UNIVERSITY OF CASSINO AND SOUTHERN LAZIO

Ph.D. in Methods, models and technologies for engineering

XXXII Cycle

UNICAS
1979>2019
40 ANNI DALLA FONDAZIONE
UNIVERSITÀ DI CASSINO E DEL LAZIO MERIDIONALE

Ph.D. thesis

Application of Machine Learning Techniques to Brain Magnetic Resonance Imaging in Hypertensive Patients

Supervisor	**Ph.D. student**
Prof. Francesco Tortorella	Lorenzo Carnevale

Coordinator
Prof.ssa Wilma POLINI
A.A. 2018/2019

Acknowledgment

I would like to thank my advisor Prof. Francesco Tortorella, which during my doctoral years supported me in the technical choices and in the finalization of this thesis work, improving it with his insightful observation.

I want to thank my mentor Prof. Giuseppe Lembo, which gave me the cultural, technical and human foundation to achieve this result and many others achieved during my collaboration with his research group. Without his support throughout the years I would not be the person I am.

Thanks to my sister Prof. Daniela Carnevale, which is the example I follow since my first steps. I could not ask for a better trail to observe and take as a model. Thanks to all my co-workers and colleagues, supporting this work and many others performed during these years.

Finally, thanks to all my family which supported me in every step of my studies. This support has been the foundation of all my journey, and no word will be good enough to recapitulate my gratitude.

Abstract

Hypertension is one of the main risk factors for vascular dementia and Alzheimer's disease. To predict the onset of these diseases, it is necessary to develop tools to detect the early effects of vascular risk factors on the brain. In this thesis we will tackle the problem of hypertensive brain organ damage characterization approaching it at multiple scales and leveraging multiple techniques. The first part of the thesis will be focused on a deep learning system to perform automatic segmentation of White Matter Hyperintensities (WMH), one of the most common form of macrostructural vascular injury in the brain, on T2-FLAIR imaging. To this aim we will leverage a public dataset and compare our results with the ones achieved in the MICCAI WMH segmentation challenge. The second part of the thesis is focused on the setup of analysis pipelines for Diffusion Tensor Imaging (DTI) and resting state functional MRI (rs-fMRI), with the aim of characterize the microstructural integrity and the functional connectivity. These pipelines have been implemented on hypertensive brains to characterize the subtle brain functional and microstructural damage associated with the hypertensive condition. Finally, both approaches have been implemented in a ongoing research program at IRCCS Neuromed in the context of the heart and brain clinical research, achieving the injury characterization for the first two recruited patients of the study and field-testing the proposed brain injury characterization framework.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION AND BACKGROUND

1.1 Introduction

With the worldwide increase of ageing population, dementia and neurodegenerative diseases became a major issue for healthcare systems that have to ensure adequate management of affected subjects. The estimates by World Health Organization report of more than 35 million people worldwide affected by dementia and is expected a 3-fold increase by year 2050[1]. One of the main modifiable risk factors leading to these pathologies is hypertension[2], a chronic condition affecting about the 30% of adult population worldwide, with incidence growing to more than 60% in the elderly. It is a consolidated knowledge in both clinical practice and pre-clinical research that high blood pressure exerts powerful detrimental effect on a variety of organs such as heart and kidneys, while its effect on cognition and brain health are less known and needs further investigation.

Recent works have leveraged post-mortem analysis tools to shed light on the pathophysiological conditions underlying the symptoms of dementia or cognitive impairment. These works evidenced that more than 50% of dementia cases shown a vascular involvement, as main factor or in combination with neurodegeneration[3]. To empower clinicians and give them proficient tools to understand whether the cognitive decline in process is codetermined by vascular factors, first of all we should investigate the effects on brain structure and functioning of primary vascular risk factors and pathological conditions, as hypertension[4, 5].

To this aim, it is fundamental to correctly stage the entity of the injuries produced by the hypertensive pathology on different brain regions and how the progression of the pathology impacts its structure and functional organization. Magnetic Resonance

Imaging (MRI) has proven itself as an unparalleled tool to investigate these characteristics in vivo in a non-invasive manner and without administering ionizing radiations to the patient. Leveraging different kind of MRI scans, we can obtain various information about the brains, investigating with the appropriate tools different markers corresponding to distinct stages of the pathology.

1.2 Hypertension

Hypertension has been classically defined as systolic blood pressure (SBP) > 140 mmHg and/or diastolic blood pressure (DBP) >90 mmHg, or the use of antihypertensive medication. Hypertension is a complex and systematic pathology, affecting primary the cardiovascular system, with a multitude of secondary organs and functions involved by the effects of long-lasting elevated blood pressure. The hypertension diagnosis and treatment is carried on according to guidelines established and published by two major international workgroups: i) the join American Heart Association and American College of Cardiology Hypertension guidelines (AHA/ACC)[6]; ii) the European Society of Cardiology Hypertension guidelines (ESC)[7]. It is worth noting that while the two workgroups have been substantially concordant on the aforementioned hypertension definition throughout the history, the last revision of the AHA/ACC guidelines has defined as hypertensive also the subject reporting a SBP > 130mmHg and/or a DBP >80 mmHg. This change reflected in a more intensive use of anti-hypertensive treatment, looking for an increased benefit on both well-known and object of research consequences of hypertension[8].

While the end-organ damage induced by hypertension in some districts like heart or kidneys is comprehensively characterized and studied, the effects of hypertension on the brain are largely unknown, with the main contribution identified as a strong risk factor for stroke and Alzheimer’s Disease. It is however becoming clearer thanks to recent mechanistic and epidemiologic studies that hypertension alone can impact cognition and can alter the cerebral homeostasis in a way capable of altering the cognitive performances of hypertensive patients, without the onset of neurological conditions as AD or acute cerebrovascular event[3, 9, 10].

1.3 The brain and the cerebrovascular system

The brain is the main organ of the nervous system, responsible for the perception and cognition. It is a complex organ made of billions of interconnected cells, which can be grouped respect to their anatomical or functional nature. One of the biggest efforts of contemporary neuroscience has been directed to map the differences between the different regions of the brain and to elucidate their distinct role in the cognition processes.

1.3.1 Anatomy of the Brain

The brain is made up of different major parts connected between them and attached to the spinal cord through the brainstem. This serves as a regulator of primary functions through the vagus nerve such as breathing, cardiac rhythm, digestive system, immune system activation. Attached to the brainstem there is the cerebellum, a neuronal structure mainly responsible for receiving sensory input from the peripheral nerves and managing the motor response. At the top of the brainstem we can find the limbic system, made of several substructures which are fundamental in several function as memory, learning and emotional responses.

The most important structure of the brain is the cerebrum, divided into left and right hemisphere interconnected by a central structure of thick neuronal fibers called corpus callosum. In the cerebrum we can distinguish the gray matter (GM) and white matter (WM), with the cerebrospinal fluid (CSF) filling the void spaces called ventriculi between different hemispheres and regions. The gray matter, named also the cortex, is the region composing the surface of the brain, is structured as a layered architecture of neuronal cell bodies interconnected between them on a single layer and between layers. The white matter instead is mainly composed by the neuronal axons connecting different brain regions. The most prominent characteristic of human brain respect to other species is the presence of convolutions, with crests named gyri and the valleys named sulci. This convoluted structure results in a greater surface of gray matter respect to the surface who would be available on a smooth sphere.

The main activity of the brain has been classically identified as cortical activity and with several techniques throughout the history we have been able to map with notable precision the regions of the cortex pairing them with their functions.

White matter, responsible for the intercommunications between distinct subcortical and cortical regions, can be divided in fascicles or bundles, which are organized groups of axons projecting in the same direction. Both grey and white matter mapping and organization will be discussed in further detail in Chapter 3.

1.3.2 The cerebrovascular tree and the neurovascular unit

While the brain is one of the most complex and energy demanding organs of our body, it is not provided with a proper energy storage, relying its functioning only on nutrients continuously coming through the cerebral circulation. Thus, to have a reliable cognitive functioning it is mandatory to get a constant and well-regulated blood flow feeding neuronal cells. To do so, the cerebrovascular tree has developed the capability to regulate its pressure and flow autonomously respect to the general circulation, namely the cerebral autoregulation. Alterations of this property following acute cerebrovascular events, due to sustained increase of blood pressure or the aging process can impact on cognitive performance, making the system less prone to distribute the nutrients from the blood flow equally in all the cerebral regions.

To provide a homogeneous perfusion of the brain tissues, the large arteries branch into progressively smaller vessels, up to the diameter of few microns composing the capillary bed. The complex of cells responsible for the mechanism of isolating the neuronal tissues from the direct blood flow running through the capillary bed is called the NeuroVascular Unit (NVU), mainly maintaining the integrity of the blood-brain barrier (BBB) and regulating the regional cerebral blood flow and oxygen and nutrient delivery[11].

Both large vessels and the NVU are part of the delicate mechanism of brain blood flow regulation, which is heavily stressed in conditions of hypertension and are two of the main target of damage in case of chronic elevated blood pressure, often reflected into an increased permeability of the BBB and other kinds of focal tissue damage, like white matter hyperintensities (discussed in detail in Chapter 2).

1.4 Magnetic Resonance Imaging

To date, the best available tool to explore the human brain is Magnetic Resonance Imaging (MRI). Originated around fifty years ago, this technology leverages the property of hydrogen nuclei which, when stimulated by radiofrequency pulses, resonate with the magnetic field where they are immersed, allowing the analysis of different tissues. By reading the signals emitted during the resonance we can obtain the time needed for a particular tissue to return to steady state on the longitudinal component or on the transverse one, respectively T1 relaxation time and T2 relaxation time. The images are obtained by weighting one of the two components, emphasizing the contrasts between tissues with different T1 or T2 relaxation times. During the decades, MRI techniques evolved from the methods that allowed to obtain images of internal tissues of the patients to techniques capable of providing functional insights of the biological systems under examination.

The support of various MRI techniques can be a fundamental addition to the clinical practice, in order to specifically characterize and diagnose different forms of cognitive impairment originated from vascular pathologies. On this notice, a modern and quantitative approach can be instrumental to extrapolate effective biomarkers for clinical and modern computer-driven analyses.

The first applications of brain MRI were developed to understand and analyse the morphological alterations induced by various pathologies impacting on white and grey matter. The first approach was aimed at obtaining a segmentation of the brain and a parcellation of the cortical areas, by hand or by using specialized software [12, 13]. The data obtained from these elaborations were used to characterize the neurodegenerative processes and put in relation the affected physical areas with the associated cognitive functions[14].

The development of new techniques specific for white matter injury evaluation paved the way to the analysis and identification of one of the most important markers of cerebrovascular damage in the brain: the white matter hyperintensities (WMH). T2-FLAIR sequence (T2-Fluid Attenuated Inversion Recovery, discussed in detail in Appendix A)[15] is useful to highlight regions of T2 prolongation in the white matter, corresponding to regions of increased water content respect to normal white matter. In this kind of sequence, areas of hyperintensity represent a region where the white matter is undergoing a process of demyelination or axonal loss. In general, these alterations are the main evidence of cerebrovascular diseases, even though they can correspond to different pathological states[16].

Characterization of WMH has been first qualitative, with a grading system based on the appearance and position of the identifiable lesions[17], then with the progress of computer-aided diagnosis (CADx) systems and improvements in the computer vision field, we can now absolutely quantify the volume of white matter lesions[18, 19]. This improvement is fundamental to define absolute and quantitative biomarkers that could help in better evaluating and predicting the onset of VCI in the population at risk.

Further techniques of MRI were implemented to describe microstructural and functional alterations in the brain, respectively the diffusion tensor imaging and the functional magnetic resonance imaging. These techniques will be leveraged to describe the hypertensive brains and will be discussed in detail in next chapters of this thesis together with the analysis pipelines necessaries to get significant insight from the raw diffusion and functional data.

1.5 The Role of Machine Learning

Machine Learning is a branch of computer science aimed at designing automated systems capable of automatic learning from examples with minimal human guidance and interference. This results in computer programs capable of absorbing data, build and refine models on the input data to maximize the system capability to predict certain outcomes and perform better decision on new data.

The model adopted for the process of learning divides the ML algorithms in two major categories: supervised and unsupervised methods.

Supervised learning is based on a labelled training data set, in which every sample is associated to a distinct category. The main aim of the algorithm is to achieve the capability to generalize the decision took from the model (the classification) from the training data to unseen data.

Unsupervised learning is based on a heterogeneous data set without any labelling. The aim of the algorithm is to identify commonalities between different samples in the dataset and organize them in homogeneous group, the clusters.

The technological advances which took place in the last decade greatly improved the capabilities of training complex mathematical models with big datasets. Previous limitations were upper bounds in term of high-speed memory, needed to load the model, and the lack of adequate computing capabilities, needed to train complex models with a sufficient amount of data. Graphical Processing Unit (GPU) improvements were key to develop affordable and reliable parallel computing capability, exploiting the intrinsic parallel architecture needed for graphical computations: many processors needed for simple parallel operations to generate and perform calculations on matrix objects (i.e. images).

Neural Networks are one of the first models of machine learning proposed in literature, and while effective have been soon put aside due to the computational complexity of the training process. The model is made up by several primitive elements (the neurons) interconnected between them and whose connections are ruled by a rule which considers both their value and the value of incoming connections. This kind of model gives the flexibility of a common operator which can approximate any nonlinear function, given a particular architecture of connections is provided.

The intrinsic parallel nature of Neural Networks gave it a strong boost with recent technological leap, facilitating the training of architectures composed by several interconnected layer, composing “deep” networks.

1.6 Outline of the work and aims

In this thesis we will split the task of characterizing the brain injury associated with hypertension in two sections, both part of the heart and brain research program carried on at Istituto di Ricovero e Cura a Carattere Scientifico Neuromed by the AngioCardioNeurology and Translational Medicine department. The first one will tackle the segmentation of white matter hyperintensities, macroscopical lesions which are evidenced by MRI scans performed in the clinical routine, hallmark of advanced stage hypertension brain damage[20, 21]. To do so, we will leverage a publicly available dataset comprising scans from different machines and clinical units in Europe. The second part will be centred on characterizing by advanced neuroimaging sequences the damage exerted on brain microstructure and functional organization by hypertension, in an early stage of pathology to design a multimodal biomarker which could be a candidate for prediction of cognitive decline[9]. This study will be conducted on patients recruited in our clinical unit, to set up all the neuroimaging pipelines to obtain a comprehensive microstructural and functional characterization of the brain. At this stage, it will be used to evaluate whether it is possible to predict the presence of a cardiovascular pathology only from neuroimaging analysis.

