling, local connections and shared weights. As already mentioned, the input into a CNN is arranged in a grid structure, and this is then fed through the layers that will preserve the spatial relationships. Each of the layer operations is operating on a small region of the previous layer. The training of a CNN is typically done using backpropagation and gradient descent [Lundervold and Lundervold, 2019].

Architecture of a typical CNN

The architecture is structured as a series of stages, where the first few stages are made up of two types of layers; convolutional layers and pooling layers [LeCun et al., 2015]. Lundervold and Lundervold [2019] describe the architecture/building blocks of a CNN similarly; "A CNN has multiple layers of *convolutions* and *activations*, often interspersed with *pooling* layers". The primary purpose of a convolutional layer is to detect distinctive local motif-like edges, lines and other visual elements [Lee et al., 2017]. The *activation layers* are usually rectified linear unit (ReLU) layers, with the purpose of introducing non-linearity to a system that has been computing linear operations during the convolutional layers [Goodfellow et al., 2016; Nair, V. and Hinton, G., 2010]. ReLU's has achieved better results as activation layers, than the previously most used sigmoid and hyperbolic tangent functions [Pereira et al., 2016]. The feature maps produced by feeding data through the convolutional layer(s), are pooled in a *pooling*

layer. Small grid regions is used as input, and single numbers is produced for each region. This single number is usually computed by max-pooling or average pooling [Lundervold and Lundervold, 2019].

Batch normalization are layers usually put after activation layers to produce normalized activation maps. This is done by subtracting the mean and dividing by the standard deviation for each training batch. The reason for doing this is to make the network periodically change its activations to zero mean and unit standard deviation as the training batch reaches these layers. This works as a regularizer for the network, makes it less dependent on careful parameter initialization and also speeds up the training process [Lundervold and Lundervold, 2019]. It is also explained by Ioffe and Szegedy [2015], that normalizing layer inputs addresses the problem of internal covariate shift, which is explained as follows; The distribution of each layer's inputs changes during training, when the parameters of the previous layers change. This slows down the speed of training by requiring lower learning rates and meticulous parameter initialization. This again makes it hard to train models with saturating nonlinearities. Batch normalization makes it possible to use much higher learning rates and requires less carefulness in the initialization [Ioffe and Szegedy, 2015].

Dropout regularization is used to reduce overfitting, which can be a problem in deep neural networks. Dropout is a technique used to randomly drop units and their connections from the neural network during training to prevent too much co-adaption [Srivastava et al., 2014].

Chapter 6

Brain networks and structures relevant in IBS

Previous research in the field of IBS brain networks using neuroimaging modalities has shown some relevant regions and networks where there are deviations between patients with IBS and healthy controls. In the review article by Mayer et al. [2019] brain networks that has shown abnormalities are compiled and described. These networks are the default mode (DMN), central autonomic control, sensorimotor processing, salience detection, emotional arousal and central executive network [Mayer et al., 2019].

The *default mode network* includes the medial frontal cortex, posterior cingulate or retrosplenial cortex, precuneus, inferior parietal cortex, lateral temporal cortex and hippocampal formation and the functions of this network are monitoring internal thoughts, external goals and future planning, self-awareness and episodic memory. Studies has shown that the alterations this network in IBS patients are higher amygdala and dorsal anterior insula (INS) functional connectivities within DMN in hypersensitive IBS [Mayer et al., 2019].

The *central autonomic network* consists of control centres in the pontine-medulla, the central nucleus of the amygdala and several cortical regions (including the anterior INS, ACC and prefrontal and motor regions). This network has central control and modulation of the autonomic nervous system as well as regulation of endocrine, cardiovascular, respiratory and digestive activities during affective, cognitive, and motor tasks and sensations. Alterations in corticotropin releasing factor (CRF), CRF

receptor 1 and the norepinephrine-adrenergic receptor signalling system has been shown in IBS patiens [Mayer et al., 2019].

The thalamus, basal ganglia, sensorimotor cortex and posterior INS are the central brain regions in the *sensorimotor network*. The functions of this network are central processing and modulation of somatic and visceral sensory information. By using imaging modalities several alterations in IBS patients have been shown, amongst them are widespread microstructural white matter changes, greater grey matter in the posterior insula correlated with symptom duration and greater volume and cortical thickness, correlated with symptom severity in female IBS [Mayer et al., 2019].

The dorsal anterior cingulate cortex (ACC) and anterior INS comprise the *salience network*. The functions are to respond to subjective expectation or experience of either interoceptive or exteroceptive stimulus and coordination of the appropriate affective, behavioural, attentional and visceral responses to the aforementioned stimulus. Alterations in IBS patients are greater engagement of anterior INS and anterior midcingulate cortex in response to both expected and actual rectal distention, increased central, affective and emotional-arousal processes and enhanced visceral stimulus perception. In male IBS there is a greater pain-related response in the INS, while in female IBS the pain-related response in greater in the ACC [Mayer et al., 2019].

The *emotional arousal network* consists of amygdala, hypothalamus, hippocampus, posterior ACC and subgenual cingulate (sgACC). The functions of this network are being activated by real or perceived disruption of homeostastis and the generation of rapid feedback inhibition of amygdala, thereby limiting network activity in magnitude and duration and also the related activity in the central autonomic network [Mayer et al., 2019]. The emotional arousal network is an important link between the central autonomic network and the salience network to targets in the periphery such as immune system, the microbiota of the gut and the gastrointestinal tract in general [Mayer et al., 2015]. In female IBS there is seen an increased responsiveness to delivered and expected visceral stimuli and also greater emotional-arousal reactivity and altered connectivity. Alterations seen in IBS patients in general is a decrease in inhibitory feedback loop that also is seen in healthy controls who has lowered central serotonin levels by acute tryptophan depletion [Mayer et al., 2019].

The lateral prefrontal cortices and posterior parietal cortex are the brain regions that make up the *central executive network*. This network is activated during tasks that involved executive functions such as working memory, planning, attention and

response selection. According to Mayer et al. [2019] it is also often coactivated with regions of the salience network. Alterations in IBS patients are deficient activation of inhibitory cortical regions that are involved in downregulation of pain and emotion and also attention during experience and expectation of GI stimuli. Another alteration is a strong negative association between grey matter density and the cortical thickness of the dorsolateral profrontal cortex and pain catastrophising [Mayer et al., 2019].

6.1 DTI in IBS brain

Previous studies with DTI of the brain in patients with IBS and healthy controls (HCs) has shown differences in fractional anisotropy in specific areas of the brain. Ellingson et al. [2013] found that IBS patients had substantially lower FA values in parts of the basal ganglia, specifically the putamen and globus pallidus. The globus pallidus of IBS patients also showed significantly lower mean diffusity (MD) compared to HCs. Lower FA values was also observed in the thalamus, somatosensory regions (both primary and secondary) and motor regions. That patients with IBS has lower FA values in thalamic regions, basal ganglia and sensorimotor association and integration regions compared to HC is also supported by Labus et al. [2019]. Higher FA values was observed in IBS patients in prefrontal white matter regions.

In a study by Chen et al. [2011] of female patients with IBS, reduced FA was found in white matter regions associated with pain perception. Higher mean FA values was found in the fornix and the external capsule next to the right posterior insula.

6.2 Specific brain regions related to IBS

As the key node of the salience network, the insula is suggested responsible for the collection, integration and processing of interoceptive stimulus, which includes visceral pain perception, as well as cognitive and emotional pain modulation [Bednarska, 2019; Uddin et al., 2017]. As previously mentioned in section 6, the insula is a part of different networks that relates to IBS in addition to the salience network. The different anatomical parts of insula are a part of the sensorimotor, default mode and central autonomic network which indicates its importance and it has been reported as the most consistently activated brain region in studies using somatic pain stimuli [Mayer

et al., 2009]. Uddin [2015] also describes insular involvement in the different networks. Activation of the insula regions was seen across IBS groups in the quantitative meta-analysis by Tillisch et al. [2011]. Insula responses to painful rectal stimuli have also been shown in relation to IBS [Kano et al., 2020; Larsson et al., 2011; Yuan et al., 2003]. The pallidum is a part of the basal ganglia, a group of subcortical nuclei that lays at the base of the forebrain and top of the midbrain. The basal ganglia have been implicated in interpretation and modulation of both acute and chronic pain [Ellingson et al., 2013], as part of the sensorimotor network along with e.g. insula and thalamus [Mayer et al., 2019]. Pallidum is referred to as a region of interest in IBS and other pain related disorders [Gracely et al., 2002; Kano et al., 2020]. Thalamus has also been used to show differences between IBS and HC, and activation of thalamus regions in IBS patients has been found in several fMRI-studies [Larsson et al., 2011; Tillisch et al., 2011; Yuan et al., 2003].

Chapter 7

Research questions and hypotheses for the experimental work

This thesis aims to explore if and how the DTI- derived metric FA value can be used as a biomarker to detect group differences between subjects with IBS and HCs in targeted regions of the brain related to IBS. A secondary aim is to explore machine learning as a tool to predict whether a subject is IBS or HC based on brain DTI data.

7.1 Is tissue microstructure assessed with FA measurements in target brain regions different in patients with IBS versus healthy controls at a group level?

The research hypothesis is that it is possible to distinguish between IBS patients and HCs based on FA values derived from DTI data and using machine learning. This gives the following null hypothesis: $H_0 : \mathrm{IBS}_{\mathrm{FA}} = \mathrm{HC}_{\mathrm{FA}}$

7.2 How well (according accuracy, precision and recall) can we predict at a subject level an IBS brain versus a healthy control brain using FA-based brain signatures and machine learning models?

This is the other research question we address in this thesis, and is methodologically linked to *personalized medicine* and the machine learning technologies explained in Chapter 5. These research questions (7.1 and 7.2) are the subject for the following Experimental part of the thesis.

Part III

Experimental work

Chapter 8

Materials and methods

The materials was collected as part of a pilot project that led to a NFR FRIMEDBIO-funded project called "*Brain-Gut Microbiota Interaction in Irritable Bowel Syndrome: A multidimentional Approach*" (https://www.braingut.no). The work on this thesis has been done in collaboration with Mohn Medical Imaging and Visualization Centre (https://mmiv.no) and The Norwegian National Center for Functional Gastrointestinal Disorders, Department of Medicine, Haukeland University Hospital.

8.1 Participants

15 IBS patients (11 females and 4 males) clinically diagnosed with IBS according to the Rome III criteria by a gastroenterologist, and 15 HC (10 females and 5 males) participated in the study. The mean age of IBS patients was 38.6 $\pm$ 12.4 years, while the mean age of HC was 35.8 $\pm$ 13.2 years. No functional gastrointestinal disorder was present in the HCs. None of the participants reported any psychiatric illnesses. The study was approved by the regional ethics committee of Western Norway, and all the participants signed a written consent before participating in the study. All participants were fasting for at least four hours before the experiment was conducted. A plot of the participants' age and gender demographics can be seen in Fig. 8.1.

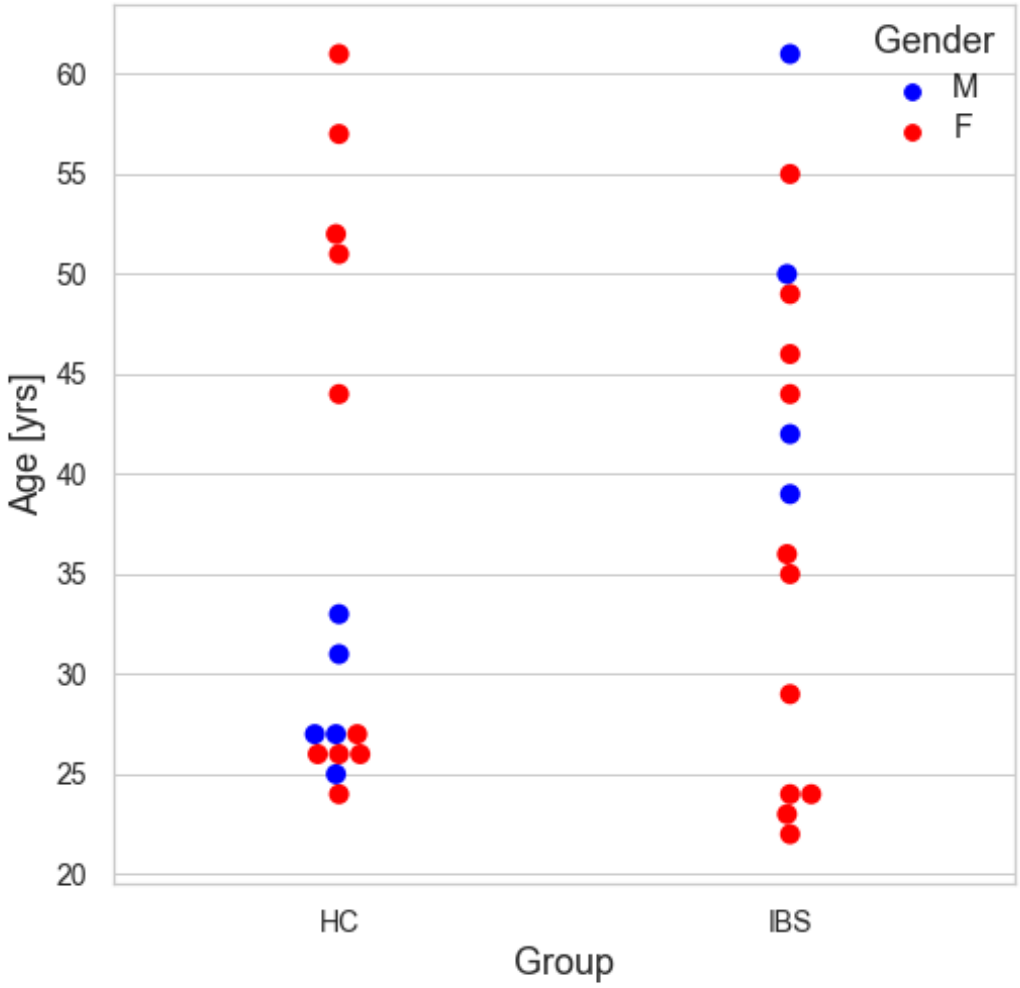

Fig. 8.1: Age and gender demographics of the 15 HCs and 15 IBS patients

8.2 MRI protocol

Of the more comprehensive original MRI protocol, the following two sequences are the focus in this thesis.

Brain MRI was acquired using a GE Signa HDxt 3T with a standard 8-channel head coil. First, a localizer was performed before a structural 3D T1-weighted (T1w) gradient-echo sequence was acquired. TR = 7.76 ms, TE = 2.95 ms, TI = 500 ms,

flip angle = 14°, matrix = 256 × 256, FOV = 256 × 256 mm^2 and 188 1.0 mm thick sagittal slices. The diffusion data was collected using a single-shot spin echo diffusion sensitized echo-planar imaging (EPI) sequence. TR = 14000 ms, TE = 92.9 ms, flip angle = 90°, matrix = 128 × 128; 50 axial slices; voxel size 1.72 × 1.72 × 2.4 mm^3. In total, 36 volumes was collected in this sequence; 6 volumes without diffusion weighting (b = 0 s/mm^2) and 30 volumes with diffusion weighting (b = 1000 s/mm^2).

8.3 File formats and file management

Since MRI scanners store the data in DICOM format, the images were converted to the NIfTI file format, which is used for scientific imaging analysis, visualization and processing. [Li et al., 2016]. The conversion to NIfTI file format was accompanied with a JSON file and was performed using the Python library `dcm2niix`. A "bug" in the software required an extra step for converting slice timing from milliseconds to seconds. The NIfTI files were organized into the Brain Imaging Data Structure (BIDS) directory structure for ease of automating processing pipelines [Gorgolewski et al., 2016]. The BIDS-validator available from `https://github.com/bids-standard/bids-validator` was used to make sure the formatting was correct.