In both sections we will apply machine learning strategies, to leverage the amount of data generated from these experimental procedures and to automate the time-consuming part of labelling and quantification of lesion, with the final aim to produce a framework to comprehensively characterize the damage in the brains of hypertensive patients, from the earliest stage of it to identify the set of markers which can identify patients at risk of cognitive decline in the later stage of the pathology to automated tools to provide fast and reliable segmentation of hypertensive lesions in the brain to stratify in an efficient way the hypertensive population for their brain damage.

Our final aim will be to implement both strategies on patients recruited in our clinical unit in the context of an ongoing prospective study, to characterize patients in a longitudinal way to predict the onset of hypertension associated cognitive impairment from macrostructural, microstructural and functional data. Due to the ongoing recruiting process, it will be only shown the application of the complete pipeline to two example cases.

CHAPTER 2. A DEEP LEARNING APPROACH FOR THE SEGMENTATION OF WHITE MATTER HYPERINTENSITIES

2.1 Introduction

The term “Deep Learning” and the concept of deep neural networks (DNN) was introduced in 2006 by Hinton in a seminal paper in which he showed a multilayer forward network in which every layer is pretrained as a Restricted Boltzmann Machine (RBM) (Figure 1) and then combined with fine tuning by a supervised back propagation[22].

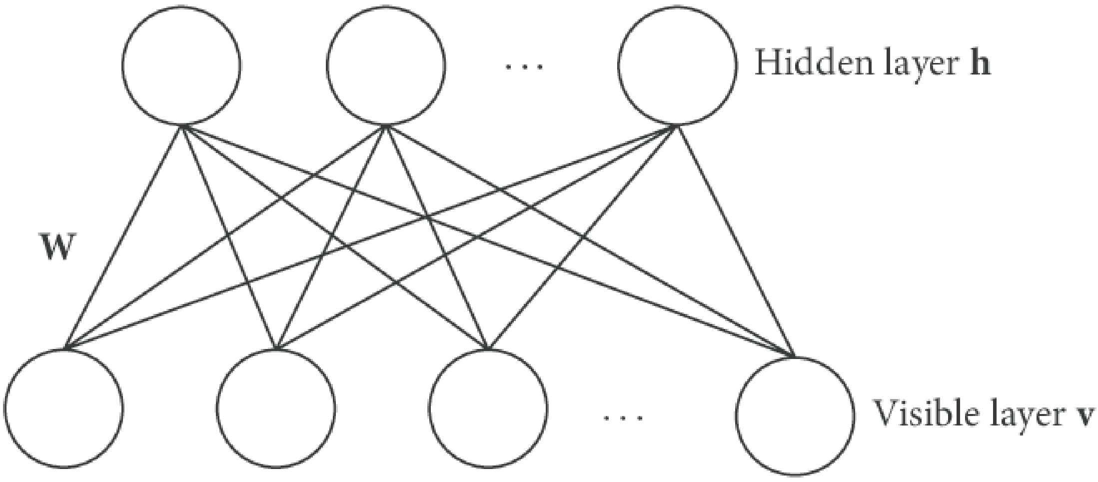


Figure 1- Restricted Boltzmann Machine Model

After this first experiments an increasing number of problems were tackled by deep learning, especially in the context of computer vision and image processing with the Convolutional Neural Network (CNN) architecture. The capability of CNNs of efficiently solving problems with results unmatched by other ML techniques soon captured the global attention of the computer vision and pattern recognition world, generating dozens of different architectures to solve different problems of different context.

The first successful applications of DNN has certainly been the image processing and recognition context: the ImageNET competition[23], a benchmark problem for image recognition, has seen a spike of the classification performance with the DNNs, from a 25% error percentage in 2010 with classical ML algorithm to a 4% classification error in 2015[24], implementing a complex CNN and achieving a performance superior to the human. Other fields of application which saw dramatic improvements with DNN are notably the Natural Language Processing (NLP), in which different kind of architectures have greatly improved the capability of machine to understand and process human written or spoken language, with great breakthroughs achieved in text sentiment analysis or speech recognition.

Medical imaging also saw great improvements from the implementation of deep learning approaches. Computer-aided diagnosis or and computer-aided detection (CADx) have always been one of the main research field in biomedical and computer engineering[25], with deep learning these systems have reached notable performances

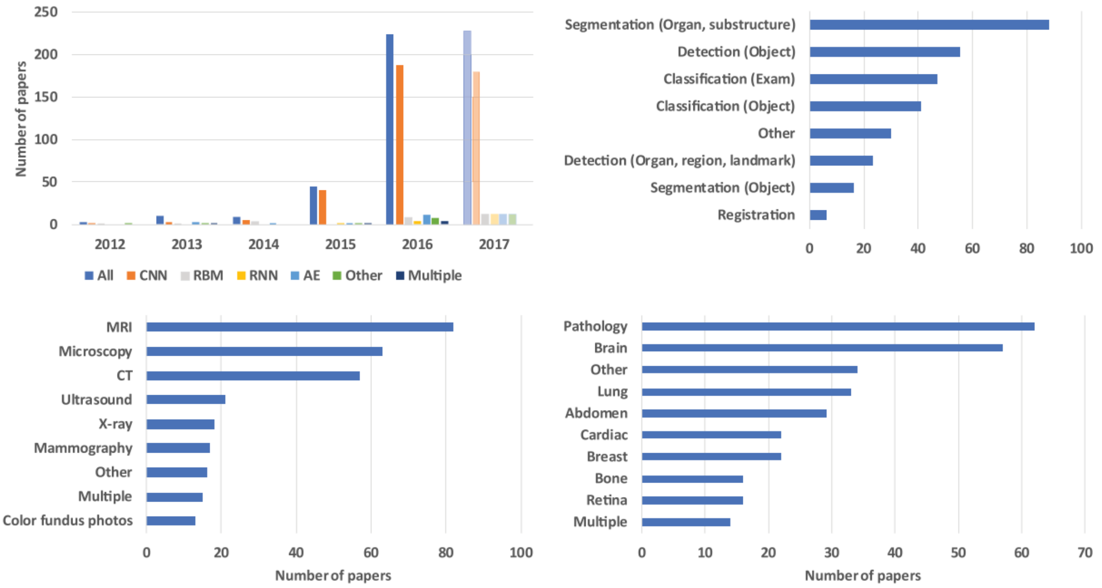


Figure 2- Occurrences of deep learning research papers in medical imaging

in selected fields of application, like a CADx system for melanoma diagnosis which outperformed trained physicians in recognizing and grading melanomas from moles pictures[26].

Mimicking the capillary diffusion that deep learning had in several computer science, the medical imaging field has seen a sudden surge in the number of systems and research papers which implement deep learning strategies. As seen in Figure 2, from Litjens et al.[27], the breakthrough seen in 2012-2013 brought a 20x times appearance of deep learning techniques first in workshops and conferences, then in peer-reviewed journals. The breakdown of the task addressed in the papers and the imaging modalities in exam show that segmentation task is the most prominent one, often paired with MRI modality (see [27, 28] for comprehensive review). The most common model of NN is the Convolutional Neural Network, given its strong performance with image recognition and pattern recognition.

In my research experience I focused on neuroimaging analysis, with images obtained by various techniques of Magnetic Resonance Imaging (MRI). The first problem tackled with deep learning was the automatic segmentation of white matter lesions evidenced by T2-FLAIR imaging.

2.2 MRI applied to evaluate macrostructural brain damage: White Matter Hyperintensities

MRI is a powerful tool for diagnosis of many different pathologies and the in-vivo inspection of internal organs. Thanks to the high contrast achieved between different soft tissues has established as the exam of choice for brain imaging, leveraging the different magnetic properties of gray matter, white matter and cerebrospinal fluid. During the last 50 years of clinical use of the MRI capabilities expanded greatly, and one of the first advanced sequences adopted in the routine examination has been the T2 Fluid Attenuated Inversion Recovery (FLAIR) (see Appendix A – MRI sequences for more information).

This kind of technique has been developed to image the brain weighting in T2 while attenuating the signal generated from the cerebrospinal fluid. In this way, while gray and white matter show a low contrast between them, lesions of the white matter presented as spot of hyperintensity are evident. While the aetiology of these marks is still debated and can be ascribed to multiple factor like massive axon demyelination, ischemic damage, hypoxia, reduced glial presence, there are strong epidemiological associations with conditions such as vascular dementia, vascular cognitive impairment and small vessel disease. Standard clinical grading of these injuries is based on their shape, extension and location, characteristics which are synthesized with a single score (Fazekas Score)[29]. In Figure 3 are reported two examples of T2 Flair showing low grade WMH and severe WMH in two different patients.

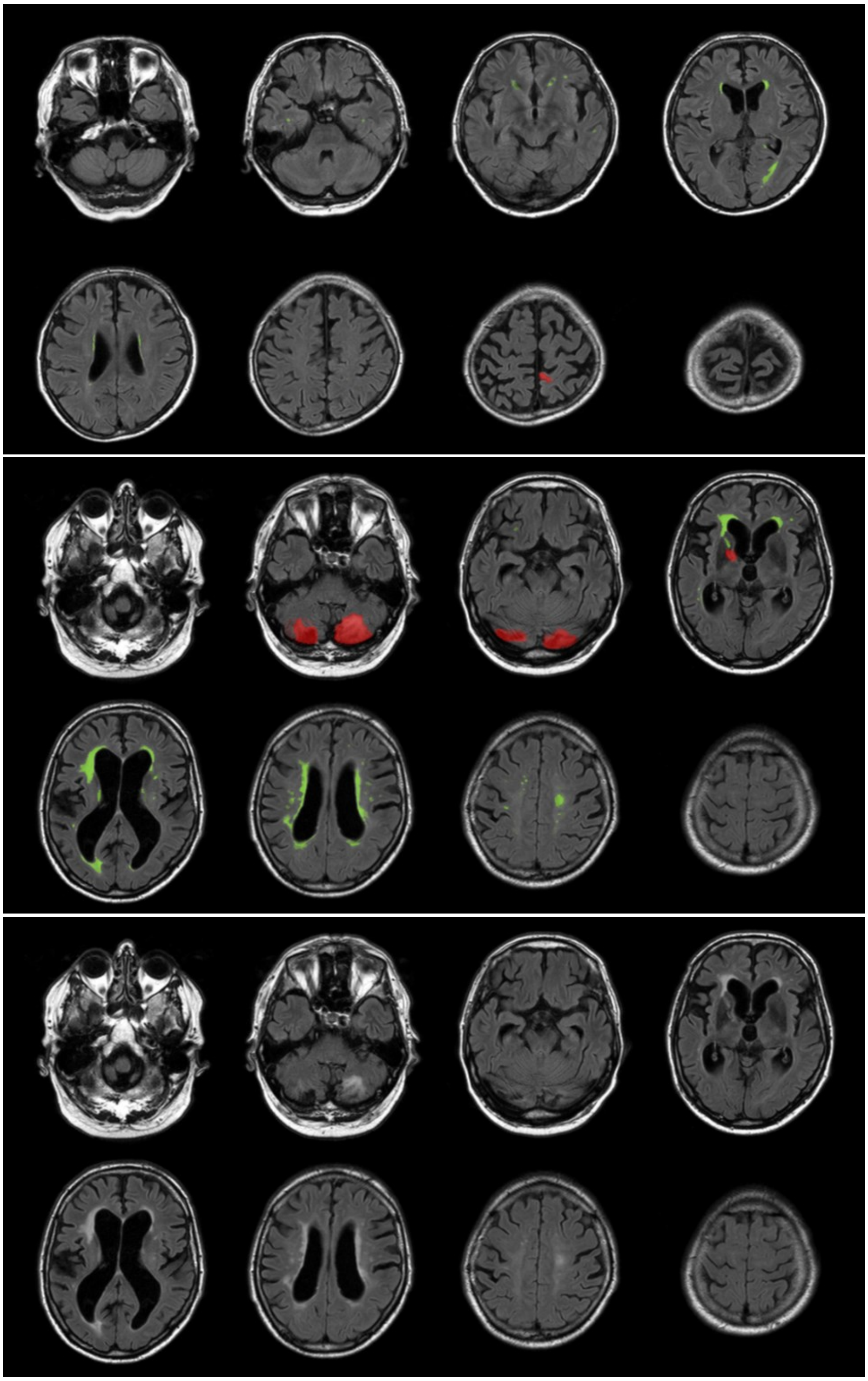


Figure 3 – Example of T2 Flair images with segmented WMH (in green) and lesions of other nature (Red). Top row: scan of a patient with a low grade WMH (Fazekas 1); Middle row: scan of a patient with severe WMH (Fazekas 4); Bottom Row: same image without segmentation overlays

Given the wide spectrum of disorders associated with White Matter Hyperintensities (WMH), it would be of utmost importance have a repeatable and objective measurement of the lesions, to better stratify the population and use the lesion volume as a proper damage biomarker.

To this aim, we will leverage the Deep Learning technologies to perform automated segmentation of T2-FLAIR brain scans to extract an absolute quantification of WMH. To do this, we will leverage a dataset collected for MICCAI 2017 – White Matter Hyperintensity Segmentation Challenge.

2.3 Methods

2.3.1 Dataset Characteristics

The MICCAI 2017 WMH Segmentation Challenge dataset is composed of brain scans of 60 patients obtained on 3 different scanners from different vendors, 20 for each scanner. Every patient has both 3D-T1 weighted and T2-FLAIR images hand labelled for WMH and different lesions[30]. Lesions have been segmented according to the standard procedures identified in the STRIVE (STandards for ReportIng Vascular changes on nEuroimaging). FLAIR images were acquired according to three different protocols, depending on the clinical unit and scanner vendor:

- UMC Utrecht – 3T Philips - 2D FLAIR sequence (48 transversal slices, voxel size: 0.96×0.95×3.00 mm3, TR/TE/TI: 11000/125/2800 ms)
- NUHS Singapore – 3T Siemens - 2D FLAIR sequence (transversal slices, voxel size: 1.0×1.0x3.00 mm3, TR/TE/TI: 9000/82/2500 ms)
- VU Amsterdam – 3T GE - 3D FLAIR sequence (132 saggital slices, voxel size: 0.98×0.98×1.2 mm3, TR/TE/TI: 8000/126/2340)

3D FLAIR images were reoriented to transversal orientation and resampled to a slice thickness of 3.00 mm, according to the 2D FLAIR specifications.

The images were pre-processed for bias correction by the dataset curators using SPM 12 and the T1 image was co-registered to the T2-FLAIR with elastix.

2.3.2 Brain Extraction

In T2 FLAIR imaging the WMH appears as bright spot in the middle of grey appearing white and grey matter, making possible to highlight the lesions. Skull surrounding skin however appear as bright as WMH, making hard to distinguish it from the lesions, especially in slices comprising almost only skin. Thus, to avoid this confounding element we extracted the brain to eliminate all the tissues not useful to learn the characteristic patterns of WMH lesions. To do so, we leveraged the widely used Brain Extraction Tool, part of one of the main neuroimaging analysis suite (FSL). In all the subsequent steps of the pipeline were all executed on masked images, as in Figure 4. All brain extraction results were quality checked to ensure a correct brain segmentation and to ensure that all the WMH lesions were inside the extracted region.

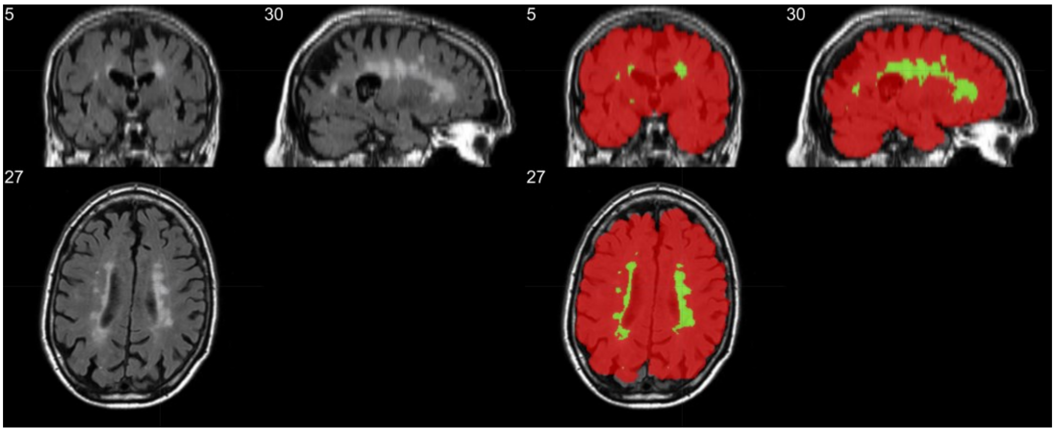


Figure 4 – Left: T2 Flair showing similar intensity between WMH and skin. Right: mask of the extracted brain in red, with WMH segmentation in green.

2.3.3 Framework

All the experiments were implemented in Python 3.6, leveraging the Tensorflow (1.13) libraries with Keras (2.3.6) functional APIs, installed on a workstation with Ubuntu 18.06 OS. The workstation was equipped with an Intel 9980XE, 64GB of RAM and a Nvidia Titan RTX, built with 576 Tensor Cores and 24GB of GDDR6 RAM.

On top of this setup, all the experiments were implemented leveraging MisCNN, a framework designed for medical image segmentation. In Figure 5 is depicted the schematics of the framework implementations, with a series of tool to load the neuroimaging data (Nifti I/O interface), preprocess data (normalization, patch extraction, one hot encoding), data augmentation (flips, rototranslation), built-in models and the possibility to expand the library with custom designed models, training and evaluation validation. Some minor implementation have been performed to modify the framework to implement functionalities as 2D patch extraction from 3D volumes.

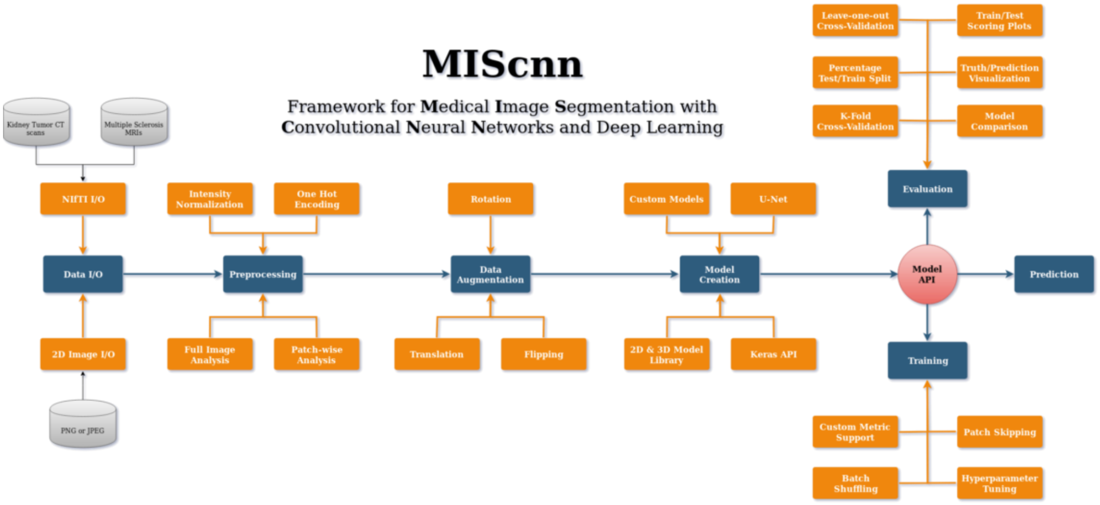


Figure 5 - Schematics of MIScnn framework implementation

2.3.4 Preliminary Experiments

In order to leverage three-dimensional shape information provided by volumetric scans of the brain, we first implemented a 3D segmentation network. The experiments were carried on 3D patches extracted from the whole volume and applied a 3D implementation of the networks discussed in 2.3.6. The results were discouraging, with Dice Scores for the WMH segmentations in the order of 0.1. Our hypothesis is that the anisotropy of the images (1x1x3 mm approximatively for the three different scanners) does not add significant information to the segmenting network while the 3D structure has a significantly increased number of weights to train, resulting in a more complex training process. In Figure 6 is shown the loss trend of the best-performing 3D segmentation network.

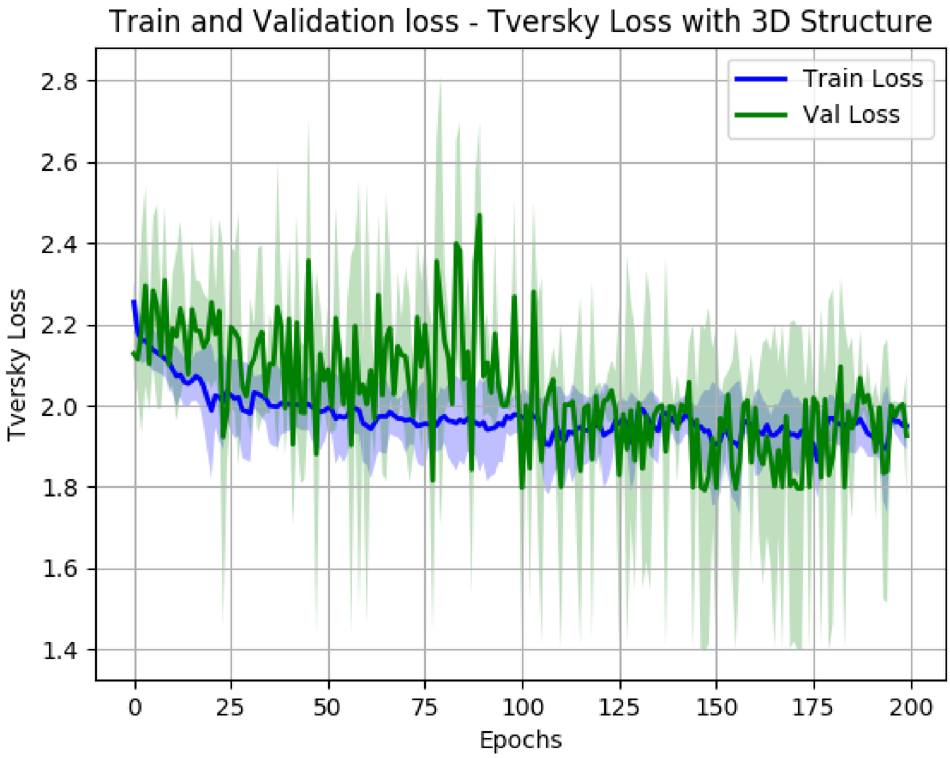


Figure 6 - 3D Train and Validation loss

Thus, we focused the subsequent experiments on the segmentation of planar patches extracted on the X-Y plane. In Figure 7 we can see an example of 3D patch used in preliminary experiments and an example of patch used in subsequent experiments.

2.3.5 Network Architectures

To tackle the problems of medical images semantic segmentation the most common approach has been U-Net and variations, since its first implementation in 2015[31]. The basic principle of U-Net is to structure a network coupling one encoder or convolutional arm which acting as feature extractor and a decoder or deconvolutional arm which leverages the extracted features to provide a dense segmentation of the input image. In this experiment we are implementing two different variations of this

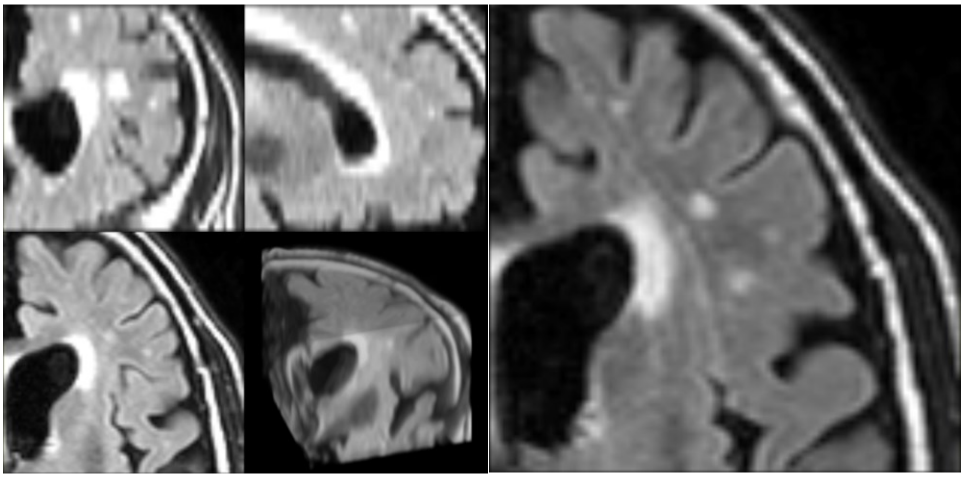


Figure 7 - Comparison between 3D patch and 2D patch

architecture, previously published in the context of semantic segmentation[32]: Unet3 and FractalNet.

U-Net3is a extension of the original architecture with focus on the training with a reduced number of samples. The contracting path is a typical convolutional network architecture, with a sequence of two 5x5 convolutions followed by a ReLU and a 2x2 max pooling with stride of 2 to perform down-sampling. After this down-sampling the features are doubled. The expansive path is made of up-sampling blocks followed by a 2x2 up-convolution, halving the features and performing a concatenation with the

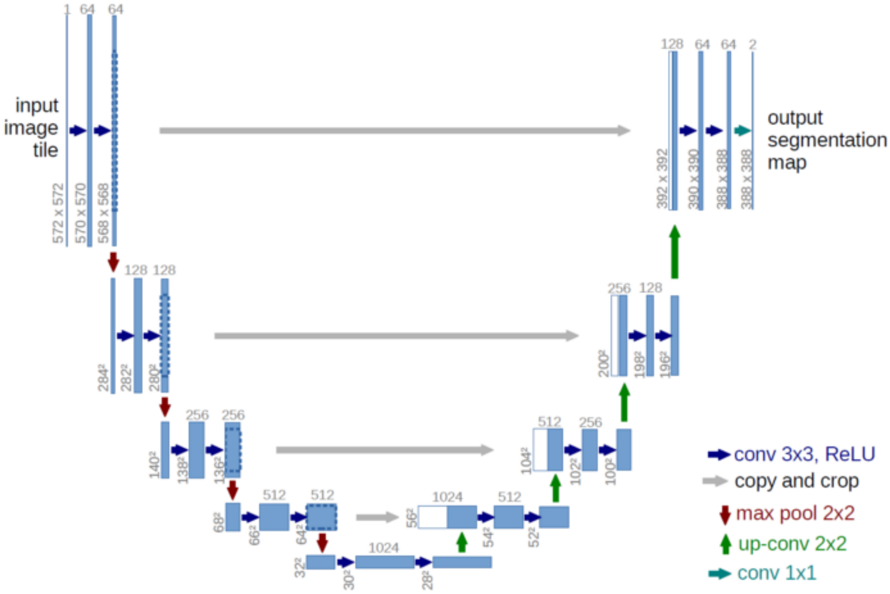


Figure 8 – U-NET standard architecture

corresponding feature map obtained from the contractile path. The output of this layer is followed by to 5x5 convolutions and a ReLU. The output layer of the network maps the final feature vector to the output number of classes. This architecture extends the original UNet by adding a batch normalization after every ReLU activation function, after each max-pooling and up-sampling layer.