8.4 Anatomical segmentation of the brain

As previously mentioned in section 4.1, anatomical segmentation is a key step for analysing medical images. Anatomical 3D T1w images was first processed through the FSL-based general pipeline, *fsl-anat*, that includes brain-extraction, tissue-type segmentation and subcortical structural segmentation. The images were reoriented to the standard (MNI) orientation and automatically cropped, before a bias-field correction was performed. Registration to standard space, brain-extraction, tissue-type segmentation and subcortical structure segmentation were then performed using the pipeline.

FreeSurfer v6.0.0 (FreeSurfer), a documented and freely available brain segmentation tool [Reuter et al., 2012], was also used for parcellation and segmentation of the brain. Two of FreeSurfers default atlases were used; the Desikan-Killiany-atlas [Desikan et al., 2006] and the Destrieux-atlas [Destrieux et al., 2010], where the former is "coarser" and

the latter is "finer", thereby analysis at two different levels of granularity is possible. In context of segmentation, an atlas is the combination of an image and a trusted reference segmentation, where the reference segmentations are generated by experts following an already established delineation protocol [Yaakub et al., 2020]. After running the subject data through FreeSurfer, the data needed to be converted from FreeSurfer space back to native space. The native space conversion was tested on one subject using the bias-corrected image from the FSL-pipeline as target image before the native space conversion was run on all subjects using the same method.

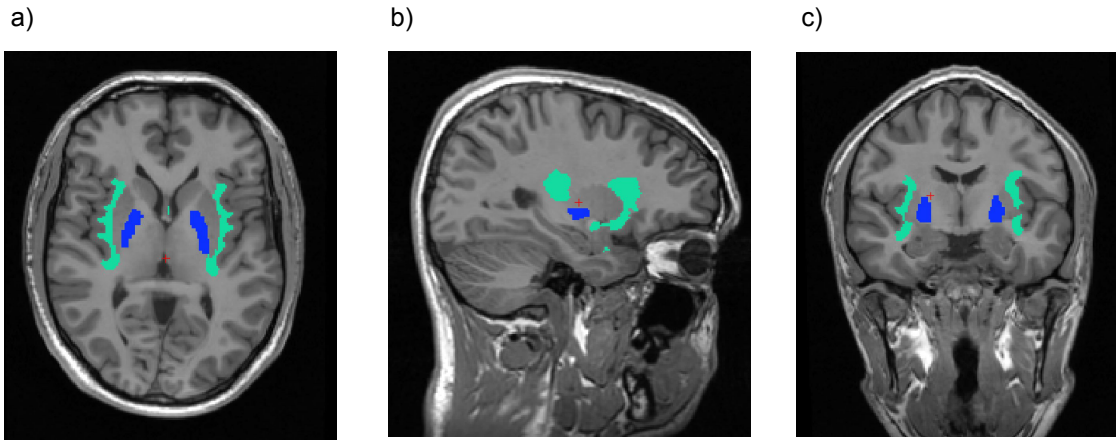

Fig. 8.2: FreeSurfer segmentation showing left and right insula (green) and left and right pallidum (blue) in one of the subjects. a) Axial view. b) Sagittal view. c) Coronal view. The segmented thalamus region is not shown in this image.

8.5 Diffusion Imaging analysis

DIPY, *Diffusion Imaging in Python*, a free and open source software that focuses on diffusion MRI was used on the diffusion weighted images. A NIfTI-file containing the diffusion weighted data and two text-files that contains the b-values (`bval`) and b-vectors (`bvec`) are necessary for the usage of DIPY.

A preliminary exploration of the first subject and session was done using DIPY, checking that the data corresponded with the dMRI protocol. The middle axial slice was visualized with and without diffusion weighting as seen in Fig. 8.3

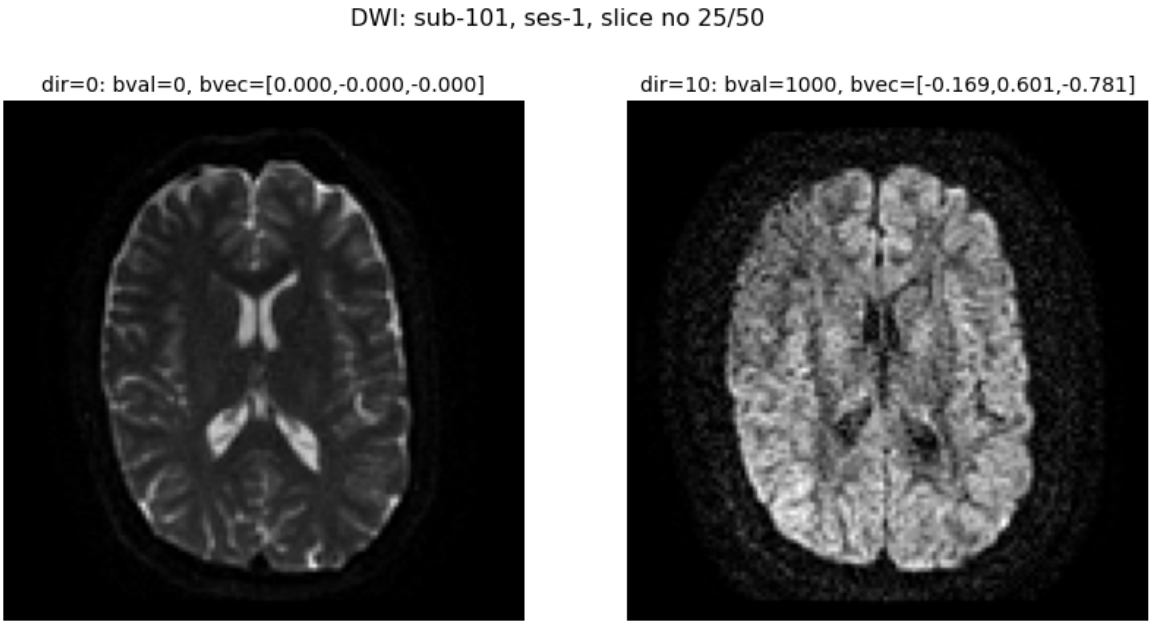

Fig. 8.3: First subject and session, slice 25/50, without (left) and with (right) diffusion weighting. Note the numbering of the diffusion-sensitizing direction (dir), the b-values (bval), and the gradient directions in scanner coordinate system (bvec).

The non-diffusion weighted data (S0) was then extracted from the subjects for later to be able to perform an affine registration in combination with the 3D T1w images. Image registration methods are widely used in medical imaging research and the basic task is to align two images, meaning to spatially reshape one image to match the other [Jenkinson, 2006]. In this case, the diffusion images was spatially reshaped to match the T1w images, to allow features from both sets of images to be utilized in the analysis. An affine registration function was made, where the mean S0 image was used as a target image. To transform the moving image toward the static image, each voxel in the image space (i, j, k) was mapped to world (scanner) coordinates (x, z, y). When applying the affine transform to the world coordinates, we got the (x', y', z') in world coordinates of the moving images. These coordinates were finally mapped to the voxel coordinates (i', j', k') in the moving image. The affine registration function was first tested on one subject, before running on all subjects.

8.5.1 Reconstruction of the diffusion tensor

The diffusion tensor model is based, as already mentioned in Chapter 3.2.2, on the work of Basser et al. [1994].

After a preliminary exploration of the first subject, *broadcasting* in NumPy was used to obtain the brain-masked 4D DWI data. *Broadcasting* is used to treat arrays of different shapes when doing arithmetic operations, since element-to-element operations are not possible if the dimensions of two arrays are dissimilar. DIPYs `TensorModel` was used and the fit method called to the mask data. The `TensorFit` object created by the fit method contains the fitting parameters used to generate tensor statistics. A function for the tensor features fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD) and a colormap (RGB) was defined. These tensor features were tested and then run on all subjects.

8.6 DTI-derived metrics in segmented ROIs

The co-registered white matter parcellation from the FreeSurfer segmentation for a given subject and session, lays the foundation for DTI feature extraction for a given anatomical region. Each region is given a unique number in FreeSurfer, which is found in the FreeSurfer look-up-table (LUT). The parameters for the function are the input directory, subject ID and session, ROI-name in the white matter parcelation and the corresponding ROI-number that is found in FreeSurfer LUT. This returns the measure names, in this case the list of DTI-derived features; FA, MD, AD and RD, the measure number which is the number of voxels in the 3D ROI region and measure value; the feature values of each feature i in the list of DTI-derived features. This function was tested on white matter left hemisphere insula (`wm-lh-insula`) on one subject first, before being run on all subjects and the chosen ROIs. Based on literature and previous research, the selected ROIs used in the analysis is given in Tab. 8.1 below.

Anatomical region	**Label number**	**Label name**
Insula (left hemisphere)	3035	`wm-lh-insula`
Insula (right hemisphere)	4035	`wm-rh-insula`
Thalamus (left hemisphere)	10	`Left-Thalamus-Proper`
Thalamus (right hemisphere)	49	`Right-Thalamus-Proper`
Pallidum (left hemisphere)	13	`Left-Pallidum`
Pallidum (right hemisphere)	52	`Right-Pallidum`

Table 8.1: Selected white matter regions with label number and label name from FreeSurfer LUT.

As previously mentioned, BIDS directory structure was used for organizing the neuroimaging files and this makes it possible to further utilize the Python package `pybids` for querying and navigating datasets. By using the tab-separated values-file (`.tsv`-file) of participants, the subjects was split into two groups; HC and IBS. This made it possible to make a *group-wise concatenation of ROI-based DTI-metrics*. The FA values and median FA values in the six different ROIs were obtained for the HC group first and then for the IBS group. The distribution of these FA values are shown in Figs. 9.1, 9.2 and 9.3.

8.7 Permutation testing of group differences

A permutation test is a statistical technique, and the idea was introduced in the 1930s by R.A. Fisher. As permutation tests are highly compute-intensive, the increasing power of modern computers now makes permutation tests practical to use routinely [Efron and Tibshirani, 1994]. The main application of permutation tests is to a *two-sample problem*, where two independent random samples, say z and y are drawn from possibly different probability distributions, e.g. IBS_{FA} and HC_{FA}. In our context the two-sample problem can be described as:

$$\text{IBS}_{\text{FA}} \rightarrow \mathbf{z} = \{z_1, z_2, \ldots, z_n\} \text{ and } \text{HC}_{\text{FA}} \rightarrow \mathbf{y} = \{y_1, y_2, \ldots, y_m\}$$

where we are interested in testing the null hypothesis $H_0 : \text{IBS}_{\text{FA}} = \text{HC}_{\text{FA}}$. If H_0 is true it means that there is no difference in the median FA values of the IBS group and the HC group. The permutation test is therefore useful to explore if it is likely that a slight difference between median FA values can happen by chance.

If the null hypothesis is true and there is no FA difference between HCs and IBS patients, both the group of HCs and the group of IBS patients can be considered to be samples from the same larger population. Therefore the FA values can be pooled together into one large group before shuffling and then split into two groups equal in size as before shuffling. Both groups now contain the same number of FA values as before the shuffle, but they are now a random mix of values from each original group. This was done to our data first, and the median was chosen as a desired return since the median FA values was already obtained. To get a large range of values and better statistics, the shuffling procedure was done 10000 times in the permutation test. By

using the inputs of the ROI name and ROI number according to the FreeSurfer LUT, and running 10000 permutations in a loop, the outputs of our function (implemented in Python) were: `theta_hat`, `diffCount` and `p_hat_perm`, for which

- `theta_hat` - the observed difference: *median*(HC_{FA}) - *median*(IBS_{FA}).
- `diffCount` - number of permutation occurrences for which a permuted difference in medians is larger than the observed difference in medians, and
- `p_hat_perm` - permutation test probability P of observing the difference *median*(HC_{FA}) - *median*(IBS_{FA}).

The results of the permutation tests for median FA differences in two groups and for each of the ROIs, compared with the observed difference in median FA between IBS_{FA} and HC_{FA} are shown in Chapter 9.

8.8 Machine learning

8.8.1 Feature extraction

The main goal of supervised learning is to use labeled training data to learn a model to make predictions about new, unseen data [Raschka, 2015]. To accomplish this, the data we have has to be split into a *training set* and a *test set*. The training set is used to train and optimize the machine learning model, while the test set is used to evaluate the model.

A group-wise subject extraction of ROI-based FA features was made. The selected features used in the model were the 50th percentile (median) FA value within each of the four ROIs: left and right insula regions and left and rigt pallidum regions in both HC-group and IBS-group, resulting in two separate data sets. These two data sets were then concatenated, giving 30 individuals or instances each assigned four features, i.e. a 30×4 data matrix, denoted $\boldsymbol{X}$. The corresponding 30×1 vector of labels or class belongings (IBS or HC) was denoted y.

8.8.2 Classification and performance measures

`Scikit-learn` open source machine learning library [Pedregosa et al., 2011] was primarily used for the supervised learning. The python package `XGBoost` (eXtreme Gradient Boosting) was also included in the machine learning process. Four different classification algorithms were trained and used for prediction, and three different performance metrics were applied in a k-**fold cross-validation** scheme, splitting the data into a training set and a test set in each fold. A mean performance score (metric) for each machine learning algorithm was accumulated across the folds, using three different metrics: **accuracy**, **precision**, and **recall**. Accuracy score is defined as

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions made}}$$

The mean precision scores for the four algorithms were also calculated to examine the precision of the predictions:

$$Precision = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

The mean recall score was calculated to examine the models' accuracy on the actual positive (IBS) class, i.e.

$$Recall = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

.

The algorithms used for training and prediction were

- **Logistic regression** (LG)
- **XGBoost** (XGB)
- **Decision Tree Classifier** (Classification and regression trees (CART))
- **Random Forest Classifier** (RF)

The `train_test_split` function from `scikit-learn` was used in the cross-validation loop, and the `test_size` was set to 0.25, meaning 25% of the samples were used for testing the algorithms, while 75% was used for training. We also applied *stratification* on the labels, making sure that the proportion of IBS instances and HC instances in

the original data set was kept balanced in the train and test splitting procedure. In our case, the randomized splitting procedure produced at each iteration a training set of 22 subjects, and test set of 8 subjects with about equal proportions of IBS and HCs. See Section 8.8.3 for implementation details. The four different algorithms were run N times (i.e. $k = N$ in the k-fold cross-validation), with different values of N's as seen in Tab. 8.2.

N (in N-fold cross-validation)					
10	50	70	100	150	200

Table 8.2: Number of times (N) the different algorithms were run with a random splitting of the data into a training set and a test set at each iteration.

We report results for $N = 100$ telling that the process of training and testing was executed 100 times for each algorithm to create the different models. Also, a **confusion matrix** was made for each machine learning algorithm to evaluate the performance and types of mis-classifications of the predictive algorithms, visualized in Fig. 9.16 and Fig. 9.17.