FractalNet is an architecture introduced by Larsson et al. [33]. This networks aims at exploit at a convolutional layer features from different visual levels, joining them to enhance the discriminative capabilities of the network. To this aim, let define C as the index of a truncated fractal $f_C(\cdot)$ and the base case of this fractal defined as a single convolutional layer: $f_1(z) = conv(z)$. Defining the expansion rule as:

$$z' = conv(z)$$

$$f_{C+1}(z) = conv(conv(z') \oplus f_C(z'))$$

It can be defined recursively for successive fractals, with $\oplus$ as a generic join operation and *conv(·)* as the convolution operator. The $\oplus$ operation can be summation, maximization or concatenation, as in this specific architecture. In order to enlarge the receptive field and enclose more contextual information, down-sampling and up-sampling operations are added in the above expansion rule. In particular, a max pooling with a stride of 2 and a deconvolution also with a stride of 2 are added. After the down-sampling operation, the receptive field of a fractal becomes broader. When combining different receptive fields through the join operation, the network can harness multi-scale visual cues and promote itself in discriminating.

2.3.6 Training

Data Augmentation was carried on in order to increase the number of training samples, adding them by applying a combination of transforms to the original dataset. In this case, we applied 3 cycles of data augmentation applying mirroring, rotations and scaling to the original images. Data augmentation is applied at image level, not to the sliced patches. Mirroring was enabled on every image axis, rotations were constrained in a random angle between -15 and +15 degrees, random scaling was applied between 0.85 and 1.25.

Training was performed with in batches, with batch size of 128 patches. Networks weights were initialized with a normal distribution for the bias and a glorot normal function for the weights[34]. Three different loss were tested in this experiment: the soft Dice coefficient loss, calculated as one minus the average dice coefficient of each segmented class in a multiclass problem; the soft Dice loss + the categorical crossentropy loss; the Tversky loss, with alpha and beta coefficients = 0.5, which is a loss function for multiclass segmentation with fully convolutional deep networks [35]. The optimization of the weights have been performed with an Adam optimizer [36], with a learning rate of 0.01, $\beta1 = 0.9$ and $\beta2=0.999$. The training process was repeated for 200 epochs, with a reduction of a ten factor in learning rate after 30 epochs without loss improvement. All experiments were performed with a 5-fold cross validation.

2.3.7 Evaluation Metrics

The evaluation of results was carried on according to the metrics considered in the WMH Segmentation Challenge, using the provided script. The five considered metrics were:

- Dice Score, calculated as $DS = 1 - dice\ dissimilarity$, with dice dissimilarity defined in scipy package
- Hausdorff Distance, 95th percentile, defined as the greatest distance from a point of one set from the closest point in another set. This parameter is calculated between the segmented lesions and ground truth boundaries.
- Average volume difference, expressed in absolute value percentage.
- Sensitivity for individual lesion, defined as the number of detected lesions divided by the number of true lesions.
- F1-score for individual lesion, defined as $2 * \frac{precision * recall}{precision + recall}$, with both precision and recall parameters calculated on the number of identified connected components in the results.

Individual lesions are identified and defined as single 3D connected components.

2.4 Results

2.4.1 Raw Data

Our first experiment was carried on raw data, which underwent only the brain extraction procedure. We compared the two different architectures presented in 2.3.5 trained with the losses discussed in 2.3.6. The discrepancy between the loss metric evaluated on the sliced patch extracted from the samples and the parameters for the global evaluation of the system shown in 2.3.7, we decided to save a checkpoint of the trained model every 20 epochs of training, to analyze how the evaluation metrics on the whole sample behaved respect to the loss calculated on the patches extracted from the sample. This analysis revealed that all the metrics show a concordant trend with the patch-based loss measured during the validation phase, with an example shown in Figure 9. We can show that the best model in the validation phase corresponds to optimal results in the image-based validation.

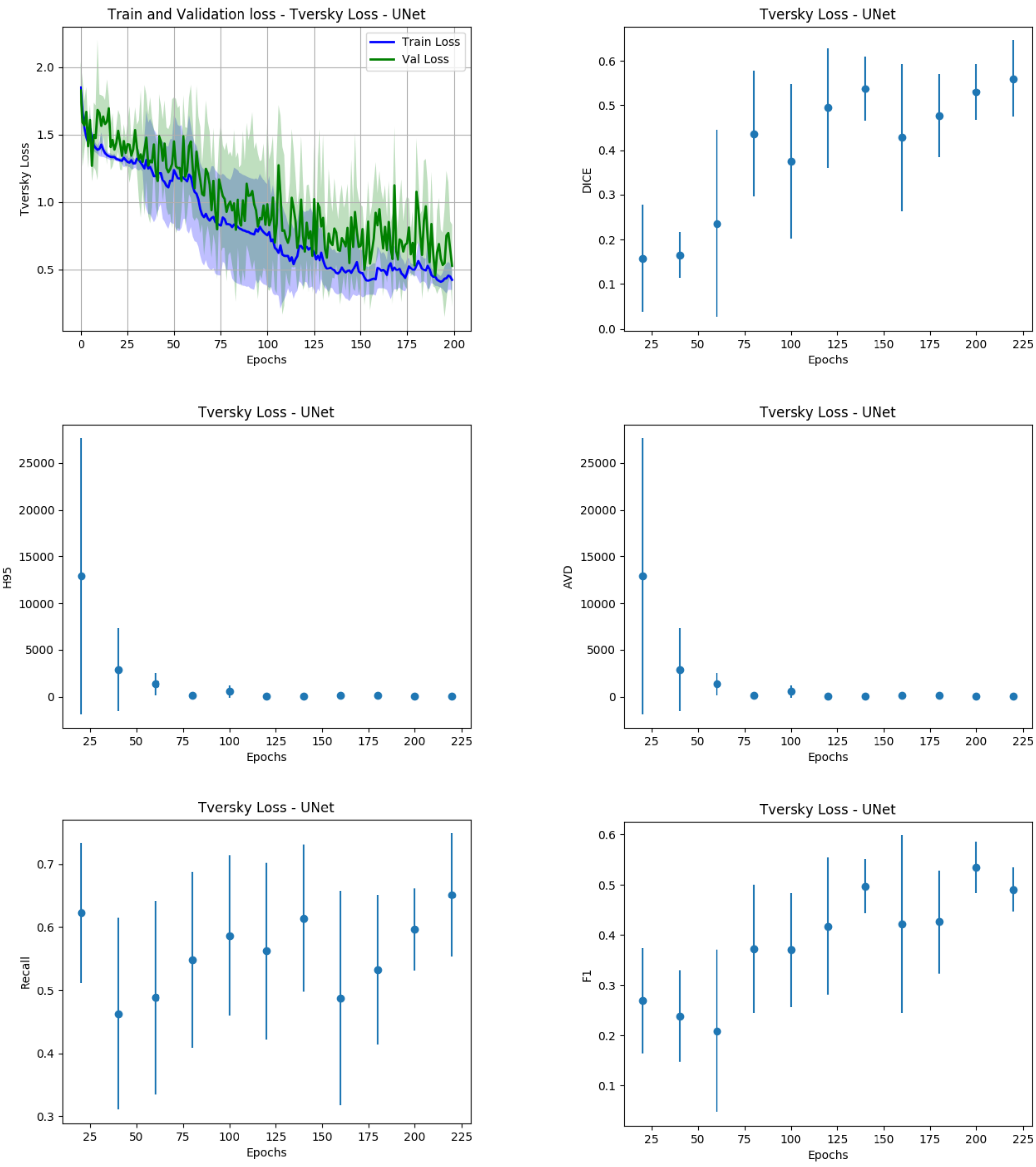


Figure 9 - Variation of final evaluation metrics with epochs

All the combinations of architectures and losses in this setting show non optimal results. To further enhance the predictive performances of the models and improve the generalization capability of the system we proposed to leverage an ensemble of networks, by combining the predictions of the 5 models obtained in the 5-fold validation, proposing the final result as the average of the 5 outcomes. In Table 1 we report the obtained results and we evaluate the performance of our system in the ranking of the WMH segmentation challenge, assuming that the system will be able to generalize the segmentation process to images obtained by different scanners.

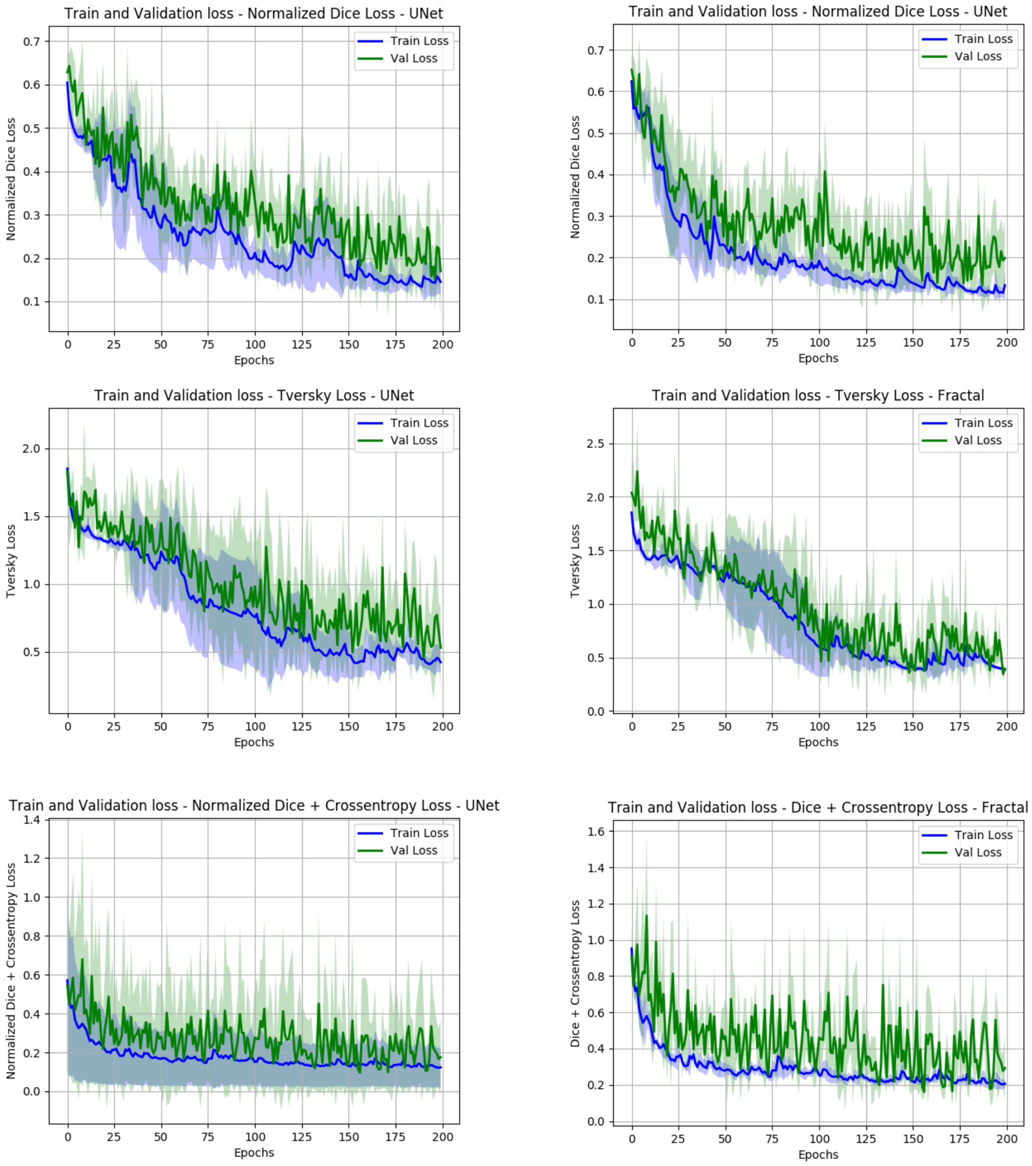


Figure 10 - Loss trend with raw data. Left column: Unet, Right column: Fractal Net

Architecture	**Loss**	**Dice**	**H95**	**AVD**	**Recall**	**F1**	**Projected ranking**
Unet	Normalized Dice	0.604	17.53	41.10	0.663	0.596	0.252 – 24th
	ND + CrossEntropy	0.416	22.63	64.55	0.430	0.373	0.485 – 37th
	Tversky	0.608	18.67	40.23	0.674	0.606	0.249 – 24th
FractalNet	Normalized Dice	0.427	30.42	140.9	0.592	0.441	0.461 – 36th
	ND + CrossEntropy	0.346	32.57	333.3	0.584	0.398	0.561 – 37th
	Tversky	0.559	27.05	108.8	0.730	0.520	0.520 – 37th