8.8.3 The Jupyter notebook used for classification

```
# https://github.com/arvidl/viola-ibs-imaging/blob/master/notebooks
# my3_12-dti-fa-feature-extraction-and-classification.ipynb  Viola Hansen / Arvid Lundervold

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from xgboost import XGBClassifier  # pip install xgboost
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score

a = [], a_p = [], a_r = []
b = [], b_p = [], b_r = []
c = [], c_p = [], c_r = []
d = [], d_p = [], d_r = []

N = 100
for i in range(N):
    X_train, X_test, y_train, y_test = train_test_split(X, yy, test_size=0.25, stratify=yy, random_state=i)
    if i == 0:
        print(f'Training data: {X_train.shape}\nTest data: {X_test.shape}')

    print(i, 'y_train:', y_train)

    lg = LogisticRegression(random_state=i)
    lg_fit = lg.fit(X_train, y_train)
    y_pred_lg = lg.predict(X_test)
    a = np.append(a, accuracy_score(y_test, y_pred_lg))
    a_p = np.append(a_p, precision_score(y_test, y_pred_lg))
    a_r = np.append(a_r, recall_score(y_test, y_pred_lg))

    xgb = XGBClassifier(random_state=i, base_score=0.5, objective='binary:logistic', booster = 'gbtree')
    xgb_fit = xgb.fit(X_train, y_train)
    y_pred_xgb = xgb.predict(X_test)
    b = np.append(b, accuracy_score(y_test, y_pred_xgb))
    b_p = np.append(b_p, precision_score(y_test, y_pred_xgb))
    b_r = np.append(b_r, recall_score(y_test, y_pred_xgb))

    cart = DecisionTreeClassifier(random_state=i)
    cart_fit = cart.fit(X_train, y_train)
    y_pred_cart = cart.predict(X_test)
    c = np.append(c, accuracy_score(y_test, y_pred_cart))
    c_p = np.append(c_p, precision_score(y_test, y_pred_cart))
    c_r = np.append(c_r, recall_score(y_test, y_pred_cart))

    rf = RandomForestClassifier(random_state=i)
    rf.fit(X_train, y_train)
    y_pred_rf = rf.predict(X_test)
    d = np.append(d,accuracy_score(y_test, y_pred_rf))
    d_p = np.append(d_p, precision_score(y_test, y_pred_rf))
    d_r = np.append(d_r, recall_score(y_test, y_pred_rf))

    print(i, 'y_test:', y_test)
    print(i, 'y_lg  :', list(y_pred_lg))
    print(i, 'y_xgb :', list(y_pred_xgb))
    print(i, 'y_cart:', list(y_pred_cart))
    print(i, 'y_rf  :', list(y_pred_rf))

print('\n Mean accuracy across n=%d train_test_split: \n LR (XGB) [CART] {RandomForest} classifications: %.3f (%.3f) [%.3f] {%.3f}'
   % (N, np.mean(a), np.mean(b), np.mean(c), np.mean(d)))
print('\n Mean precision across n=%d train_test_split: \n LR (XGB) [CART] {RandomForest} classifications: %.3f (%.3f) [%.3f] {%.3f}'
   % (N, np.mean(a_p), np.mean(b_p), np.mean(c_p), np.mean(d_p)))
print('\n Mean recall across n=%d train_test_split: \n LR (XGB) [CART] {RandomForest} classifications: %.3f (%.3f) [%.3f] {%.3f}'
   % (N, np.mean(a_r), np.mean(b_r), np.mean(c_r), np.mean(d_r)))
```

8.8.4 Feature importance

To assess which of the features that was most important to distinguish between IBS and HC, feature importance was calculated for each of the models CART, RF and XGBoost. The results of this can be seen in Figs. 9.18, 9.20 and 9.21. A graph showing the decision tree structure of the CART model applied to the split of data at the last (100th) iteration was also made for visualization purposes.

Permutation importance was also applied to inspect and assess which features that were important for the prediction, given a classification algorithm (e.g. CART, RF, or XGB) and a metric (accuracy). The idea behind this data-driven approach is to measure the decrease in model accuracy when *randomly shuffling the values of each feature separately multiple times*, assuming that permutation of an important feature has a larger effect on model accuracy performance than for less important features. For this we used the `ELI5` Python library and the `sklearn.permutation_importance` with standard number (`n_iter` = 5) of random shuffle iterations. The permutation importances for the features in the three different classifiers are shown in Fig. 9.22.

Chapter 9

Experimental results

9.1 Results of group and ROI-wise feature analysis

A median FA value of 0.425 in IBS patients and a median FA value of 0.443 in HCs is seen in the left insula (wm-lh-insula) region. In the right insula region (wm-rh-insula) a median FA value of 0.415 in IBS patients and a median FA value of 0.431 in HCs is seen.

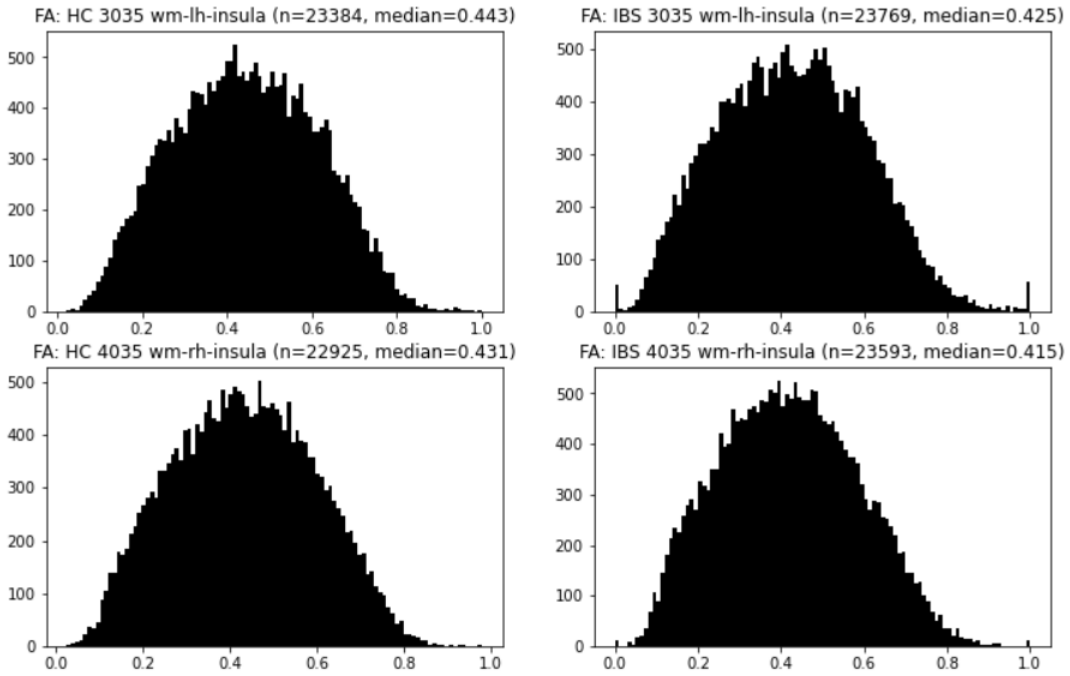

Fig. 9.1: FA values in insula regions of HC and IBS

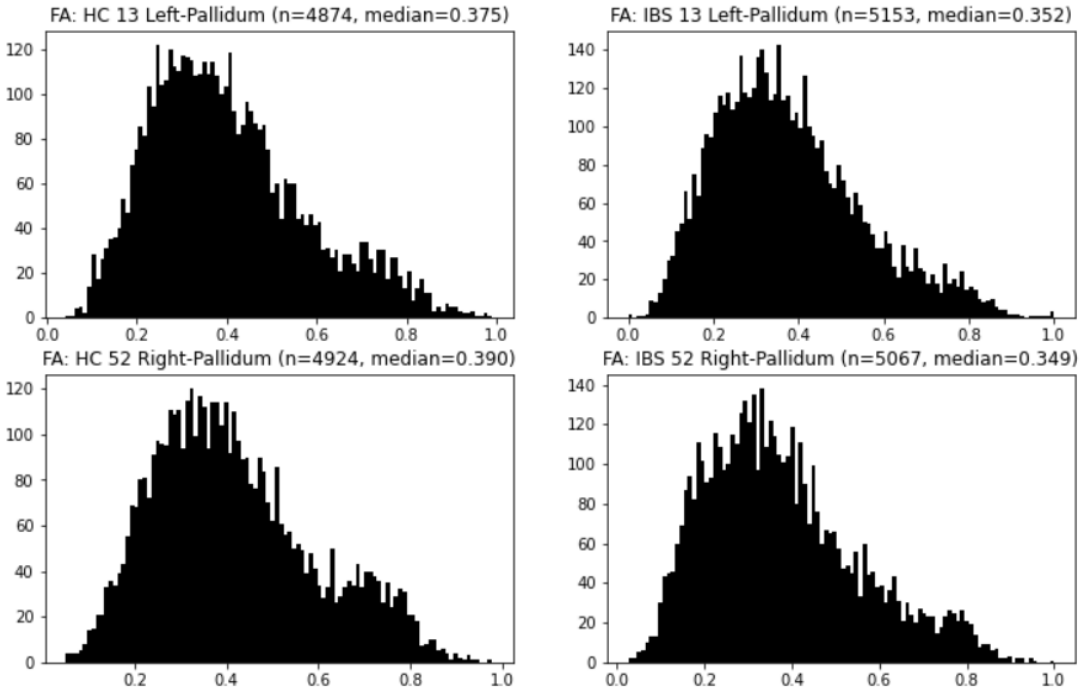

Fig. 9.2: FA values in pallidum regions of HC and IBS

A slightly lower FA value in IBS patients compared to HCs can also be observed in both left and right pallidum regions, as visualized in Fig. 9.2 above. Left hemisphere pallidum region had a median FA value of 0.375 in HC and 0.352 in IBS, while the right hemisphere pallidum region's FA values were 0.390 in HC and 0.349 in IBS. FA values in the thalamus regions shows only a minimal difference, with median FA-value of 0.331 and 0.322 in HC for left and right thalamus regions respectively, and 0.330 and 0.321 in IBS patients as seen in Fig. 9.3. A summary of the median FA values for each region and group can be seen in Table 9.1

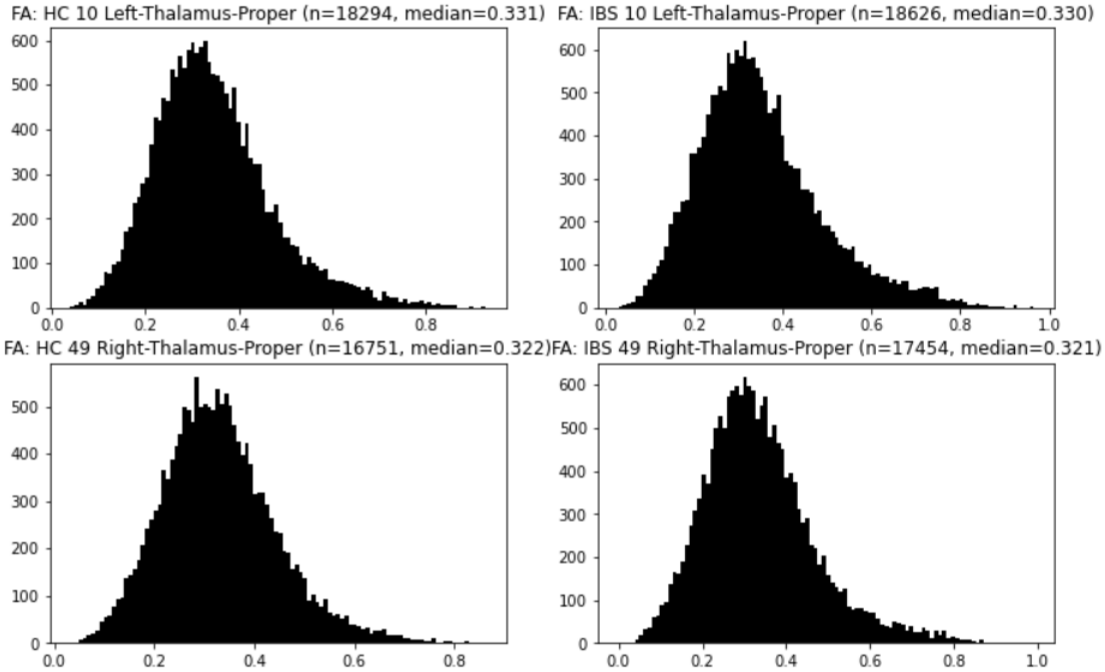

Fig. 9.3: FA values in thalamus regions of HC and IBS

Anatomical region	median FA value IBS	median FA value HC
Insula (left)	0.425	0.443
Insula (right)	0.415	0.431
Thalamus (left)	0.330	0.331
Thalamus (right)	0.321	0.322
Pallidum (left)	0.352	0.375
Pallidum (right)	0.349	0.390

Table 9.1: An overview of median FA values in the targeted ROIs in IBS patients and HC.

The distribution of FA values in HC and IBS within the different regions can be visualized to get a preliminary insight to how the FA values in the two groups compare to each other. This is depicted in Fig. 9.4 through Fig. 9.9.

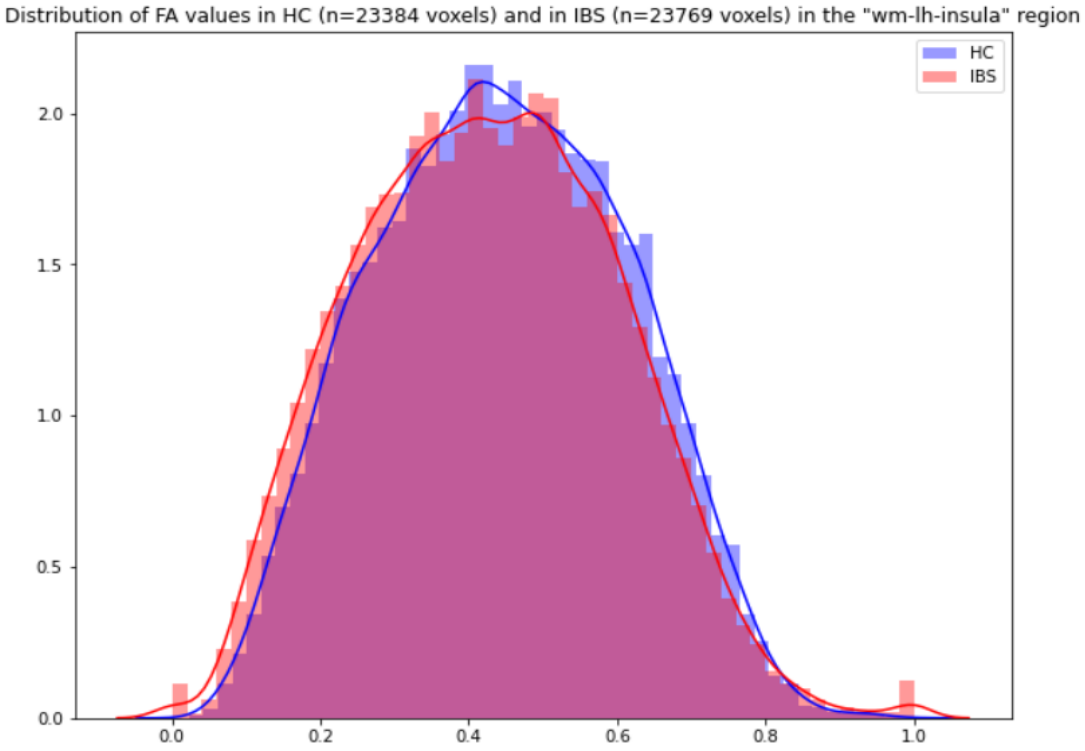

Fig. 9.4: Distributions of FA values in left hemisphere insula region of HC and IBS patients