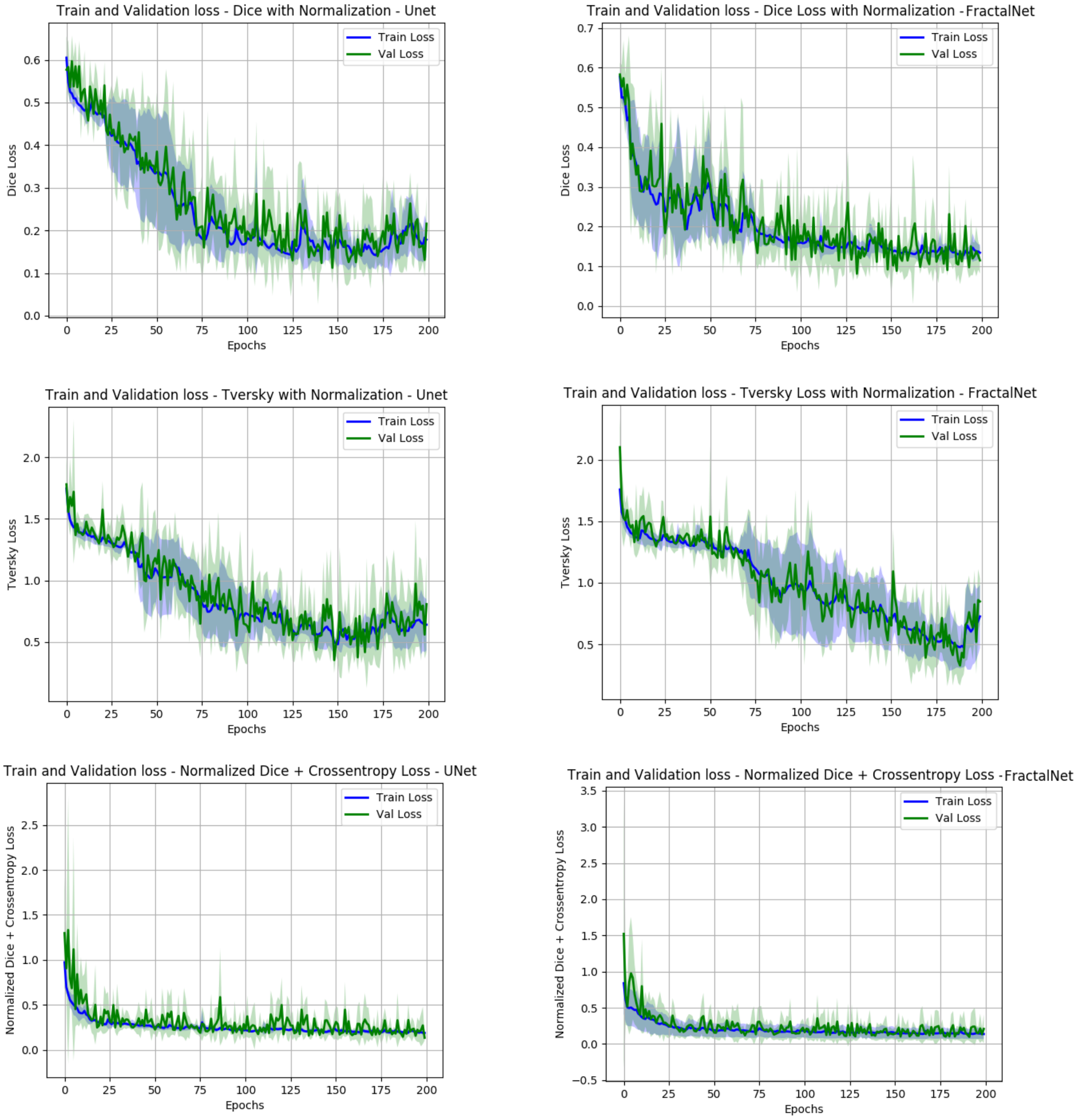
Table 1 - Results for Raw Data

2.4.2 Normalized Data

To improve the segmentation performances of the system we implemented a basic preprocessing strategy, the data normalization. As implemented by MIScnn framework normalization is carried on the whole image, before splitting it up in patches. The performances of all the networks greatly improved with this strategy, as shown in Table 2 and Figure 11. In Figure 12 and Figure 13 we show respectively a comparison of all the networks results against the ground truth and the segmentation obtained with the model with the best projected ranking.

Architecture	**Loss**	**Dice**	**H95**	**AVD**	**Recall**	**F1**	**Projected ranking**
Unet	Normalized Dice	0.755	14.33	29.52	0.729	0.618	0.155 – 20th
	ND + CrossEntropy	0.738	**9.878**	27.85	0.599	0.647	0.175 – 22nd
	Tversky	0.754	14.78	29.60	0.693	0.611	0.171 – 22nd
FractalNet	Normalized Dice	0.779	10.51	**21.70**	0.763	**0.693**	0.096 – 13th
	ND + CrossEntropy	0.745	17.40	42.03	**0.802**	0.600	0.157 – 22nd
	Tversky	**0.780**	14.21	22.96	0.777	0.658	0.118 – 18th

Table 2 - Results for Normalized Data



Train and Validation loss - Dice with Normalization - Unet

Train and Validation loss - Dice Loss with Normalization -FractalNet

Train and Validation loss - Tversky with Normalization - Unet

Train and Validation loss - Tversky Loss with Normalization - FractalNet

Train and Validation loss - Normalized Dice + Crossentropy Loss - UNet

Train and Validation loss - Normalized Dice + Crossentropy Loss -FractalNet

Figure 11 - Loss trend with normalized data. Left column: Unet, Right column: Fractal Net

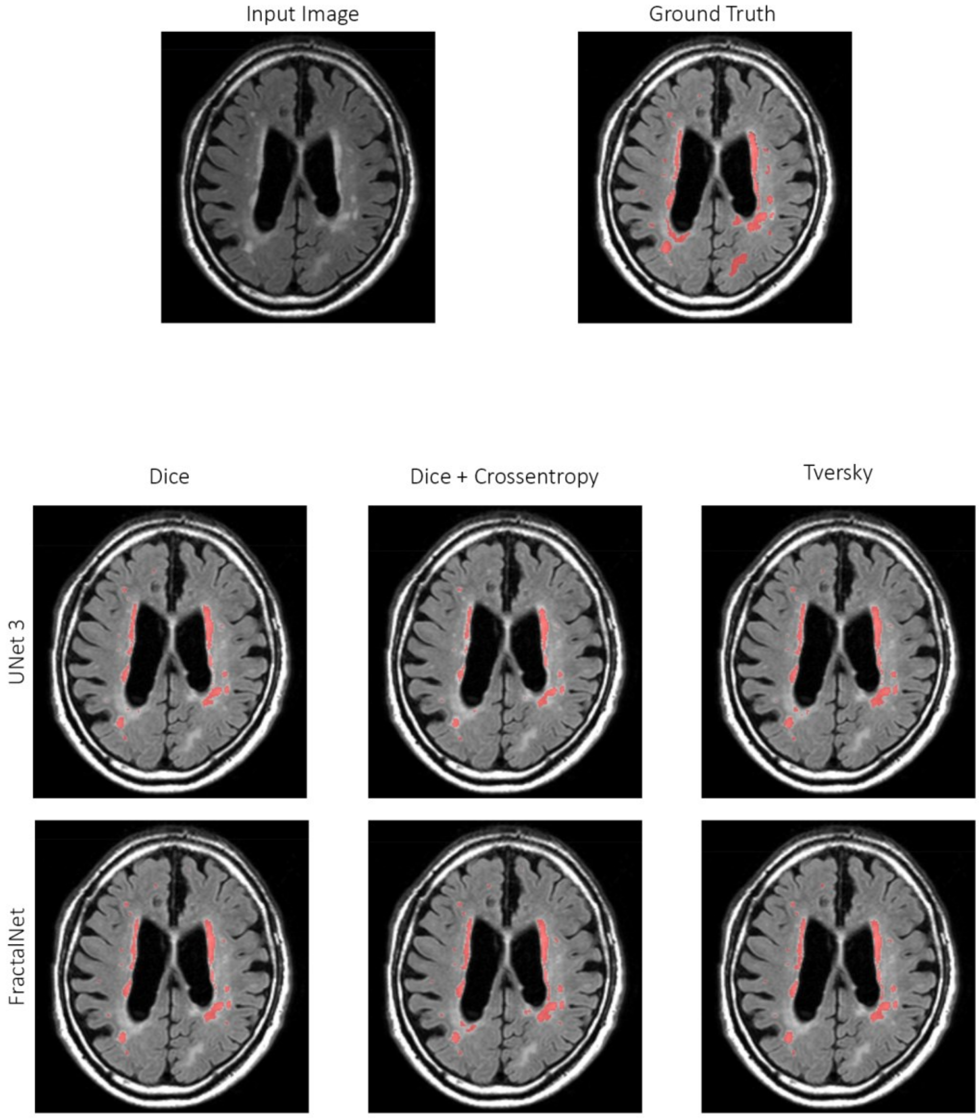


Figure 12 - Segmentation results of all the implemented models on normalized data

2.5 Discussion

With this work we designed an end to end system to segment white matter hyperintensities in T2-FLAIR imaging leveraging fully convolutional networks. The segmentations evaluation describe a system which is capable of achieving a performing score in terms of dice similarity coefficient and average volume difference

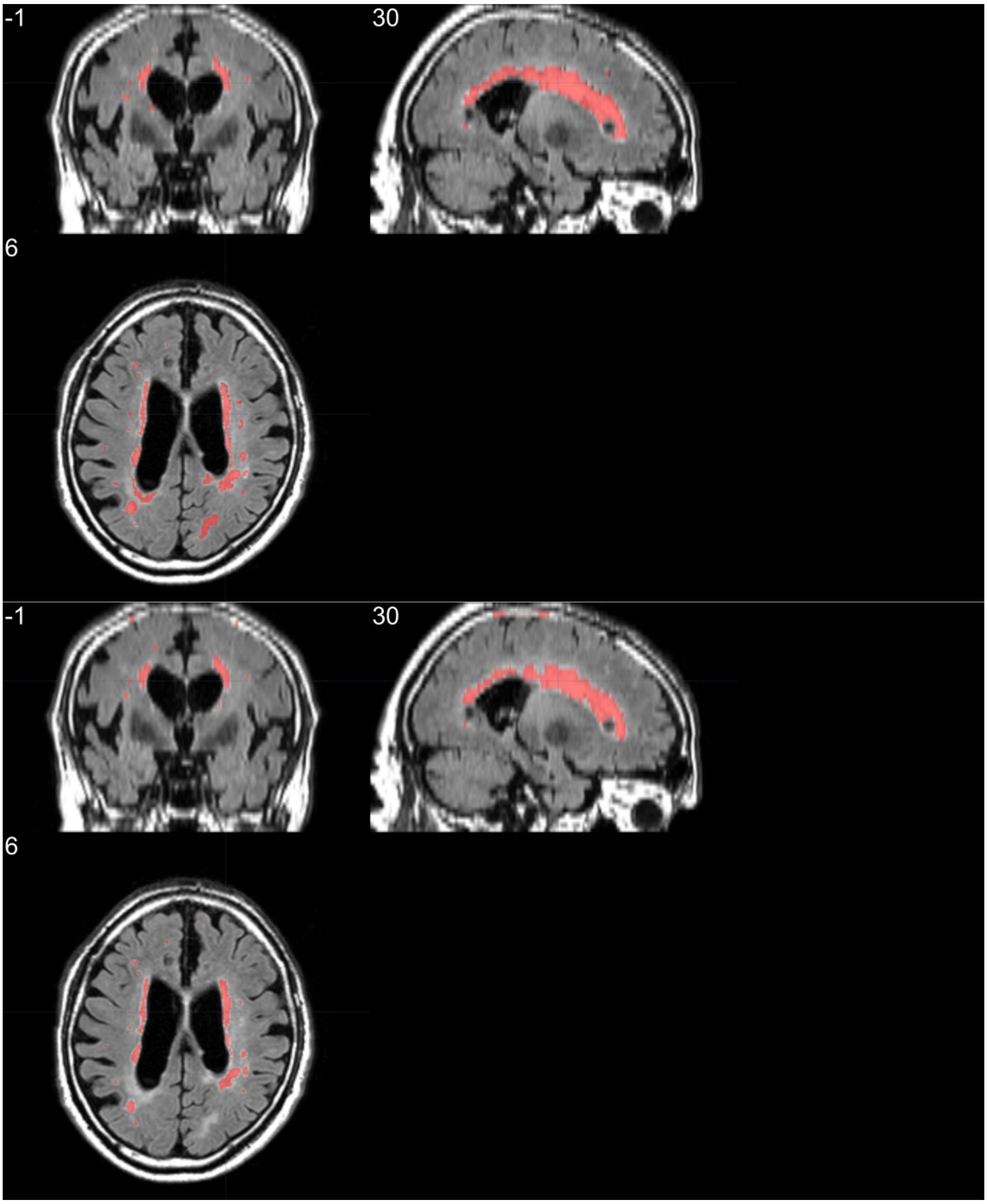


Figure 13 - Triplane comparison between ground truth (top row) and segmentation results of our proposed system (bottom row)

percentage, while suffering from a low F1-score calculated on the single lesions. This parameter in particular is suggesting the tendency of the network to segment as WMH lesions areas originally labelled as other lesions. Moreover, a visual inspection of the segmentation suggested us that our models show a tendency to underestimate the lesions, as shown in Figure 12 and Figure 13, where smaller isolated lesions are not recognized and where bigger lesions have smaller boundaries. To improve this issue we are currently implementing WMH segmentation leveraging multimodal imaging, to enhance the discrimination between WMH and other lesions using also T1-weighted images, where other lesions have other distinctive patterns to identify them.

Observing the loss per epoch in Figure 11 we can notice the very stable trend showed by Normalized Dice + Crossentropy loss, which reaches the plateau of learning very early in the process, without significant improvements till the end. This suggest us that this is the best performance we could obtain with this combination of loss and network. The Tversky loss showed a better learning trend, while more unstable across different folds. However, the best results and the best projected ranking was achieved by Dice Normalized Loss. This experiment showed the most stable learning trend and reached the best F1- score on the single lesions, the most impacting evaluation on the projected rankings of our networks. The future works will be focused on this combination of architecture and loss: the first step will be the inclusion of a second channel on the contractile path to take into account also the T1-weighted images, other works will be directed to add preprocessing steps to enhance the quality of T2-FLAIR images. Preliminary results have been obtained with image histogram equalization, with drastic improvements on some images showing contracted histograms, while no impact or performance degradation on images with well-distributed histograms.

Finally, it is ongoing the building of the official evaluation docker to test in real field the proposed architectures and obtain an objective ranking in the WMH segmentation challenge.

CHAPTER 3. HYPERTENSIVE PATIENTS AND EARLY BRAIN DAMAGE: A STRUCTURAL AND FUNCTIONAL CONNECTIVITY CHARACTERIZATION

3.1 Introduction

The WMH segmentation is a fundamental task to obtain a quantification of the white matter lesion burden, being a representative marker of brain injury associated to ongoing cognitive decline and later-stage hypertension. Thus, it's nature of established damage phenotype make it useful to stage the degree the pathology, not a good candidate to be a predictive biomarker of early brain damage and vascular cognitive impairment.

To that aim, in our work we looked for alternate parameters to be assessed on hypertensive patients in a early stage of brain damage, in which no damage is diagnosed by a routine clinical MRI exam. We designed a prospective study in which this class of patients, recruited in our outpatient facility at I.R.C.C.S. Neuromed, underwent standard clinical assessments, echocardiography, cognitive assessment and advanced neuroimaging.