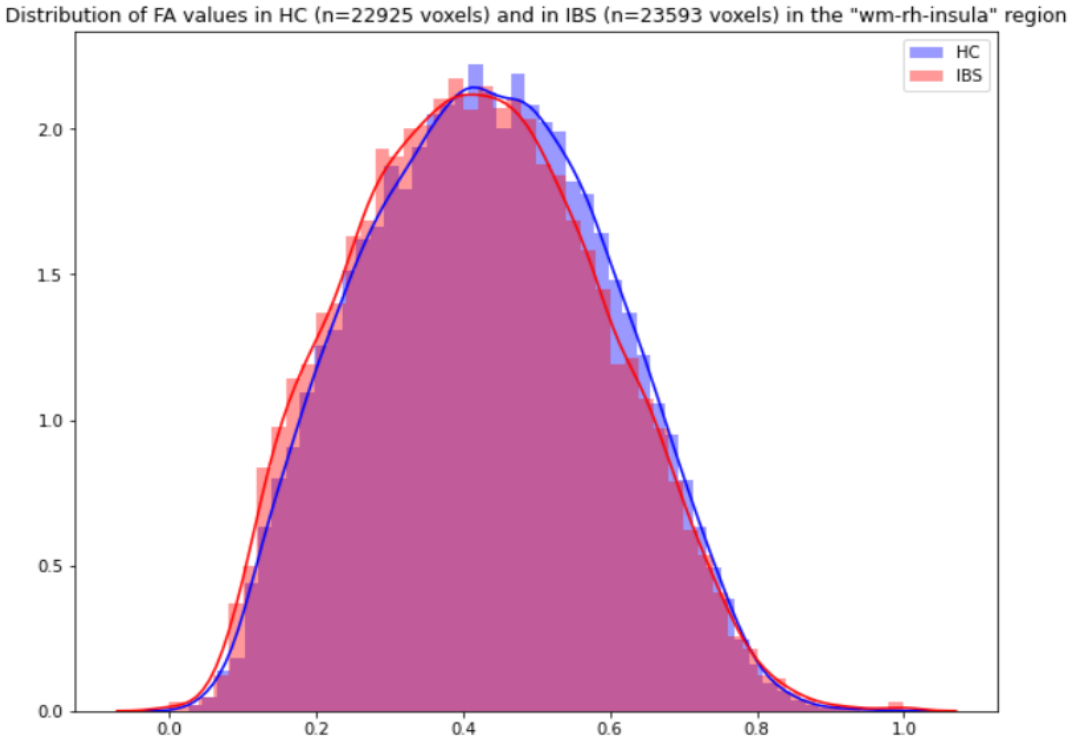

Fig. 9.5: Distribution of FA values in right hemisphere insula region of HC and IBS patients

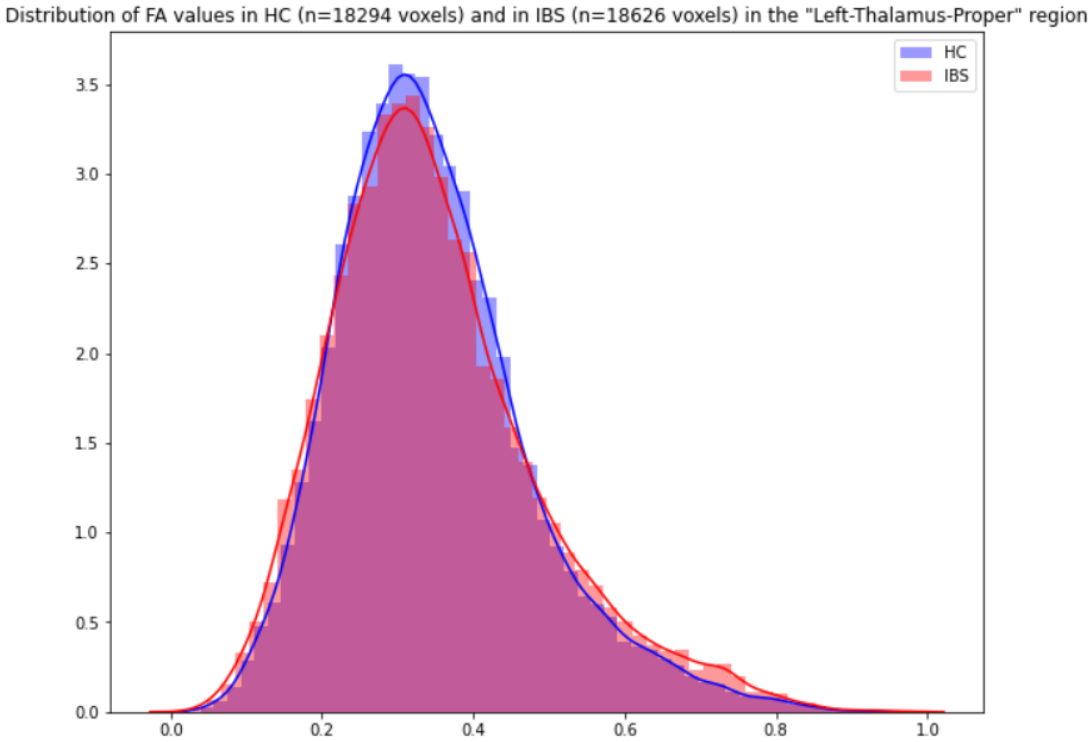

Fig. 9.6: Distributions of FA values in left thalamus region of HC and IBS patients

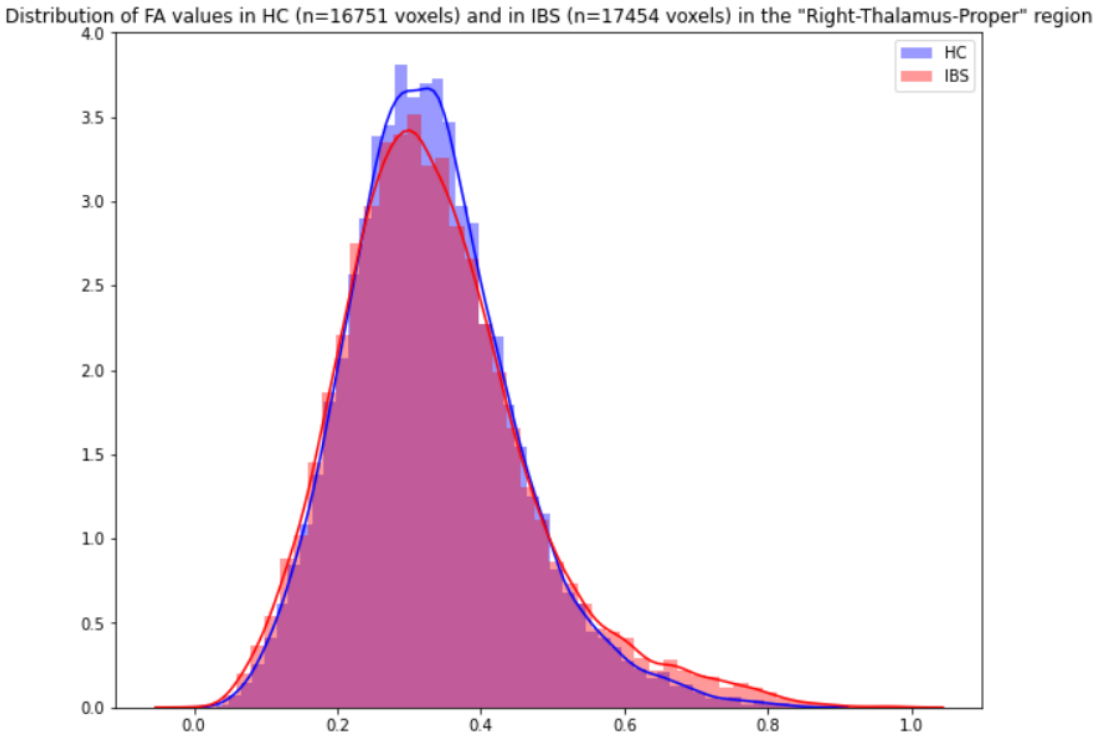

Fig. 9.7: Distributions of FA values in right thalamus region of HC and IBS patients

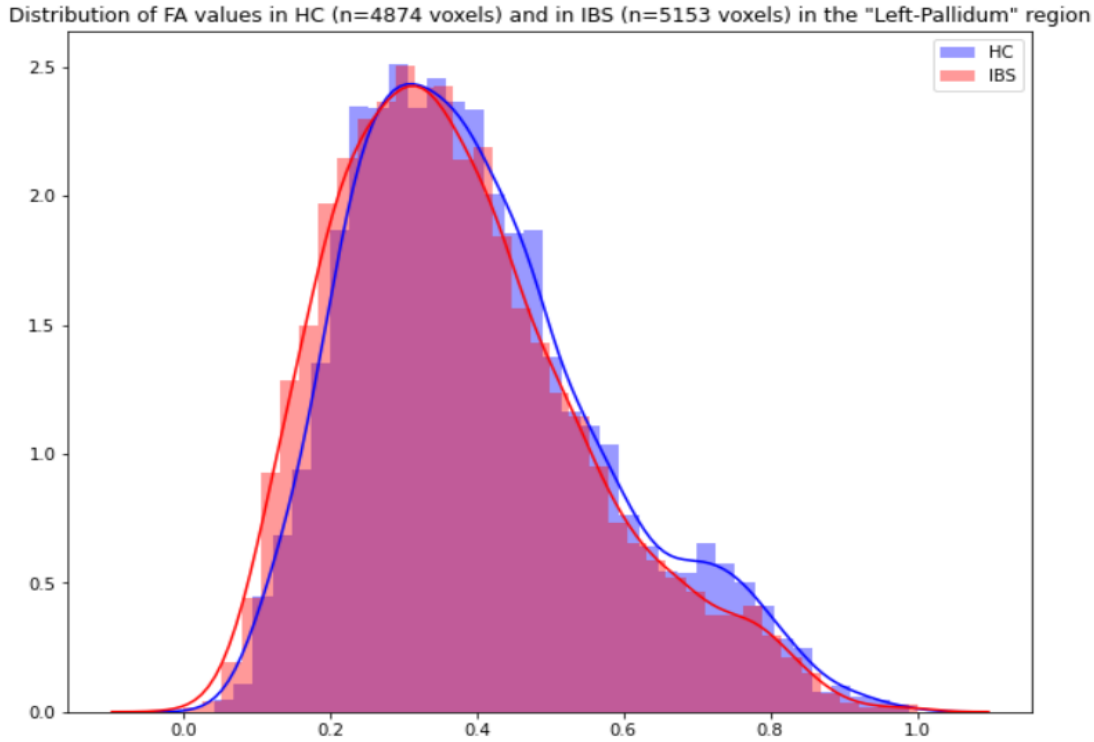

Fig. 9.8: Distribution of FA values in left pallidum region of HC and IBS patients

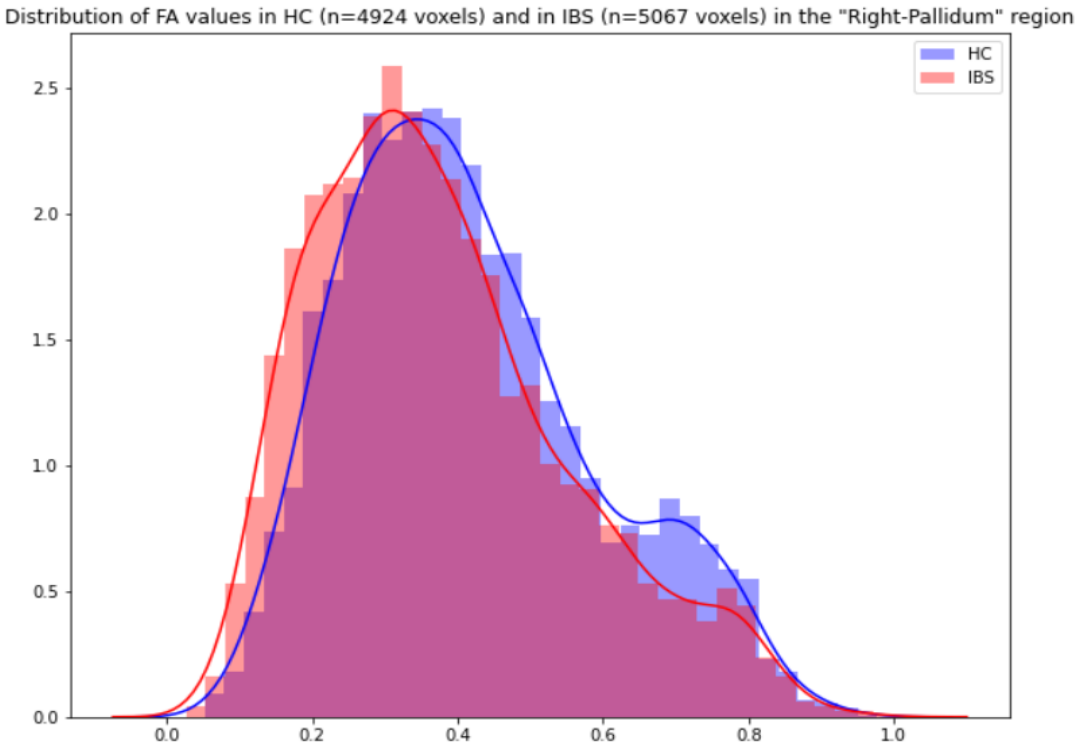

Fig. 9.9: Distributions of FA values in right pallidum region of HC and IBS patients

9.2 Group differences in FA-value distributions and permutation testing

The results of the permutation testing on left hemisphere insula region shows an observed difference in median FA values between the HC and IBS of 0.0187 (`theta_hat`). The `diffCount` was 0, showing that none of the permuted medians was larger than the observed median, which gave $p < 0.0001$ (= 1 / number of permutations) indicating a statistical significant finding, assuming a significance level $\alpha = 0.01$. This is illustrated in Fig. 9.10.

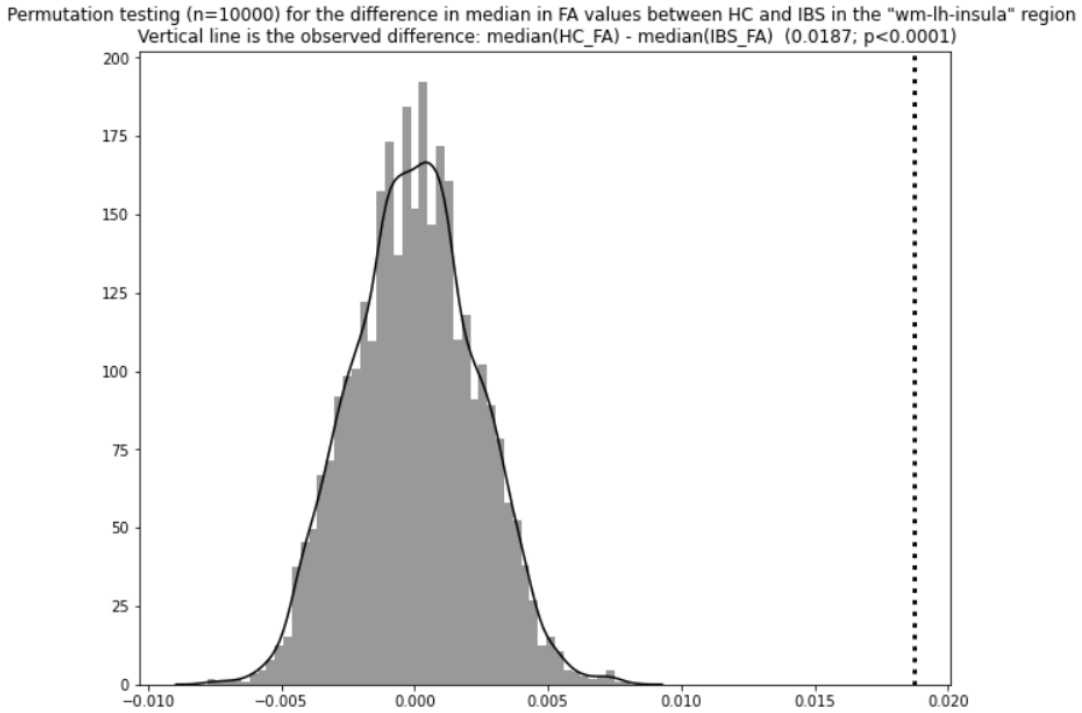

Fig. 9.10: Permutation testing analysis for FA within the "wm-lh-insula" region

Permutation testing on right insula region resulted in a `theta_hat` of 0.0163, `diffCount` of 0 giving a $p < 0.0001$ shown in Fig. 9.11.

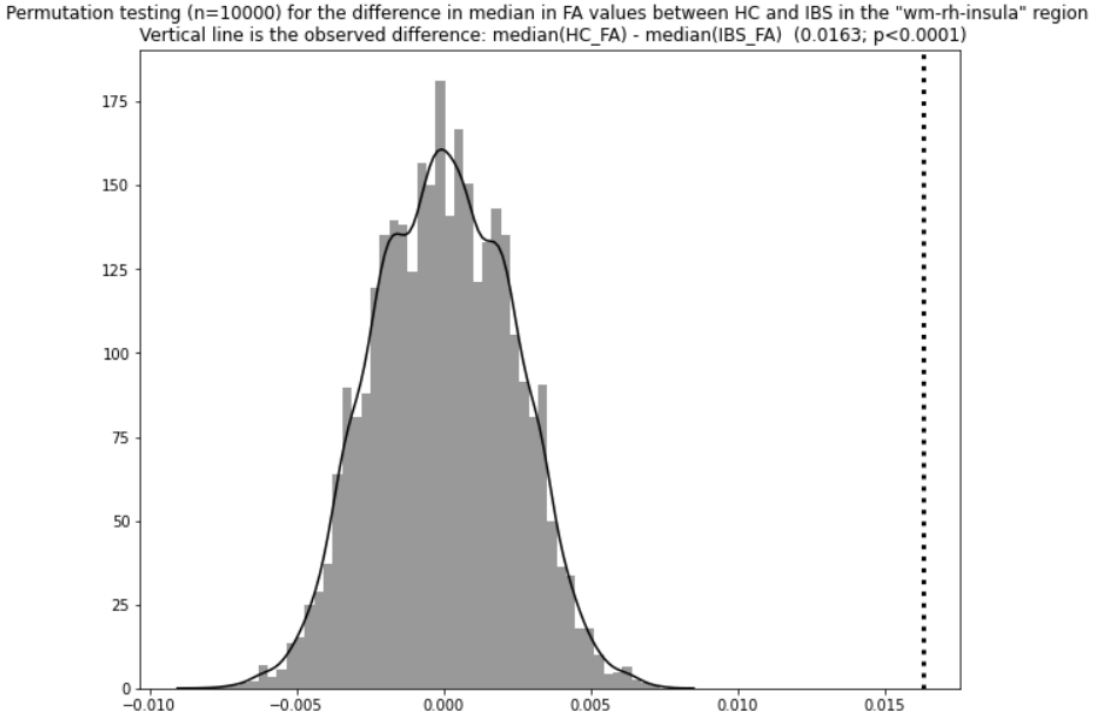

Fig. 9.11: Permutation testing analysis for FA within the "wm-rh-insula" region

The permutation testing on left (Fig. 9.12) and right (Fig. 9.13) thalamus regions returned a `theta_hat` of 0.0011 and 0.0016, respectively. The `diffCount` for left thalamus region was 2495, which is visualized as the gray areas to the right of the vertical line in Fig. 9.12. For the right thalamus region, the `diffCount` was 1612. This shows that a at considerably number of times (i.e. `diffCount`), the permutation gave a median difference that was larger than the observed median. Accordingly, the p-value of left and right thalamus region was $p = 0.2496$ and $p = 0.1613$ respectively, indicating that these findings were not statistically significant, given our significance level.

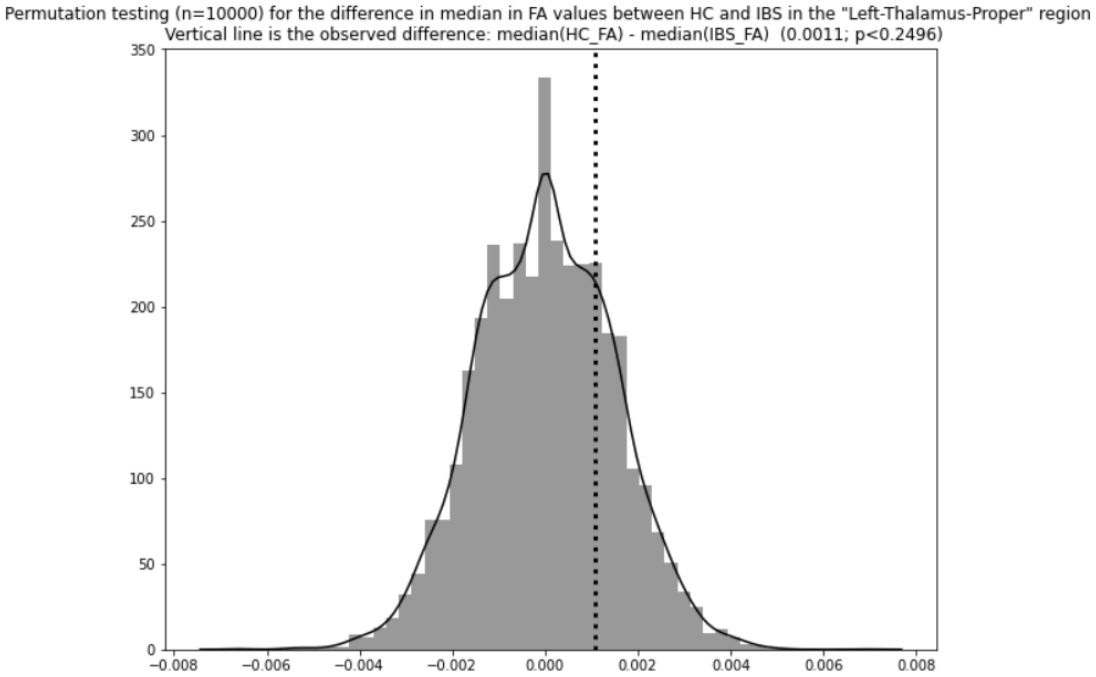

Fig. 9.12: Permutation testing analysis for FA within the "Left-Thalamus-Proper" region

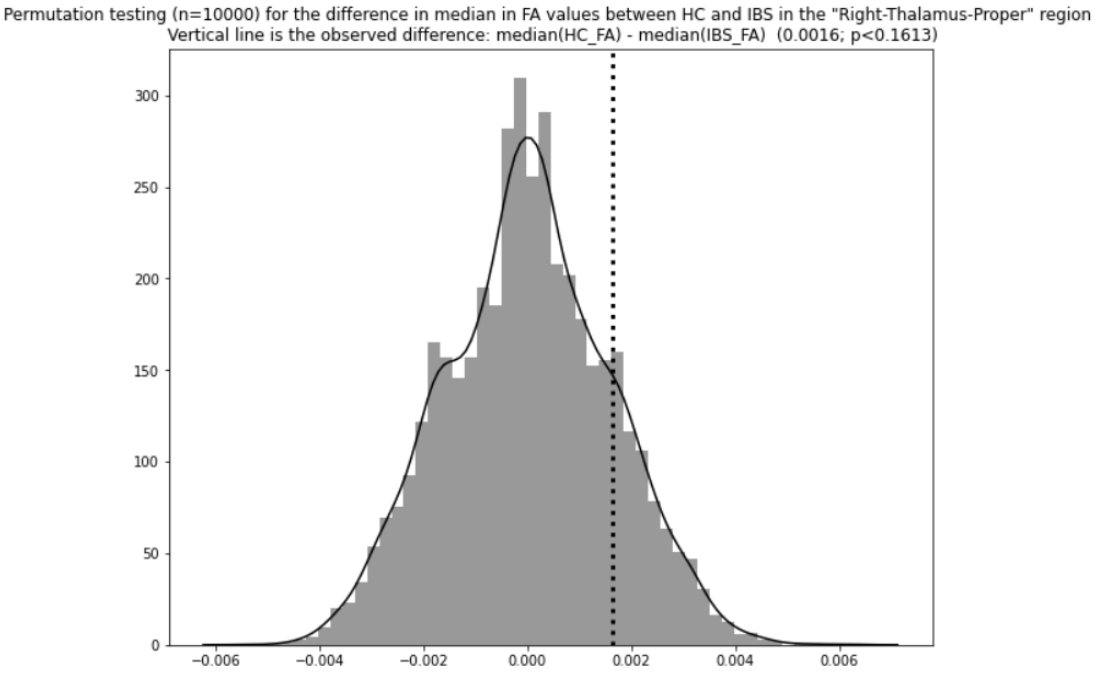

Fig. 9.13: Permutation testing analysis for FA within the "Right-Thalamus-Proper" region

The next two figures show the results of permutation testing on left and right pallidum regions. Left pallidum shows an observed difference in median FA values

between the HC and IBS (`theta_hat`) of 0.0233, and the `diffCount` was 0. Thus, the p-value was reported to be $p < 0.0001$. The `theta_hat` for right pallidum regions was 0.0409, and similarly to left pallidum regions, had a `diffCount` of 0, giving a $p < 0.0001$. These results are show in Figs. 9.14 and 9.15 below.

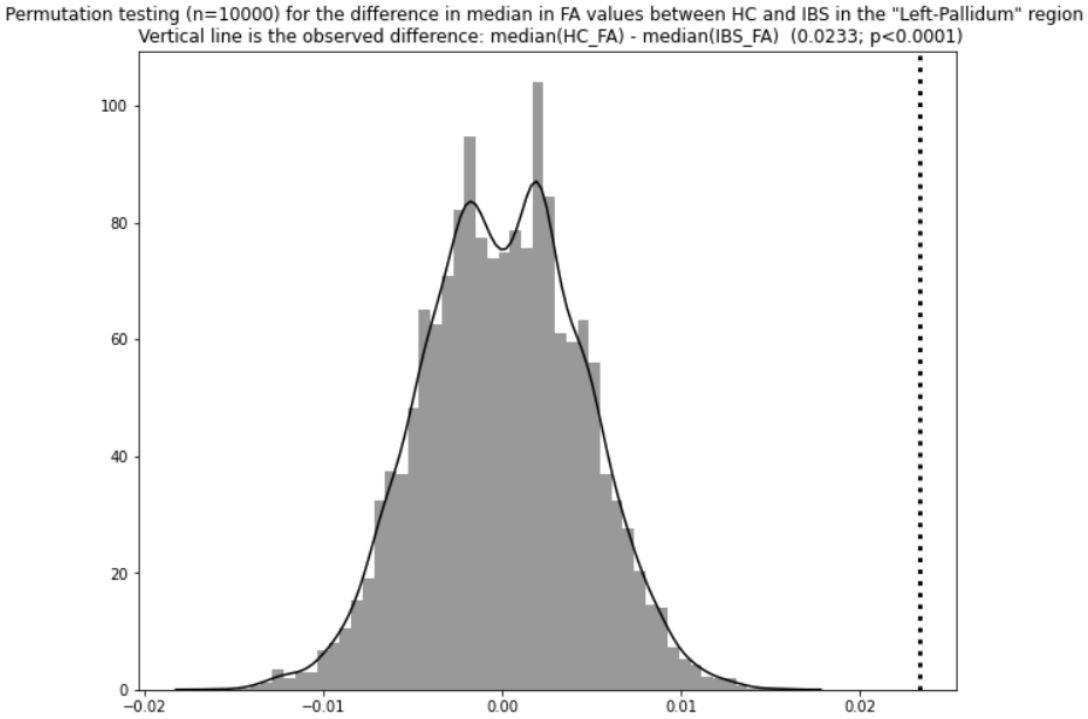

Fig. 9.14: Permutation testing analysis for FA within the "Left-Pallidum" region

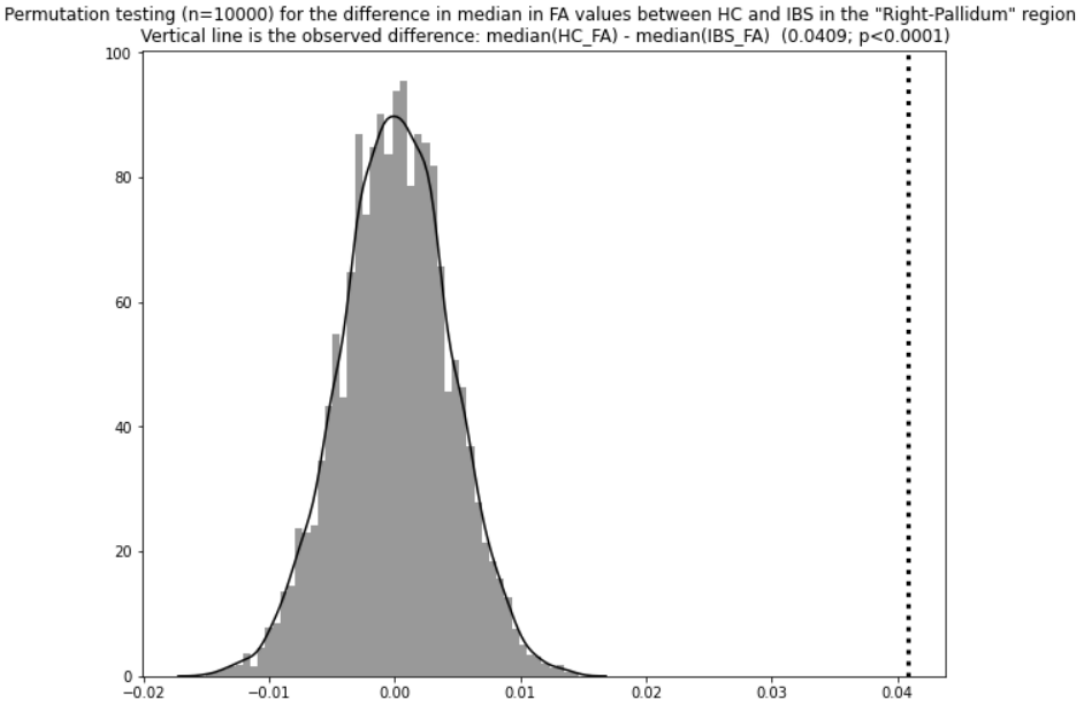

Fig. 9.15: Permutation testing analysis for FA within the "Right-Pallidum" region

9.3 Classification performance

An overview of the different N's used for `train_test_split` and the mean accuracy score for each N and different algorithms is shown in Tab. 9.2.

N	LogReg	XGBoost	CART	RF
10	0.562	0.562	0.713	0.700
50	0.555	0.545	0.640	0.642
70	0.557	0.541	0.634	0.629
100	0.551	0.566	0.646	0.627
150	0.557	0.566	0.637	0.621
200	0.553	0.562	0.629	0.619

Table 9.2: Accuracy performance score of the four different algorithms and its dependency of N (number of folds in the cross-validation procedure).

With N = 100, a *classification accuracy* of 0.646 (64.6%) and 0.627 (62.7%) was achieved for CART and RF models respectively. For XGBoost and Logistic Regression the classification accuracy was slightly lower than CART and RF across the different N used.

The *precision score* for N = 100 in logistic regression, XGBoost, CART and RF was 0.596, 0.551, 0.648 and 0.647 respectively.

The *recall score* for N = 100 in logistic regression, XGBoost, CART and RF was 0.477, 0.662, 0.695, 0.627 respectively.

To get a better understanding and evaluation of the classification accuracy, the *confusion matrices* was computed for all four models. In Fig. 9.16, the confusion matrices for the different decision trees is shown. The visualization shows that the three different decision tree classifiers predicts the same number of correct labels for the two different groups when the models are applied to the test set. The logistic regression classifier classified the test samples a little bit differently than the decision tree classifiers, and this is shown in the confusion matrix in Fig. 9.17. However, it is important to note that the confusion matrices shows only the last run of each of the algorithms.