3.1.1 Patient Sample

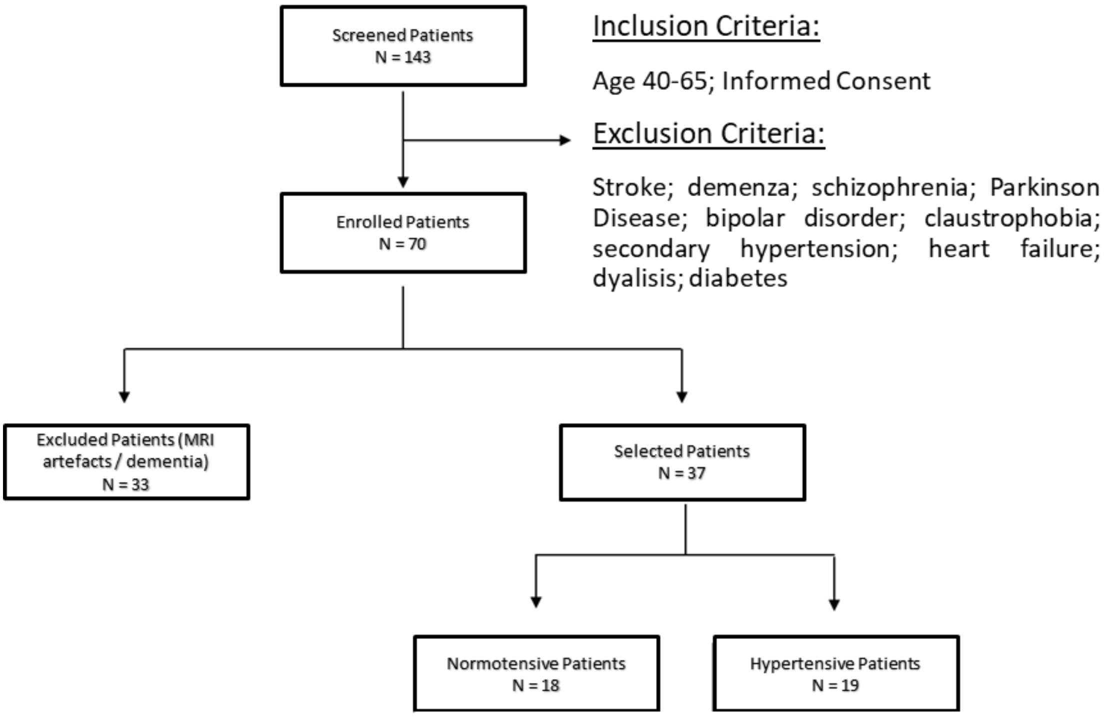


Figure 14 - Enrollment Flowchart

To stage the hypertensive disease, we subjected patients to standard organ damage assessment for hypertension target organs, such as vasculature, heart and kidneys. In Figure 14 it is shown the flowchart of the enrolment process of the study, with the main inclusion criteria and the breakdown of excluded patients. The hypertensive population show a hypertrophic cardiac remodelling, characterized by the thickening of heart walls, without loss of function, evidenced by the Ejection Fraction. This suggests a process of adaptive remodelling of the heart, excluding heart failure. No significant vasculature remodelling is evidenced, as shown by the Intima-Media Thickness (IMT) of the carotids. It is worth noting the absence of atherosclerotic plaques in the carotid tract, which could independently contribute to cognitive impairment. Renal function is still unaltered in hypertensive patients, excluding chronic kidney disease (Table 3). The chosen sample represent a homogeneous sample with early damages in organs target of chronic elevated BP levels and no failure in any of those organs. Thus, is a candidate population to explore whether hypertension exerts a damage on the brain and how we can assess it in an early stage, when no sign of WMH or structural damage is present.

Sample Characteristics	**Normotensive** **n= 18**	**Hypertensive** **n=19**	***p* value**
Demographic			
Age. mean (SD)	52 (8)	55 (7)	0.256
Sex. number of females (percentage)	10 (55.55%)	10 (52.63 %)	0.863
Smokers. number (percentage)	3 (16.66%)	4 (21.05%)	0.742
BMI. mean (SD)	25.38 (4.76)	29.96 (4.50)	**<0.01
Blood Pressure			
Systolic Blood Pressure - mmHg mean (SD)	123 (9.19)	138 (10.97)	***<0.001
Diastolic Blood Pressure - mmHg mean (SD)	77 (6.24)	87 (9.71)	***<0.001
Cardiac Remodeling			
LV end-diastolic diameter - mm. mean (SD)	4.85 (0.30)	4.96 (0.33)	0.257
IV septal thickness- mm. mean (SD)	0.91 (0.12)	1.16 (0.14)	***<0.001
LV posterior wall thickness- mm. mean (SD)	0.91 (0.14)	1.09 (0.09)	***<0.001
LV mass index ($LVMI^{2.7}$) - g/m^2. mean (SD)	37.69 (9.11)	54.66 (8.28)	***<0.001
Relative wall thickness - RWT. mean (SD)	0.37 (0.05)	0.44 (0.04)	***<0.001
Diastolic dysfunction (percentage)	4 (22.22%)	15 (78.94%)	***<0.001
LV Ejection fraction - %. mean (SD)	65.44 (6.20)	68.63 (7.03)	0.153
Carotid Artery (CA) Thickening			
Internal CA (right) - IMT. mean (SD)	0.76 (0.18)	0.86 (0.23)	0.137
Common CA (right) - IMT. mean (SD)	0.80 (0.14)	0.87 (0.13)	0.118
Internal CA (left) - IMT. mean (SD)	0.76 (0.17)	0.87 (0.21)	0.082
Common CA (left) - IMT. mean (SD)	0.81 (0.13)	0.90 (0.18)	0.094
Renal damage			
Creatinine - mg/dL. mean (SD)	0.75 (0.17)	0.74 (0.15)	0.848
Microalbuminuria - mg/24hrs. mean (SD)	11.84 (16.08)	17.30 (17.96)	0.338
Estimated GFR – mL/min. mean (SD)	114.84 (34.95)	127.12 (41.50)	0.338

Table 3 - Patient Sample Characteristics

Cognitive tests revealed that hypertensive population show a significantly reduced MoCA score (Montreal Cognitive Assessment, the gold standard for vascular cognitive impairment diagnosis) (Table 4)[37]. The damage is focused on executive function, as can be seen in Table 5 in which are reported the values obtained in different subscales of MoCA test, to better differentiate the cognitive domains with impaired function.

	Normotensive **n= 18**	**Hypertensive** **n=19**	***p* value**
Cognitive Assessment			
IADL – score. mean (SD)	7.7 (0.75)	7.7 (0.65)	0.950
MoCA – score, mean (SD)	26.00 (2.43)	22.36 (2.73)	***<0.001
Semantic Verbal Fluency – score. mean (SD)	48.67 (11.44)	43.68 (12.24)	0.210
Paired-Associate Learning – score. mean (SD)	13.92 (4.04)	9.55 (5.31)	**<0.01
Stroop Color Word Test – score. mean (SD)	0.22 (0.65)	0.97 (1.51)	0.059
Stroop Interference Test – time. mean (SD)	16.68 (7.90)	25.99 (12.11)	**<0.009

Table 4 - Cognitive Assessment

	Normotensive **n= 18**	**Hypertensive** **n=19**	***p* value**
MoCA cognitive subdomains			
Visuospatial – score. mean (SD)	3.2 (1)	3.35 (1)	0.680
Executive Functions – score, mean (SD)	3.6 (0.7)	2.4 (1.2)	**<0.01
Language – score. mean (SD)	5.3 (0.8)	4.9 (0.7)	0.216
Attention – score. mean (SD)	5.2 (1.3)	4.7 (1.5)	0.324
Memory – score. mean (SD)	2.9 (1.7)	1.4 (1.3)	**<0.01

Table 5 - MoCA cognitive subdomains

3.2 Structural Connectivity: Diffusion Tensor Imaging applied to Hypertensive Patients

Our first aim in this study was to characterize the microstructural damage that hypertension exerts on cerebral white matter, the region of the brain which connects different areas of the cerebral cortex between them and with the subcortical structures. Damages to this region have classically been associated and diagnosed with the WMH, lacking methods to investigate its integrity in early stages of damage. To this aim, the use of Diffusion Tensor Imaging (DTI) is a powerful tool. Briefly, by analysing the Brownian motion of the water in the brains we can reconstruct a model of the neuronal fibers, thus evidencing the connections existing in the brains and their microstructural properties[38, 39]. (see Appendix A – MRI Sequences for more insights)

3.2.1 Diffusion parameters

To fully characterize the diffusion at each voxel the first measure extracted is the *Mean Diffusivity* (**MD**), which is the sum of the diagonal elements of D, is a measure of the magnitude of diffusion and is a rotationally invariant measure[40]. From MD we can thereby extract the most widely used measure of anisotropy, the *Fractional Anisotropy* (**FA**) [40].

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

Note that the diffusion anisotropy does not describe the full tensor shape or distribution. This is because different eigenvalue combinations can generate the same values of FA. Although FA is likely to be adequate for many applications and appears to be quite sensitive to a broad spectrum of pathological conditions, the full tensor shape cannot be simply described using a single scalar measure. However, the tensor shape can be described completely using a combination of spherical, linear and planar shape measures. The last two measures are secondary combination or amplitude evaluation of the eigenvalue: the *Axial Diffusivity* (**AD**) [40], which is the greatest eigenvalue (λ_1) and the *Radial Diffusivity* (**RD**) [40] which is the average of the others

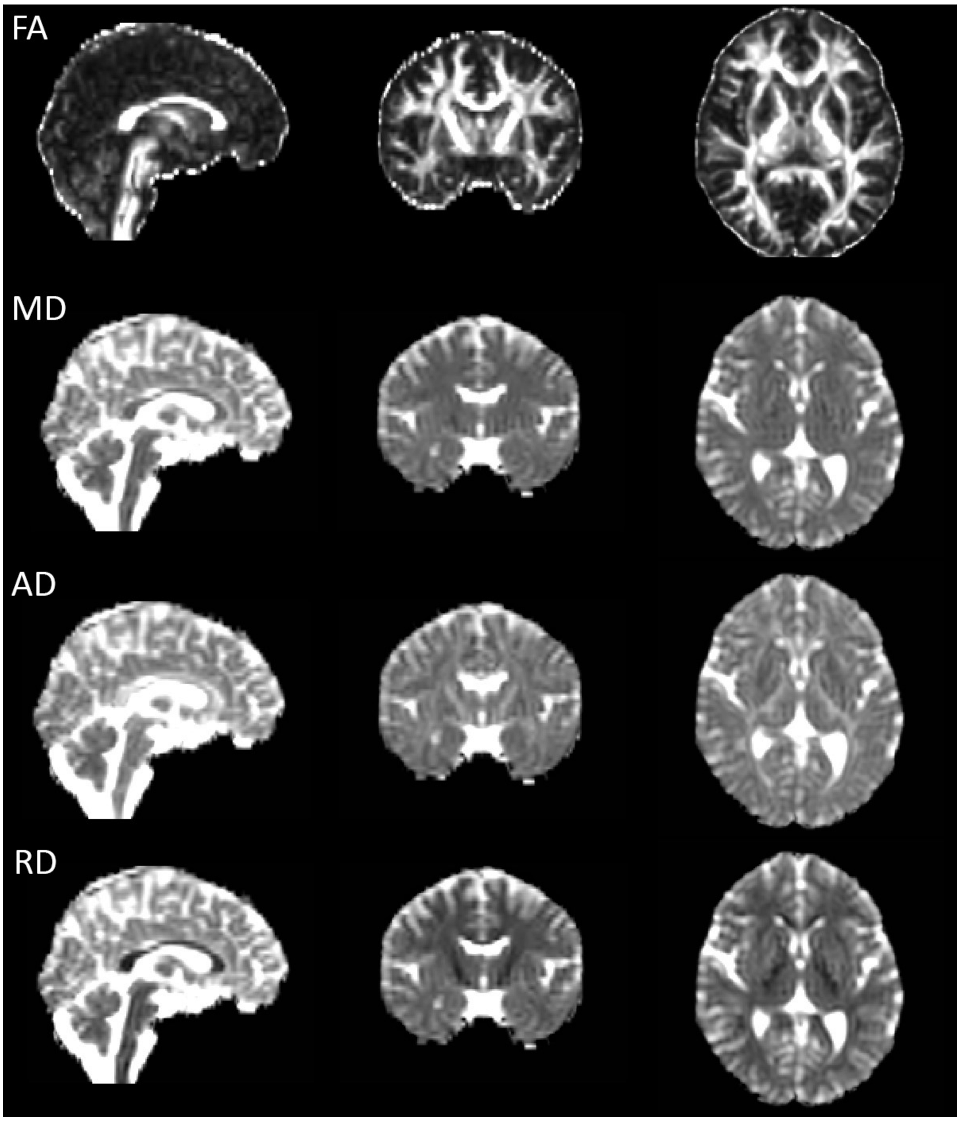


Figure 15 - Diffusion Parameters Map

eigenvalues ($\frac{\lambda_2+\lambda_3}{2}$). An example of the maps is shown in Figure 15. The scalar maps can't represent the principal directions of the diffusion. A standard color coding, superimposed to an FA map, has been vastly used to represent the principal eigenvector in each voxel (Figure 16). The green represents diffusion on the Posterior-Anterior direction, red represents diffusion on the Left-Right direction, blue represents diffusion on the Upper-Lower direction.