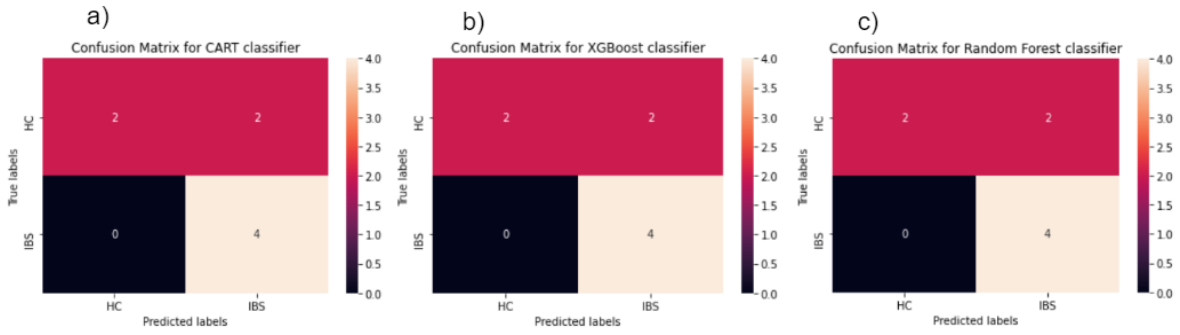

Fig. 9.16: Confusion matrix for a) CART, b) XGBoost and c) Random Forest.

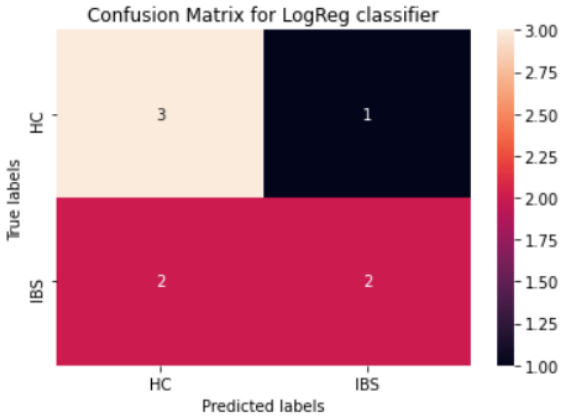

Fig. 9.17: Confusion matrix for logistic regression.

The *feature importance* assessment for the CART-model and Random Forest-model yielded slightly different results. In CART, the feature with the highest relative importance was the median FA value of left pallidum region (FA_13_p50), with a value of 0.41. The median FA value of right pallidum (FA_52_p50) was the second most important with a relative importance of 0.38. Left insula region's median FA value (FA_3035_p50) had a relative importance of 0.21, while the right insula region's median FA value (FA_4035_p50) was of no relative importance for the CART, giving a value of 0.0. This is visualized in figure 9.18. Figure 9.19 shows the feature importances in a decision tree structure. Important note in the decision tree is that it is the visualization of the last run algorithm, number 100 out of 100.

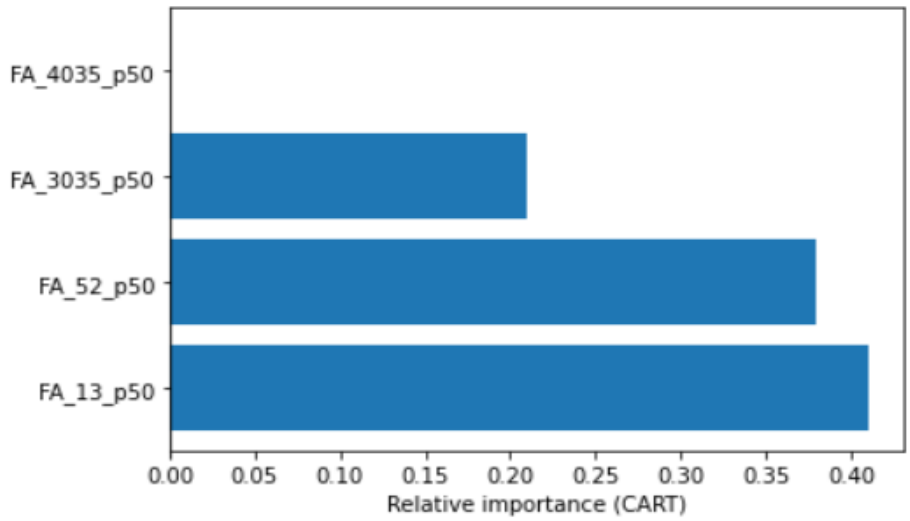

Fig. 9.18: Graphical visualization of the feature importances in the CART model.

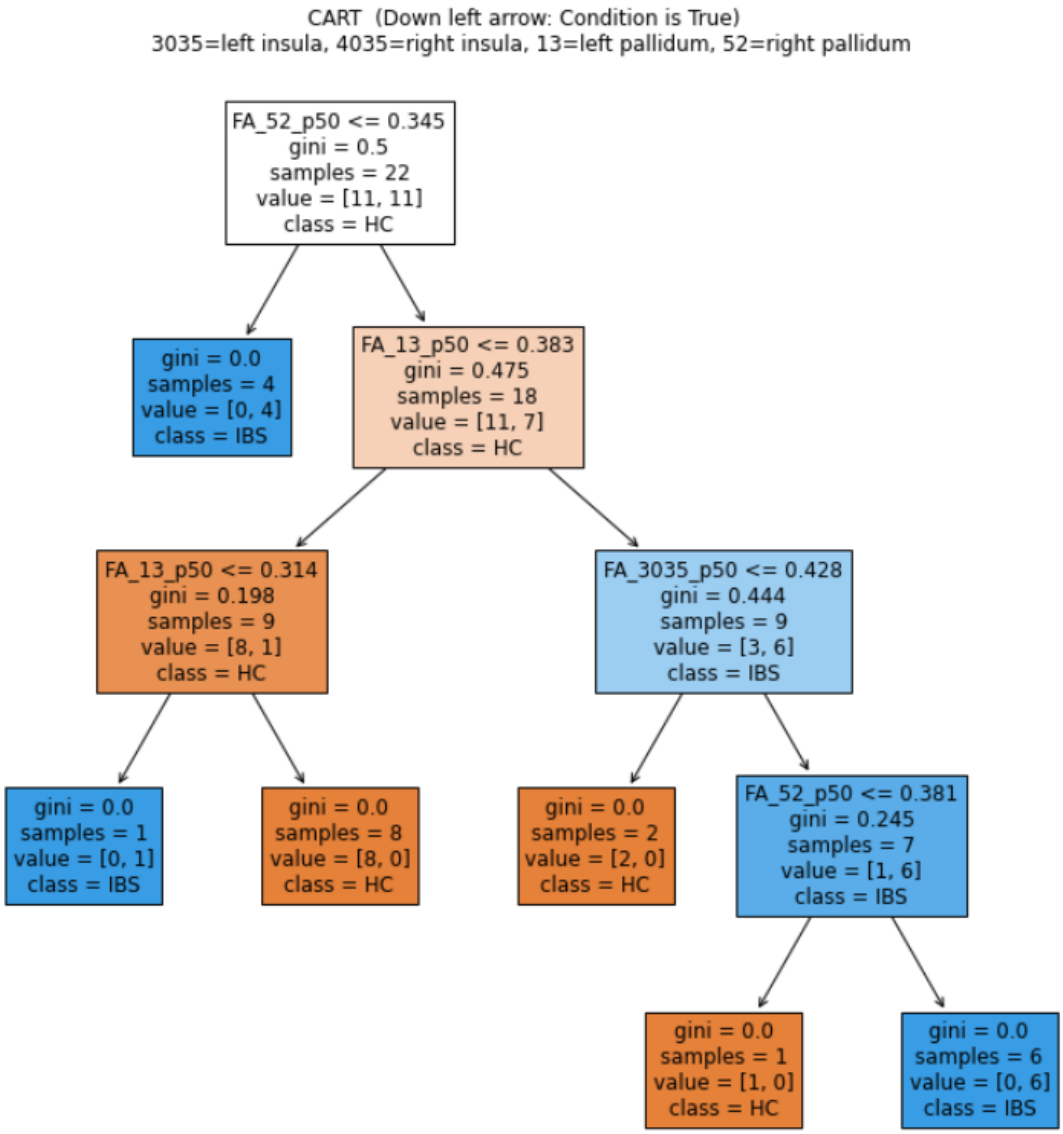

Fig. 9.19: Tree visualization of the feature importances in the CART model.

In the Random Forest model, the feature importance assessment is shown as a graph in Fig. 9.20. From the highest relative importance to lowest relative importance, the values were 0.32 for right pallidum region, 0.31 for left pallidum region, 0.20 for right insula region and 0.16 for left insula region.

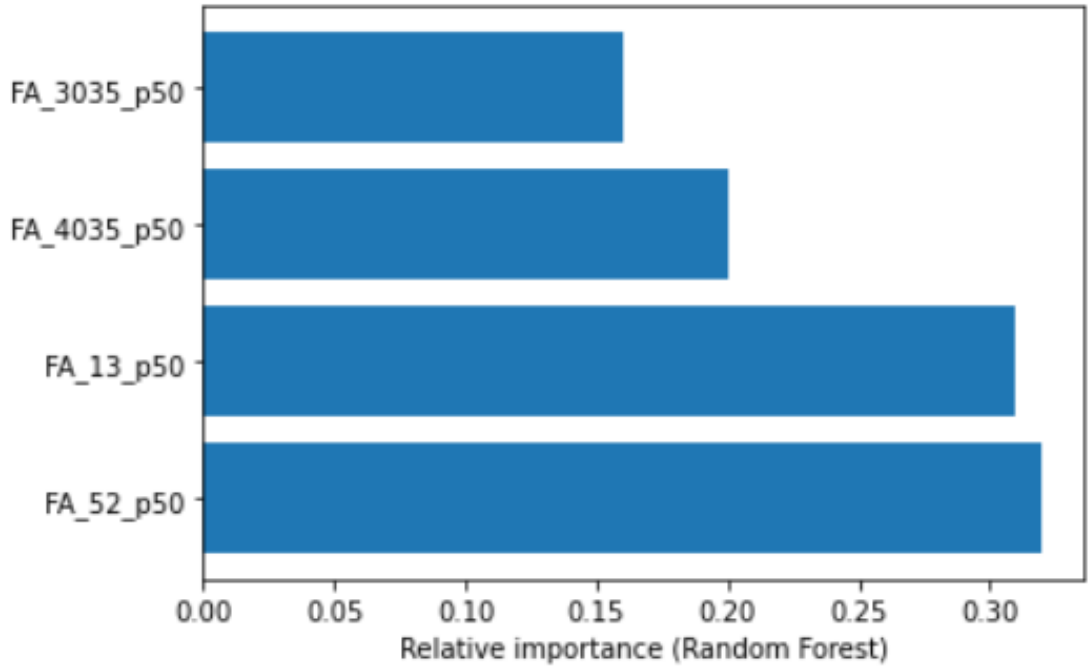

Fig. 9.20: Graphical visualization of the feature importances in the Random Forest model.

The feature importances in XGBoost model shows that the FA values of left pallidum region (FA_13_p50) was the most important feature in distinguishing between HC and IBS by far.

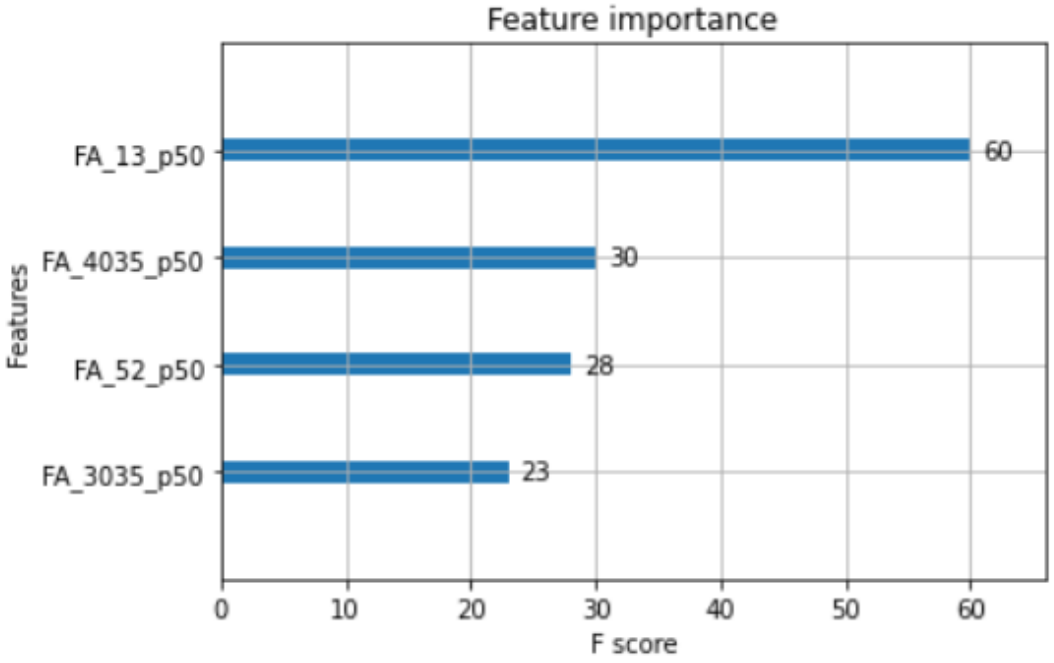

Fig. 9.21: Graphical visualization of the feature importances in the XGBoost model.

Based on the combined feature importances, the FA values of left and right pallidum regions seems to be the most important features in distinguishing between IBS and HC.

The results of the permutation importance for the three different classifiers shown in the three lists in Fig. 9.22. The feature at the top of each models' list has the highest importance, and the features are listed in order of decreased importance for prediction. The first number in each row shows the reduction in the performance of the model when that specific feature is shuffled. These results indicates that the FA values of the right pallidum region is the most important feature used in the prediction for all three different models. The second most important feature was FA values in left pallidum region for all three models. The FA values of the two insula regions were calculated as least important for the model performance in all models, though slightly more important for the Random Forest and XGBoost models compared to CART.

Weight	Feature
0.2000 ± 0.2550	FA_52_p50
0.1500 ± 0.2449	FA_13_p50
0 ± 0.0000	FA_4035_p50
0.0000 ± 0.2236	FA_3035_p50

(a) CART

Weight	Feature
0.3750 ± 0.2236	FA_52_p50
0.2250 ± 0.1000	FA_13_p50
0.0750 ± 0.1225	FA_4035_p50
0.0250 ± 0.1871	FA_3035_p50

(b) RF

Weight	Feature
0.3000 ± 0.2000	FA_52_p50
0.2000 ± 0.1225	FA_13_p50
0.1500 ± 0.1000	FA_4035_p50
0.0000 ± 0.2236	FA_3035_p50

(c) XGBoost

Fig. 9.22: Permutation importance for three different classifiers (CART, RF, XGBoost).

Chapter 10

Discussion

Due to the close dependency between the Results and the Methods used to obtain them, it was found most appropriate for the author to discuss these sections together.

The findings of decreased median FA values in the bilateral insula and pallidum regions in IBS compared to HC indicates the presence of microstructural tissue changes in these regions. In the study performed by Ellingson et al. [2013], they found substantially lower FA values in IBS patients compared to HC in the globus pallidus. Mayer et al. [2019] pointed out that lower FA values in globus pallidus were also found in female IBS patients, which can indicate that the sex of the subject might influence the measured FA values. In the data used in this thesis there were participants of both sexes, but a considerable higher number of females (11 IBS, 10 HC) than males (4 IBS, 5 HC). It is plausible that this has influenced the results. However, the IBS diagnosis is considered to be most common in females [Quigley, 2018], thus indicating the representativeness of the samples.

The permutation testing analysis gave several statistical significant findings. The number of observations that is *equal to* or *more extreme* than the initial observation divided by the total number of permutation observations gives the p-value. In left and right insula and pallidum regions, the p-value of $p < 0.0001$ shows that the probability of H_0 being true is less than 0.01%. This can be interpreted as an indication of changes in white matter integrity in IBS compared to HC at group level. As previously described, IBS is associated with abdominal pain [Aziz et al., 2018; Canavan et al., 2014a] and both the insula and pallidum regions have been implicated in the interpretation and modulation of pain [Ellingson et al., 2013; Mayer et al., 2019] so it might be that the

IBS leads to the lower FA values in the IBS patients. An important consideration when looking at the findings however, is that even though the FA metric has a high sensitivity for diffusion to both the presence and integrity of white matter, it might be slightly confounded by the *partial volume effect* since a voxel can contain water from different types of tissue. There is a possibility that there can be other reasons for the lower FA values found in the IBS patients, since other pain related disorders can also cause lower FA values in insula and pallidum [Gracely et al., 2002; Kano et al., 2020]. As FA often is used as a measure of white matter integrity, other factors and conditions that results in a loss of white matter integrity might also result in lower FA values. This can be factors such as aging or neurological and neurodegenerative diseases [Timmers et al., 2016].