Every parameter's alteration can be associated to a specific type of damage of the analysed area: FA and MD are associated with primary axon degeneration. Lower FA values indicate disorganized fibers, which are affected by microstructural processes such as demyelination, axonal degradation, or gliosis. Of the two, MD, is a more

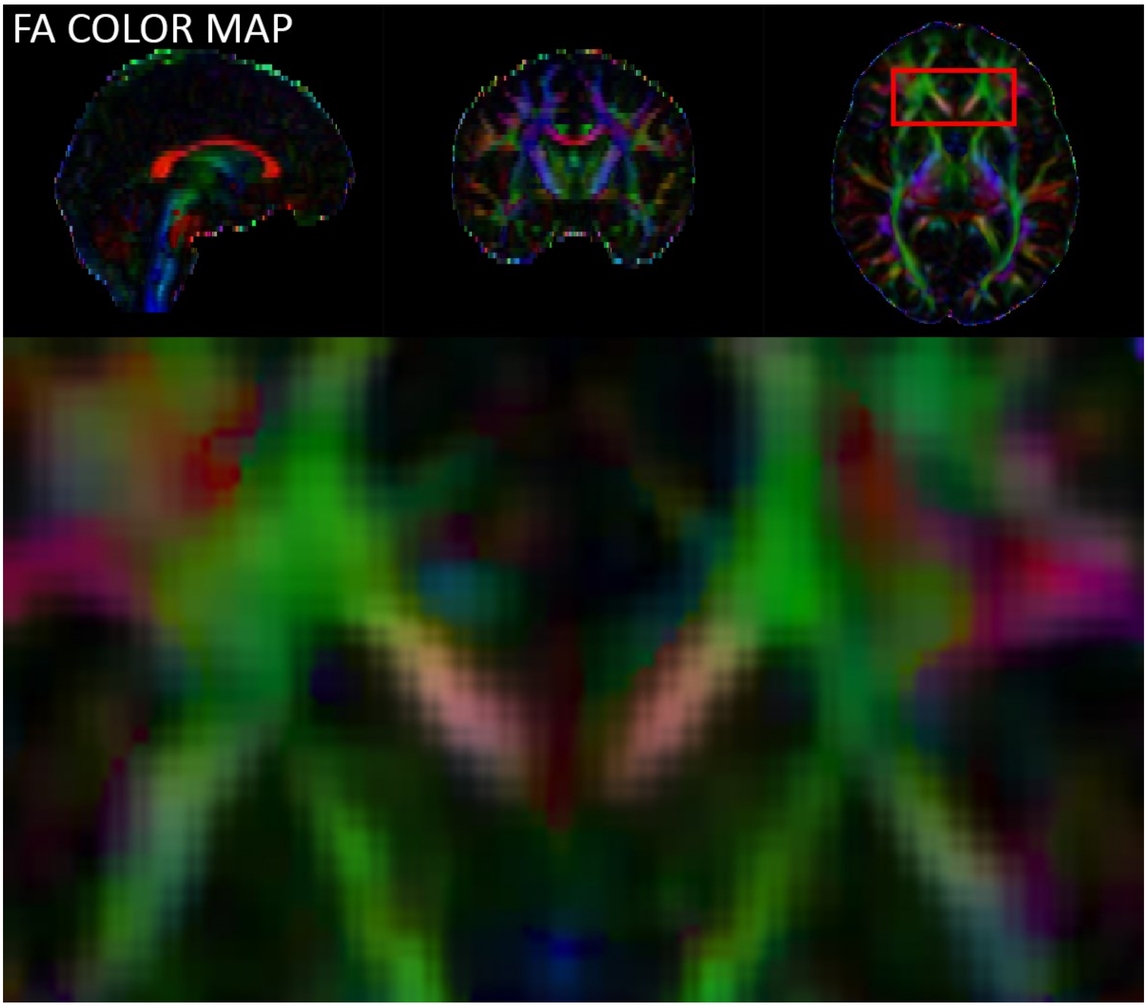


Figure 16 - Fractional Anisotropy Color Map

sensitive measure even though it is less specific, and its results can be increased by any pathological process affecting the neuronal cell membranes. Incremental variations in RD are associated with myelin breakdown, whereas in AD describe secondary axon degeneration.

3.2.2 From tensorial model to probabilistic diffusion modelling

The main limitation of DTI deterministic methods can be found in the simplistic modelling of a single fiber per voxel. White matter fibers are microscopical structures and brain networks are intricate and it is often needed the connection between distant areas. The typical spatial resolution of a MRI dataset is 1x1x1 mm, so one voxel is very likely to contain more than a bundle of fiber. To overcome this limitation a probabilistic approach has been developed to model the crossing fibers in each voxel: BEDPOSTX (Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques. X stands for Crossing Fibers)[41]. This technique exploits the partial volume model (also called ball and stick model) which assumes that a fraction of diffusion is along a single dominant direction, and that the remainder is isotropic. The algorithm also estimates how many fibers can be modelled in each voxel of the space, creating a multi-fiber distribution parameter estimation. The output of BEDPOSTX is a series of scalar maps per estimated fiber. For each fiber we have the estimates of theta, phi (polar coordinates identifying the main diffusion direction) and anisotropic volume fraction. These maps are necessary to perform probabilistic tractography. The information from multiple fibers can be superimposed to have a global view of the modelled fibers distribution. In Figure 17 the modelling of the first fiber, in red, the second fiber, in blue, and their superimposition is shown. In detail the zoom of the green rectangle: it’s possible to see how, in high anisotropy regions, the models are mixed and the two fibers point out different diffusion directions.

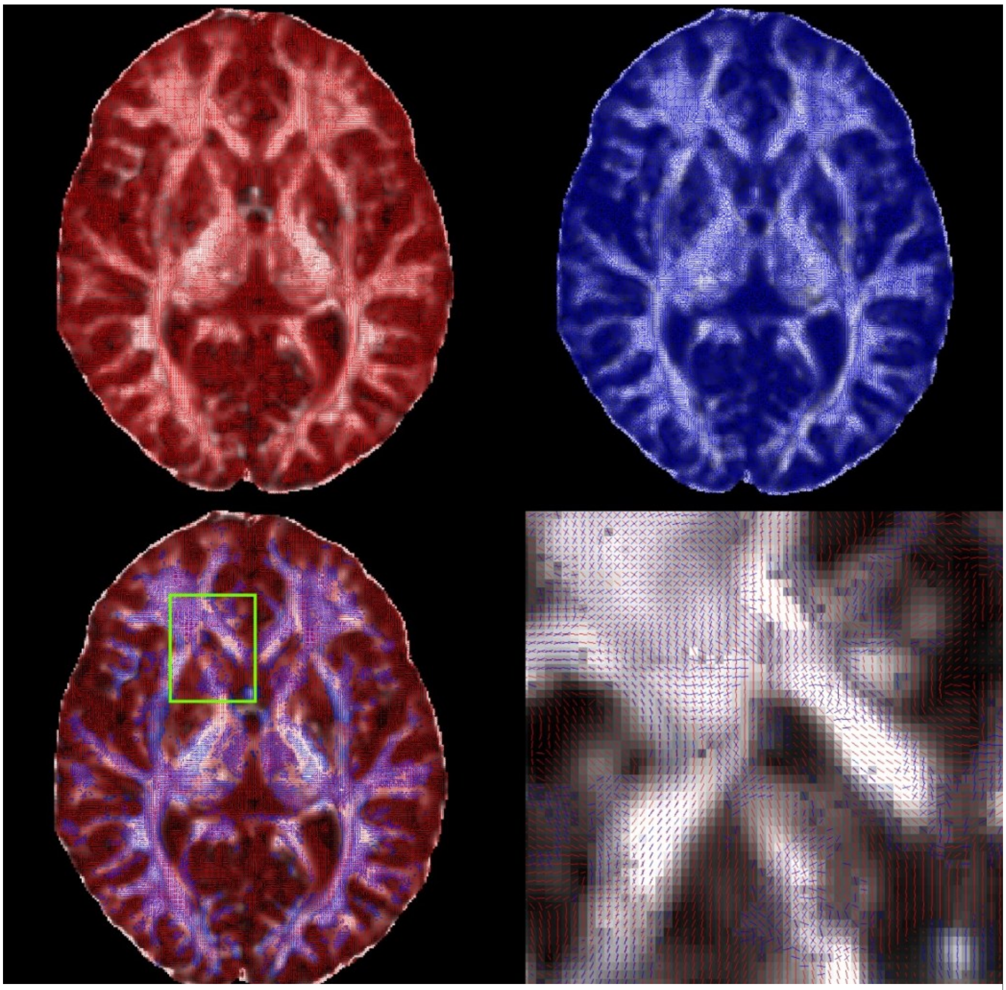


Figure 17 - Probabilistic Diffusion Modelling

3.2.3 Probabilistic tractography

After the multiple fibers modelling, classical tractography methods are unfeasible. Having no longer a single principal direction to follow means that a deterministic tractography algorithm is not able to decide which fiber a streamline should follow. To solve this issue, PROBTRACKX (Probabilistic tracking with crossing fibres) has been introduced[42]. To perform probabilistic tractography we need to define a seed voxel (or a group of seed voxels) and a target voxel (or a group of target voxels). This method repetitively samples starting at the seed voxels from the distributions of voxel-wise principal diffusion directions, each time computing a streamline through these local samples to generate a *probabilistic streamline* or a *sample* from the distribution on the location of the true streamline. By taking many such samples PROBTRACKX is able to build up the histogram of the posterior distribution on the streamline location or the *connectivity distribution* between the seed region and the target region. In Figure 18 the spatial histogram of the connectivity distribution of anterior thalamic radiation tract is shown. The colour scheme represents the confidence of the connection: yellow stands for a high connectivity region, red stands for a low connectivity region.

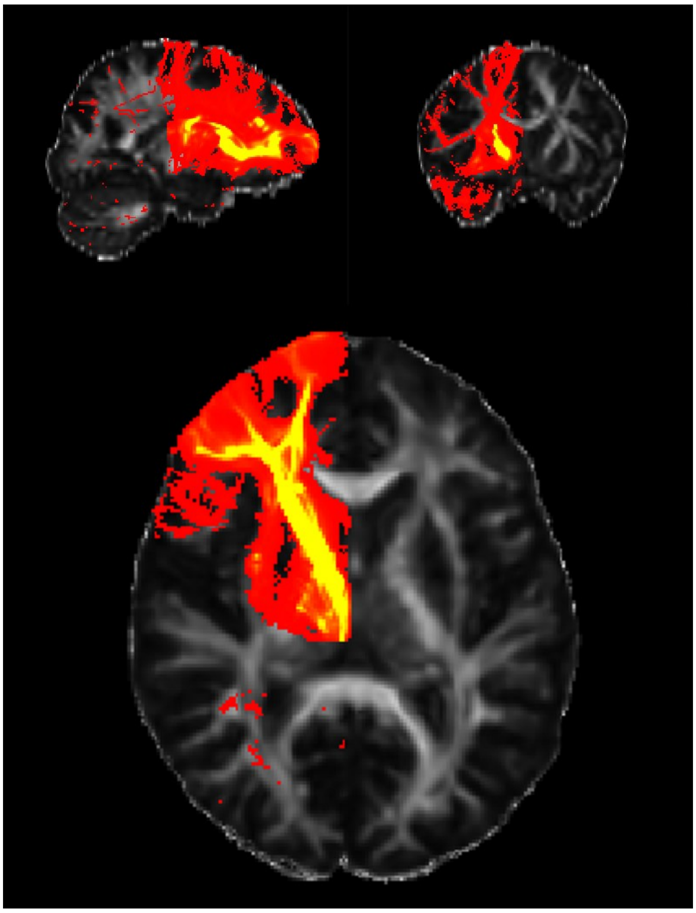


Figure 18 - Probabilistic Tractography Heathmap

3.2.4 Analysis Pipeline

In order to process data collected from the study and make them comparable, it has been necessary to develop and implement a rigorous analysis pipeline. The stages of this workflow are here presented and discussed (Figure 19).

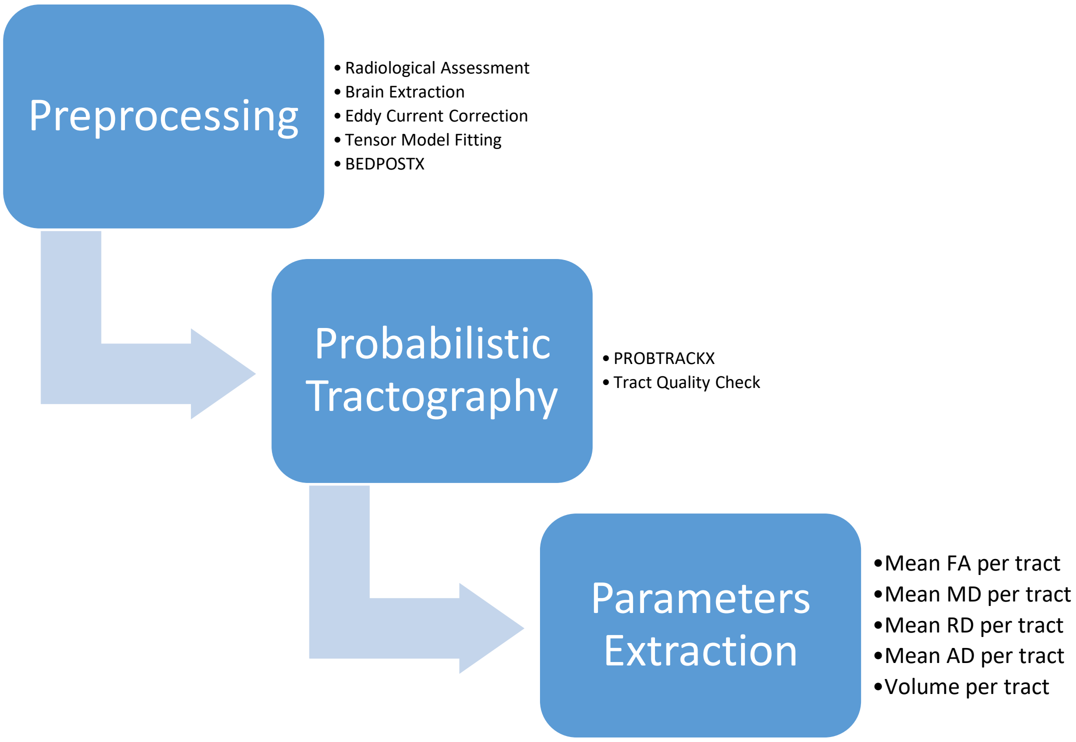


Figure 19 – DTI Analysis Pipeline

First, structural and DWI images are examined by an expert radiologist to exclude pathologies which could affect the outcome of the study (gliosis, white matter hyper intensities, ischemia). After the radiological response the raw DWI data begins the processing with BET (Brain Extraction Tool)[43], a tool which deletes from the image all non-brain tissues and creates a binary mask to identify the brain. The images so obtained are then processed with the eddy current correction tool[44]. Eddy currents are characteristics parasite currents induced by the magnetic field, resulting in image artefacts like shading and blurring. Corrected images are then co-registered to a common atlas in a standard space. On the resulting image is fitted the tensorial model, to obtain the standard parameter maps, then BEDPOSTX is performed in order to produce data for tractography[41].