Frøkjær et al. [2011] found decreased mean FA values in the anterior insula bilaterally in patients suffering from chronic pancreatitis with abdominal pain ($p<0.0001$) compared with HC, which supports the importance of the gut-brain axis and abdominal pain related white matter integrity changes in the insula regions. When considering that the brain and its connectivity can be viewed as a network, to approach the findings in this thesis from a network perspective is also relevant. The statistical significant decreased median FA values might suggest that damage to white matter integrity may influence the regulation of networks, especially the sensorimotor- and salience networks that already have been connected to IBS Labus et al. [2019]; Mayer et al. [2019].

Ellingson et al. [2013] found that IBS patients had substantially lower FA values in the thalamus regions, while in our analysis there was only a minimal difference between the FA values in the thalamus regions for HC and IBS. As visualized in the figures 9.4 to 9.9, the distribution of FA values for HC and IBS overlaps to great extent, indicating that the FA values of the two groups are largely similar. This is not surprising when considering all the similarities in the two groups, since the goal is that only the presence of IBS should be influencing the FA values. Despite the findings of Ellingson et al. [2013], the median FA values in the thalamus regions in our study did not discriminate well between HC and IBS at group level. This might be due to a number of reasons. One explanation might be the difference in the number of subjects in the study of Ellingson et al. [2013] compared to ours. Ellingson et al. [2013] had IBS=33 and HC=93, more than twice the number of IBS patients compared to this study. Another reason could be the difference in the methodological approach to the DTI data.

With the small sample size available in this study, the permutation testing was used to increase the internal validity.

The difference in median FA values in insula and pallidum regions between HC and IBS at a group level, suggested that these features might be used to discriminate between subjects with IBS and HC. When using these features to make predictions about new data, in this case to use the features to classify in a test set whether a subject was in the IBS group or HC group, the moderate accuracy, precision and recall of the machine learning models show that this is not easy.The highest accuracy score of 64.6% was obtained using CART-model, which may be described as moderately successful. As seen in Labus et al. [2019], supervised learning has been used on structural brain imaging data using morphometric features to discriminate IBS versus HC with a predictive accuracy of 70%. This led to the statement that it is not sufficient for diagnosis, but that it underscores the presence of brain alterations in IBS and the utility of data-driven analyses such as machine learning. Although the scores alone are not sufficient for diagnosis, our results still gives an indication that the features and target regions are relevant for classification of IBS versus HC.

Results also gave an insight into which of the features that was most successful in discriminating between IBS and HC. In this thesis study, the median FA values in the pallidum region had the highest importance for the classification, while the FA values in insula regions had lower importance. Given the prominent role insula regions has in the processing and perception of visceral pain [Bednarska, 2019; Mayer et al., 2019; Ruffle et al., 2017], one might have had the expectation that the importance of this feature should have had more positive impact in the classification. The underlying reasons for this lower importance can be due to a number of reasons pertaining to e.g. methodological choices or the limited sample size.

The inclusion of rs-fMRI data and/or symptom scores for analysis might also have improved the three performance metrics, however this would have meant the inclusion of other features due to the nature of the data. This was not done on account of the time and resource contstraints. Also, if a higher number of IBS patients and HCs was included in the study, it would have given the machine learning model more data to learn from and possibly make predictions with higher accuracy.

On that note, classification accuracy works well as a score when the number of samples in each class are equal. If the classes had been unbalanced, for example if the model was trained on a training set where 80% of samples belonged to IBS class and

20% of the training samples belonging to HC class, the model could get an accuracy score of 80% if it just predicted "IBS" on all samples in the test set. In this case, an accuracy score of 80% would not necessarily mean that the model is good, because when applied to a balanced test set with 50% in each class, the accuracy score would drop to 50% and imply that the model is just "guessing" and has not actually learned from the true labels. When reviewing the training- and test sets used in this thesis it becomes apparent that the test set contained four subject with IBS and four HC subjects, meaning that the model was trained on 11 subjects from each group. An accuracy score better than 50% indicates that the model actually has learned from the true labels. This is because of the *stratification* applied to the labels as described in section 8.8.2.

When it comes to the supervised machine learning algorithms, the purpose of splitting the available data into training set and test set is to prevent the model from making the methodological mistake of both learning and predicting on the same data. If the data set is not split into a training set and a test set, the model would be able to just repeat the labels it has already seen, thus resulting in a perfect score. The model would have learned the data "by heart", and can therefore make a perfect prediction of whether the patients belongs to one group or the other. This is called *overfitting* and is not useful for a machine learning model because it would make it unable to make useful predictions on new and unseen data. Another way to describe overfitting is that if the researcher works too hard to find the best fit to the training data, the risk is that "noise" is fitted in the training data by memorizing anomalies instead of learning a general predictive rule [Dietterich, 1995].

To get better accuracy and/or precision score, manual tuning of the machine learning algorithms' settings, called *hyperparameters*, might have yielded better results for the different models. The danger of tuning these settings is that there is a danger of overfitting the model to the data, because it is possible to tune the parameters of the model until an optimal performance of the model is reached by what is called "data leakage" from the training set to the test set. If this happens, the metrics would not reflect generalization performance. To avoid this while training the models in this thesis, the hyperparameters was set at 'default'.

As visualized in the confusion matrices in Fig. 9.16, the trained machine learning models did not make 100% correct classifications of HC and IBS. Theses models predicted that two out of eight subjects in the test set belonged to the IBS group, while

they in fact belonged to HC group, i.e. "false negative". On the other hand, the same three models did not predict that any of the HCs belonged to the IBS group, i.e. "false positives". But as mentioned previously, it is important to note that the confusion matrices is a visualization of one out of the $N = 100$ predictions. The recall quantifies the number of positive class predictions made out of all positive class examples in the dataset, and the CART model's recall score of 0.695 shows that 69.5% of the actual IBS patients were detected.

Another methodological aspect that might have influenced the results is the image processing steps followed in our study. As described by Zwanenburg et al. [2019] the image processing steps that leads to image biomarker calculations can be performed in a number of different ways. Choosing differently in some of the steps might have given other results, e.g. feature calculation (see Fig. 4.1) on other ROIs than the insula, pallidum and thalamus. Choices regarding which ML models to use could also have an impact on the outcome, as there are several different models available (e.g. artificial neural networks) for classification tasks.

Finally, as there is no specific test that can diagnose IBS, the ROME III criteria was used to diagnose IBS in patients in the sample used in the study. There might therefore be a possibility that the IBS sample is insufficiently charaterized or with difference in symptom severity that could influence the outcome. However, the lack of specific tests for IBS is one of the reasons for searching for image biomarkers.

Future perspectives

One methodological aspect related to future image-based biomarkers research could be the application of *deep learning methods* (see section 5.2.1 for reference). Since deep learning methods allows a machine to use the raw data for automatical discovery of the representations needed for classification, the step of manually designing features from the raw data is bypassed. This could possibly eliminate the problem of choosing suboptimal features for training. Deep learning methods, especially CNNs that are designed to process data that come in the form of multiple arrays, such as 3D for video or volumetric images, can also be used for prediction and classification, in addition to fast and accurate image segmentation.

The use of multimodal brain imaging analyses is still promising, and combining different features in different brain regions and networks with computational methods and machine learning could be pursued in the search for quantitative image-based biomarkers in IBS.

It has become apparent that there is a need for multidisciplinary effort in the endeavour to establish biomarkers for IBS. Medical- and neuroscience skills, combined with technical-, imaging-, computing-, and statistical skills and techniques could provide new insights into the multifaceted disorder that is IBS.

Chapter 11

Conclusion

The aims of the experimental part of this thesis were (i) to investigate if tissue microstructure assessed with fractional anisotropy (FA) measurements in target brain regions is different in patients with IBS versus HC at a group level, and (ii) how accurate it can be predicted whether a subject is IBS or HC using FA-based brain signatures and machine learning models. As the median FA values in the targeted regions were found to be lower in IBS patients than HC, it indicates a loss of white matter integrity in these regions. Median FA values in the insula and pallidum regions were found to be statistically significant ($p < 0.0001$).On a group level, these FA values can be used to detect differences and discriminate between IBS and HC. The white matter changes in the targeted regions supports the implication of the brain-gut-axis in IBS, but further research is needed to understand the mechanisms that leads to these changes in white matter.

The machine learning models highest accuracy score of 64.8%, precision score of 64.6% and recall score 69.5% can be considered as moderately successful, since the predictions are clearly better than "chance level". However, it indicates that the features used might not be best suited for classification at subject level. Further research into quantitative imaging biomarkers should be pursued to enable prediction of IBS at subject levels, e.g. by incorporating information such as quantitative network properties derived from the resting state fMRI recordings that was part of the MRI examination protocol.

References

Abramson, R., Burton, K., Yu, J.-P., Scalzetti, E., Yankeelov, T., Rosenkrantz, A., [...], and Subramaniam, R. (2015). Methods and Challenges in Quantitative Imaging Biomarker Development. *Acad Radiol*, 22(1):25–32, DOI: 10.1016/j.acra.2014.09.001.

Alexander, A., Lee, J., Lazar, M., and Field, A. (2007). Diffusion Tensor Imaging of the Brain. *Neurotherapeutics*, 4(3):316–329, DOI: 10.1016/j.nurt.2007.05.011.

Andica, C., Kamagata, K., Hatano, T., Saito, Y., Ogaki, K., Hattori, N., and Aoki, S. (2019). MR Biomarkers of Degenerative Brain Disorders Derived From Diffusion Imaging. *J Magn Reson Imag*, DOI: 10.1002/jmri.27019.

Arzani, M., Jahromi, S., Ghorbani, Z., Vahabizad, F., Martelletti, P., Ghaemi, A., Sacco, S., and Togha, M. (2020). Gut-brain Axis and migraine headache: a comprehensive review. *The Journal of Headache and Pain*, 21(15), DOI: 10.1186/s10194-020-1078-9.

Assaf, Y. and Pasternak, O. (2008). Diffusion Tensor Imaging (DTI)-based White Matter Mapping in Brain Research: A Review. *J Mol Neurosci*, 34:51–61, DOI: https://doi.org/10.1007/s12031-007-0029-0.

Aziz, I., Törnblom, H., Palsson, O., Whitehead, W., and Simrén, M. (2018). How change in IBS Criteria from Rome III to Rome IV Impacts on Clinical Characteristics and Key Pathophysiological Factors. *Am J Gastroenterol*, 113(7):1017–1025, DOI: 10.1038/s41395-018-0074-z.

Baliyan, V., C.J., D., Sharma, R., and Gupta, A. (2016). Diffusion weighted imaging: Technique and application. *World J Radiol*, 8(9):785–798, DOI: 10.4329/wjr.v8.i9.785.

Basser, P., Mattiello, J., and LeBihan, D. (1994). MR Diffusion Tensor Spectroscopy and Imaging. *Biophys J*, 66(1):259–267, DOI: 10.1016/S0006-3495(94)80775-1.

Bednarska, O. (2019). *Peripheral and Central Mechanisms in Irritable Bowel Syndrome - in search of links*. PhD thesis, Linköping University.

Borghi, J. and Van Gulick, A. (2018). Data Management and sharing in neuroimaging: Practices and perceptions of MRI researchers. *PLoS ONE*, 13(7), DOI: https://doi.org/10.1371/journal.pone.0200562.

Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic Resonance Imaging: Basic Principles and Sequence Design*. John Wiley and Sons, Inc., New Jersey, USA, second edition, ISBN: 9781118633984.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, ISSN: 1471-003X, 1471-0048, DOI: 10.1038/nrn2575, http://www.nature.com/articles/nrn2575.

Canavan, C., West, J., and Card, T. (2014a). Review article: the economic impact of the irritable bowel syndrome. *Alimentary Pharmacology and Therapeutics*, 40(9):1023–1034, DOI: https://doi.org/10.1111/apt.12938.

Canavan, C., West, J., and Card, T. (2014b). The epidemiology of irritable bowel syndrome. *Clin Epidemiol*, 6:71–80, DOI: 10.2147/CLEP.S40245.

Carabotti, M., Scirocco, A., Maselli, M., and Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology*, 28(2):203–209.

Chen, J. Y.-W., Blankstein, U., Diamant, N., and Davis, K. (2011). White matter abnormalities in irritable bowel syndrome and relation to individual factors. *Brain Research*, 1392:121–131, DOI: 10.1016/j.brainres.2011.03.069.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, ISSN: 1053-8119, DOI: 10.1016/j.neuroimage.2006.01.021.

Despotović, I., Goossens, B., and Philips, W. (2015). MRI Segmentation of the Human Brain: Challenges, Methods, and Applications. *Computational and Mathematical Methods in Medicine*, 2015, DOI: https://doi.org/10.1155/2015/450341.

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1):1–15, ISSN: 10538119, DOI: 10.1016/j.neuroimage.2010.06.010, https://linkinghub.elsevier.com/retrieve/pii/S1053811910008542.

Dietterich, T. (1995). Overfitting and Undercomputing in Machine Learning. *ACM Computer Surveys*, 27(3):326–327, DOI: https://doi.org/10.1145/212094.212114.

Douek, P., Turner, R., Pekar, J., Patronas, N., and Le Bihan, D. (1991). MR Color Mapping of Myelin Fiber Orientation. *Journal of Computer Assisted Tomography*, 15(6):923–929, DOI: 10.1097/00004728-199111000-00003.

Drossman, D., Morris, C., Schneck, S., Hu, Y., Norton, N., Norton, W., Weinland, S., Dalton, C., Leserman, J., and Bangdiwala, S. (2009). International Survey of Patients With IBS Symptom Features and Their Severity, Health Status, Treatments, and Risk Taking to Achieve Clinical Benefit. *J Clin Gastroenterol*, 43(6):541–550, DOI: 10.1097/MCG.0b013e318189a7f9.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, USA.

Ellingson, B., Mayer, E., Harris, R., Ashe-McNally, C., Naliboff, B., Labus, J., and Tillisch, K. (2013). Diffusion tensor imaging (dti) detects microstructural reorganization in the brain associated with chronic irritable bowel syndrome (ibs). *Pain*, 154(9):1528–1541, DOI: 10.1016/j.pain.2013.04.010.

Flores-Alvarez, E., Durand-Muños, C., Cortes-Hernandez, F., Muños-Hernandes, O., Moreno-Jimenes, S., and Roldan-Valadez, E. (2019). Clinical Significance of Fractional Anisotropy Measured in Peritumoral Edema as a Biomarker of Overall Survival in Glioblastoma: Evidence Using Correspondence Analysis. *Neurol India*, 67:1074–1081, DOI: 10.4103/0028-3886.266284.

Ford, A., Lacy, B., and Talley, N. (2017). Irritable bowel syndrome. *N Engl J Med*, 376:2566–2578, DOI: 10.1056/NEJMra1607547.

Frøkjær, J., Andersen, L., Brock, C., Simrén, M., Ljungberg, M., Søfteland, E., Dimcevski, G., Yavarian, Y., Gregersen, H., and Drewes, A. (2013). Altered Brain Microstructure Assessed by Diffusion Tensor Imaging in Patients With Diabetes and Gastrointestinal Symptoms. *Diabetes Care*, 36:662–668, DOI: 10.2337/dc12-1131.

Frøkjær, J., Olesen, S., Gram, M., Yavarian, Y., Bouwense, S., Wilder-Smith, O., and Drewes, A. (2011). Altered brain microstructure assessed by diffusion tensor imaging in patients with chronic pancreatitis. *Gut*, 60:1554–1562, DOI: 10.1136/gut.2010.236620.

Giannelli, M., Cosottini, M., Michelassi, M. C., Lazzarotti, G., Belmonte, G., Bartolozzi, C., and Lazzeri, M. (2010). Dependence of brain DTI maps of fractional anisotropy and mean diffusivity on the number of diffusion weighting directions. *J Appl Clin Med Phys*, 11(1):179–190, DOI: 10.1120/jacmp.v11i1.2927.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., Poline, J.-B., Rokem, A., Schaefer, G., Sochat, V., Triplett, W., Turner, J. A., Varoquaux, G., and Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3:160044, ISSN: 2052-4463, DOI: 10.1038/sdata.2016.44, https://www.nature.com/articles/sdata201644.

Gracely, R., Petzke, F., Wolf, J., and Clauw, D. (2002). Functional magnetic resonance imaging evidence of augmented pain processing in fibromyalgia. *Arthritis and Rheumatology*, 46(5):1333–1343, DOI: 10.1002/art.10225.

Grusso, F. and Wheeler-Kingshott, C. (2018). The Diffusion of Water (DTI). In Cercignani, M., Dowell, N., and Tofts, P., editors, *Quantitative MRI of the Brain - Principles of Physical Measurement*, chapter 8, pages 111–138. CRC Press, Boca Raton, second edition, ISBN: 9781315363578.

Guresen, E. and Kayakutlu, G. (2011). Definition of artificial neural networks with comparison to other networks. *Procedia Computer Science*, 3:426–433, DOI: 10.1016/j.procs.2010.12.071.

Han, L., Wang, Z., Tong, E., Williams, L., Zaharchuk, G., Zeineh, M., Goldstein-Piekarski, A., Ball, T., Liao, C., and Wintermark, M. (2018). Resting-State Functional MRI (rs-fMRI): Everything that Non-Experts Have Always Wanted to Know. *Am J Neuroradiol*, 39(8):1390–1399, DOI: 10.3174/ajnr.A5527.

Hao, X., Xu, D., Bansal, R., Dong, Z., Liu, J., Wang, Z., [...], and Peterson, B. (2013). Multimodal magnetic resonance imaging: The coordinated use of multiple, mutually informative probes to understand brain structure and function. *Hum Brain Mapp*, 34:253–271, DOI: 10.1002/hbm.21440.

Honey, G. and Bullmore, E. (2004). Human Pharmacological MRI. *TRENDS in pharmacological sciences*, 25(7):366–374, DOI: 10.1016/j.tips.2004.05.009.

Huisman, T. (2010). Diffusion-weighted and diffusion tensor imaging of the brain, made easy. *Cancer Imaging*, 10(Spec No 1A):163–171, DOI: doi:10.1102/1470-7330.2010.9023.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France. PMLR.

ISRRT and EFRS (2020). Artificial intelligence and the radiographer/radiological technologist profession: A joint statement of the international society of radiographers and radiological technologists and the european federation of radiographer societies. *Radiography*, 26(2):93–95, DOI: 10.1016/j.radi.2020.03.007.

Jenkinson, M. (2006). Image Registration and Motion Correction. *Proc Int Soc Magn Reson Med*, 14.

Jones, D. (2009). Gaussian Modeling of the Diffusion Signal. In Johansen-Berg, H. and Behrens, T., editors, *Diffusion MRI: From Quantitative Measurement to In vivo Neuroanatomy*, pages 37–54. Elsevier, London, first edition, ISBN: 978-0-12-374709-9.

Jones, D., Knösche, T., and Turner, R. (2013). White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *Neuro Image*, 73:239–254, DOI: 10.1016/j.neuroimage.2012.06.081.

Kano, M., Dupont, P., Aziz, Q., and Fukudo, S. (2018). Understanding Neurogastroenterology From Neuroimaging Perspective: A Comprehensive Review of Functional and Structural Brain Imaging in Functional Gastrointestinal Disorders. *Journal of Neurogastroenterology and Motility*, 24(4):512–527, DOI: https://doi.org/10.5056/jnm18072.

Kano, M., Muratsubaki, T., Yagihashi, M., Morishita, J., Mugikura, S., Dupont, P., Takase, K., Kanazawa, M., Van Oudenhove, L., and Fukudo, S. (2020). Insula Activity to Visceral Stimulation and Endocrine Stress Responses as Associated With

Alexithymia in Patients With Irritable Bowel Syndrome. *Psychosomatic Medicine*, 82(1):29–38, DOI: 10.1097/PSY.0000000000000729.

Kessler, L., Barnhart, H., A.J., B., Choudhury, K., Kondratovich, M., Toledano, A., [...], and Group., Q. T. W. (2014). The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Statistical Methods in Medical Research*, 24:9–26, DOI: https://doi.org/10.1177/0962280214537333.

Koslow, S. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*, 3:863–865, DOI: https://doi.org/10.1038/78760.

Labus, J., Tun, G., Kilpatrick, L., Rao, S., Mayer, E., and Tillisch, K. (2019). Neuroimaging and biomarkers in functional gastrointestinal disorders: What the scientists and clinicians need to know about basic neuroimaging, biomarkers, microbiome, gut and brain interactions. In Rao, S., Lee, Y., and Ghoshal, U., editors, *Clinical and Basic Neurogastroenterology and Motility*, chapter 3, pages 31–54. Elsevier.

Larsson, M., Tillisch, K.and Craig, A., Engström, M., Labus, J., Naliboff, B., [...], and Walter, S. (2011). Brain responses to expectation and delivery of a visceral stimulus in IBS reflect visceral sensitivity thresholds. *Gastroenterology*, 142:463–472, DOI: 10.1053/j.gastro.2011.11.022.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nat Rev*, 521, DOI: 10.1038/nature14539.

Lee, J., Jun, S., Cho, Y.-W., Lee, H., Kim, G., Seo, J., and Kim, N. (2017). Deep Learning in Medical Imaging: General Overview. *Korean J Radiol*, 18(4):570–584, DOI: 10.3348/kjr.2017.18.4.570.

Li, J., Pan, P., Song, W., Huang, R., Chen, K., and Shang, H. (2012). A meta-analysis of diffusion tensor imaging studies in amyotrophic lateral sclerosis. *Neurobiol Aging*, 33:1833–1838, DOI: 10.1016/j.neurobiolaging.2011.04.007.

Li, X., Morgan, P. S., Ashburner, J., Smith, J., and Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, 264:47–56, ISSN: 1872-678X, DOI: 10.1016/j.jneumeth.2016.03.001.

Liu, W., Yang, J., Burgunder, J., Cheng, B., and Shang, H. (2016). Diffusion imaging studies of huntington's disease: A meta-analysis. *Parkinsonism Relat Disord*, 32:94–101, DOI: 10.1016/j.parkreldis.2016.09.005.

Lundervold, A. and Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on MRI. *Zeitschrift für Medizinische Physik*, 29:102–127, DOI: https://doi.org/10.1016/j.zemedi.2018.11.002.

Marques, T. M., Holster, S., Wall, R., König, J., and Brummer, R. J. (2016). Correlating the Gut Microbiome to Health and Disease. In Hyland, N. and Stanton, C., editors, *The Gut-Brain Axis. Dietary, Probiotic, and Prebiotic Interventions on the Microbiota*, chapter 12, pages 261–291. Elsevier.

Mayer, E., Aziz, Q., Coen, S., Kern, M., Labus, J., Lane, R., Kuo, B., Naliboff, B., and Tracey, I. (2009). Brain Imaging Approaches to the study of Functional GI Disorders: A Rome Working Team Report. *Neurogastroenterol Motil*, 21(6), DOI: 10.1111/j.1365-2982.2009.01304.x.

Mayer, E. A., Labus, J., Aziz, Q., Tracy, I., Kilpatrick, L., [...], and Borsook, D. (2019). Role of Brain Imaging in disorders of brain-gut interaction: a Rome Working Team Report. *Gut*, 68(9):1701–1715, DOI: 10.1136/gutjnl-2019-318308.

Mayer, E. A., Labus, J. S., Tillisch, K., Cole, S. W., and Baldi, P. (2015). Towards a systems view of IBS. *Nature reviews. Gastroenterology & hepatology*, 12(10):592–605, ISSN: 1759-5045, DOI: 10.1038/nrgastro.2015.121, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5001844/.

Mukhtar, K., Nawaz, H., and Abid, S. (2019). Functional gastrointestinal disorders and the gut-brain axis: What does the future hold? *World Journal of Gastroenterology*, 25:552–566, DOI: 10.3748/wjg.v25.i5.552.

Mungovan, K. and Ratcliffe, E. M. (2016). Influence of the Microbiota on the Development and Function of the "Second Brain" - The Enteric Nervous System. In Hyland, N. and Stanton, C., editors, *The Gut-Brain Axis. Dietary, Probiotic, and Prebiotic Interventions on the Microbiota*, chapter 19, pages 403–421. Elsevier.

Nair, V. and Hinton, G. (2010). In *Rectified Linear Units Improve Restricted Boltzmann Machines*, volume 27, pages 807–814.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9868–9872, ISSN: 0027-8424, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC55275/.

Pawela, C. and Biswal, B. (2011). Brain Connectivity: A New Journal Emerges. *Brain Connect*, 1(1):1–2, DOI: 10.1089/brain.2011.0020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, DOI: 10.1109/TMI.2016.2538465.

Pesapane, F., Codari, M., and Sardanelli, F. (2018). Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp*, 2(35), DOI: 10.1186/s41747-018-0061-6.

Quigley, E. (2018). The Gut-Brain Axis and the Microbiome: Clues to Pathophysiology and Opportunities for Novel Management Strategies in Irritable Bowel Syndrome (IBS). *Journal of Clinical Medicine*, 7(8):1–8, DOI: 10.3390/jcm7010006.

Raschka, S. (2015). *Python Machine Learning*. Packt Publishing, Birmingham, UK, ISBN: 978-1-78355-513-0.

Reuter, M., Schmansky, N., Rosas, H., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61(4):1402–1418, DOI: 10.1016/j.neuroimage.2012.02.084.

Ruffle, J., Frøkjær, J., and Farmer, A. (2017). Neuromaging of Viceral Pain. In Saba, L., editor, *Neuroimaging of Pain*, pages 341–374. Springer, Cham, ISBN: 978-3-319-48046-6.

Schmulson, M. J. and Drossman, D. A. (2017). What Is New in Rome IV. *J Neurogastroenterol Motil*, 23(2):151–163, DOI: 10.5056/jnm16214.

Soares, J., Marques, P., Alves, V., and Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Front Neurosci*, 7(31), DOI: https://doi.org/10.3389/fnins.2013.00031.

Sporns, O. (2011). *Networks of the Brain*. MIT Press, Cambridge, Massachusetts, ISBN: 978-0-262-01469-4. OCLC: ocn551342282.

Sporns, O. (2013). Structure and function of complex brain networks. *Dialogues in clinical neuroscience*, 15(3):247–262.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

Stejskal, E. and Tanner, J. (1965). Spin Diffusion Measurements: Spin Echoes in the Presence of a Time Dependent Field Gradient. *J. Chem. Phys.*, 42:288–292, DOI: 10.1063/1.1695690, https://mriquestions.com/uploads/3/4/5/7/34572113/stejskal_and_tanner1965.pdf.

Sullivan, D., Obuchowski, N., Kessler, L., Raunig, D., Gatsonis, C., Huang, E. ., and RSNA-QIBA Metrology Working Group (2015). Metrology Standards for Quantitative Imaging Biomarkers. *Radiology*, 277(3):813–825, DOI: 10.1148/radiol.2015142202.

Szczepankiewicz, F., Lasič, S., van Westen, D., Sundgren, P. C., Englund, E., Westin, C.-F., Stålberg, F., Lätt, J., Topgaard, D., and Nilsson, M. (2015). Quantification of microscopic diffusion anisotropy disentangles effects of orientation dispersion from microstructure: Applications in healthy volunteers and in brain tumors. *NeuroImage*, 104:241–252, DOI: 10.1016/j.neuroimage.2014.09.057.

Tack, J., Stanghellini, V., Mearin, F., Yiannakou, Y., Layer, P., Coffin, B., Simren, M. Mackinnon, J., Wiseman, G., Marciniak, A., and Group, I.-C. S. (2019). Economic burden of moderate to severe irritable bowel syndrome with constipation in six European countries. *BMC Gasteroenterol*, 19(69), DOI: 10.1186/s12876-019-0985-1.

Tillisch, K., Mayer, E. A., and Labus, J. S. (2011). Quantitative Meta-analysis Identifies Brain Regions Activated During Rectal Distension in Irritable Bowel Syndrome. *Gastroenterology*, 140(1):91–100, ISSN: 00165085, DOI: 10.1053/j.gastro.2010.07.053, https://linkinghub.elsevier.com/retrieve/pii/S0016508510011510.

Timmers, I., Roebroeck, A., Bastiani, M., Jansma, B., Rubio-Gozalbo, E., and Zhang, H. (2016). Assessing Microstructural Substrates of White Matter Abnormalities: A Comparative Study Using DTI and NODDI. *PLoS ONE*, 11(12), DOI: 10.1371/journal.pone.0167884.

Uddin, L., Nomi, J., Hebert-Seropian, B., Ghaziri, J., and Boucher, O. (2017). Structure and function of the human insula. *J Clin Neurophysiol*, 34(4):300–306, DOI: 10.1097/WNP.0000000000000377.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, 16(1):55–61, ISSN: 1471-003X, 1471-0048, DOI: 10.1038/nrn3857, http://www.nature.com/articles/nrn3857.

van den Heuvel, M. P. and Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, ISSN: 0924977X, DOI: 10.1016/j.euroneuro.2010.03.008, https://linkinghub.elsevier.com/retrieve/pii/S0924977X10000684.

Vandvik, P., Lydersen, S., and Farup, P. (2006). Prevalence, comorbidity and impact of irritable bowel syndrome in Norway. *Scand J Gastroenterol*, 41(6):650–656, DOI: 10.1080/00365520500442542.

Weaver, K., Sherwin, L., Walitt, B., Melkus, G., and Henderson, W. (2016). Neuroimaging the brain-gut axis in patients with irritable bowel syndrome. *World J Gastrointest Pharmacol Ther*, 7(2):320–330, DOI: http://dx.doi.org/10.4292/wjgpt.v7.i2.320.

Winklewski, P., Sabisz, A., Naumczyk, P., Jodzio, K., Szurowska, E., and Szarmach, A. (2018). Understanding the Pathophysiology Behind Axial and Radial Diffusivity Changes - What Do We Know? *Front Neurol*, 9(92), DOI: 10.3389/fneur.2018.00092.

Winston, G. (2012). The physical and biological basis of quantitative parameters derived from diffusion mri. *Quant Imaging Med Surg*, 2(4):254–265, DOI: 10.3978/j.issn.2223-4292.2012.12.05.

Yaakub, S., R.A., H., Keller, S., McGinnity, C., Weber, B., and Hammers, A. (2020). On brain atlas choice and automatic segmentation methods: a comparison of MAPER and FreeSurfer using three atlas databases. *Sci Rep*, 10(2837), DOI: https://doi.org/10.1038/s41598-020-57951-6.

Yuan, Y.-Z., Tao, R.-J., Xu, B., Sun, J., Chen, K.-M., Miao, F., Zhang, Z.-W., and Xu, J.-Y. (2003). Functional brain imaging in irritable bowel syndrome with rectal balloon-distention by using fmri. *World J Gastroenterol*, 9(6):1356–1360, DOI: 10.3748/wjg.v9.i6.1356.

Zwanenburg, A., Leger, S., Vallières, M., and Löck, S. (2019). *Image Biomarker Standardisation Initiative*, https://arxiv.org/abs/1612.07003.