The second stage of the pipeline is the tractographic analysis. In this step we create a per subject segmentation of the white matter based on connectivity features[45]. The main tract of the white matter has been characterized by a team of expert neuroanatomists[46]. In order to perform the segmentation, each tract has been described with a series of binary masks in standard space: the *seed mask*, the region from which the tract is originated; the *target mask*, the region where the tract is headed to; the *stop mask*, a region which describes a terminator for streamlines; the *exclusion mask*, if streamlines crosses this region are eliminated from the model. PROBTRACK is performed tract by tract to obtain a connectivity distribution for every tract specified in the protocol[42]. After that the tracts are thresholded to optimize the repeatability of the parameters' extraction. A different threshold is used for visualization. Tracts are identified by their functionality: in Figure 20A the right hemisphere (RH) fraction of the callosal tracts is shown, in Figure 20B the RH of the limbic system tracts is shown, in Figure 20C the RH of the associative tracts is shown, in Figure 20D the RH of the projection tracts is shown. On a per patient basis a visual inspection and tract quality check is performed, to exclude from the study potential outliers (interrupted tracts, errors due to misaligning of masks and anatomic structures). Once we have obtained the white matter segmentation for the patient, we proceed to extract the diffusion parameters associated with each tract. After a co-registration of segmentation maps and parameters maps, segmented areas are used as ROI in which the diffusion parameters are averaged and extracted. For each tract, volume is also calculated in order to exclude tract atrophy. After the clinic study the first data analysis approach has been the univariate statistical comparison, to explore where the hypertension could have caused damages and how was this damage characterized.

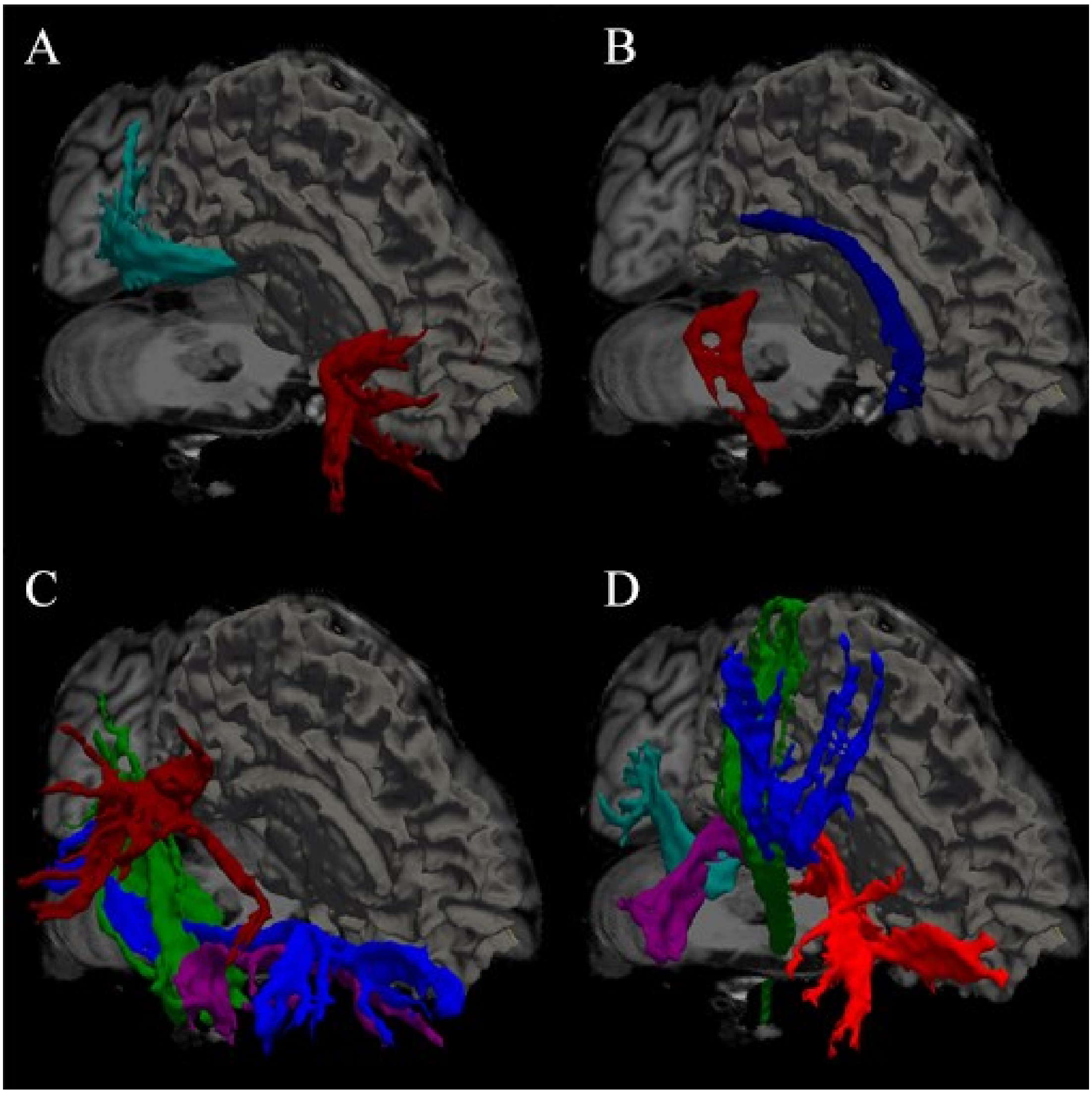


Figure 20 - White matter tracts reconstruction

3.3 Functional Connectivity: resting state functional MRI applied to Hypertensive Patients

Our second aim in this study was to characterize not only the damage that hypertension exerted on the brain structure, but which was the consequent functional damage. While mainly limited to psychiatric or neurodegenerative disorders, resting state functional MRI is a potent tool to assess the functional organization of the brain. By acquiring the blood-oxygen-level dependent signal (BOLD signal) we can image the relative concentration of oxygenated and deoxygenated blood in each voxel, and sampling this signal repeatedly let us reconstruct the regional neuronal activations[47]. (see Appendix A – MRI Sequences for more insights)

3.3.1 Functional Network Analysis

It has been extensively demonstrated that in normal subjects the brain in resting state shows a consistent pattern of synchronous alterations. Since the first works using this technique in small populations of healthy controls, it was very evident the activation during rest time of an ensemble of regions which were conversely negatively associated and non active during every kind of task previously administered: the Default Mode Network (DMN)[48-50].

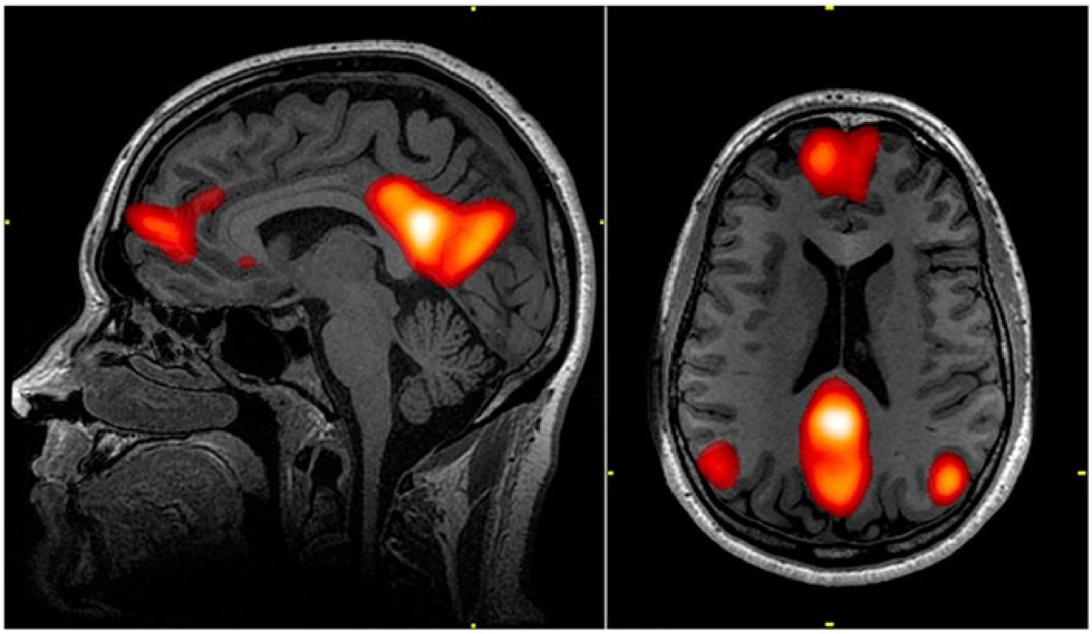


Figure 21 - Default Mode Network activation regions

Subsequent efforts were aimed at resolving all the synchronous connections through different regions of the brains, and these efforts resulted in precise mapping of each brain region associated to a group of synchronous activations dedicated to generic type of brain activity. One of the pivotal works in rs-fMRI research mapped 7 coarse networks of activation which could be eventually divided in 17 fine networks, each one consistent between one thousand subjects included in the study. To do this, Yeo at al. applied a clustering algorithm to group the voxels taking into account both their functional time course and their distance and regional profile[51].

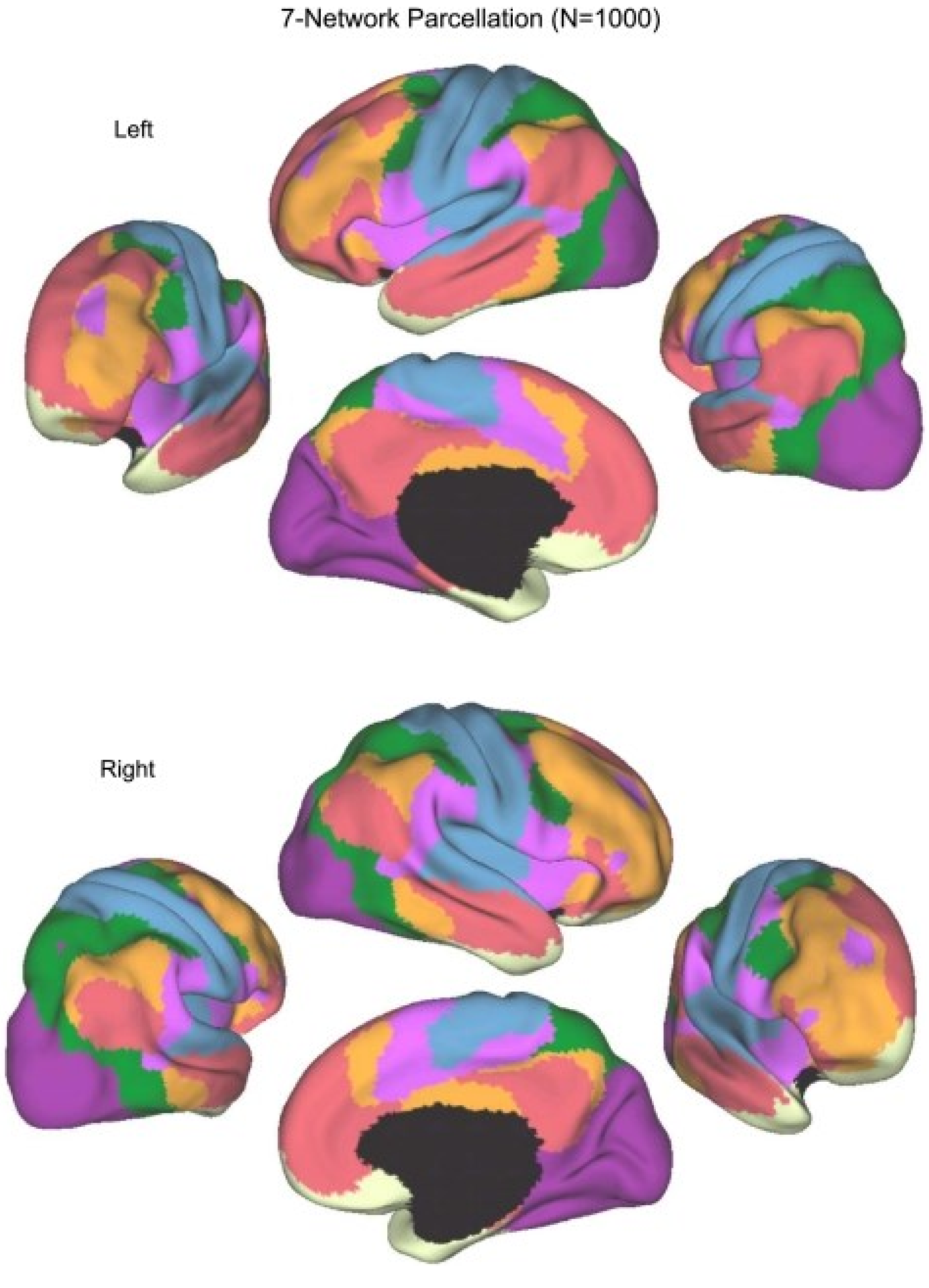


Figure 22 - Yeo et al. coarse (7-network) cortex parcellation

Other approaches were able to extrapolate functional networks from the raw data: Independent Component Analysis (ICA) is one of the main tools exploited in functional MRI analysis[52]. ICA is used to find a set of statistically independent maps (which are the maps depicting the regions of the functional networks) with time courses associated between them (which is the measure of their functional connectivity). This technique let us extrapolate the spatial and temporal information of the brain networks from the data, without any a-priori knowledge, and without any a-priori constraint. Usually the results of ICA, while explaining the quasi-totality of the starting dataset, need to be cleaned to rule out the spurious signals from regions not involved in functional connectivity (such as CSF or, in a classical vision of the problem, white matter). Once obtained the regional maps and their associated time course, we can proceed with the analysis of functional connectivity between different

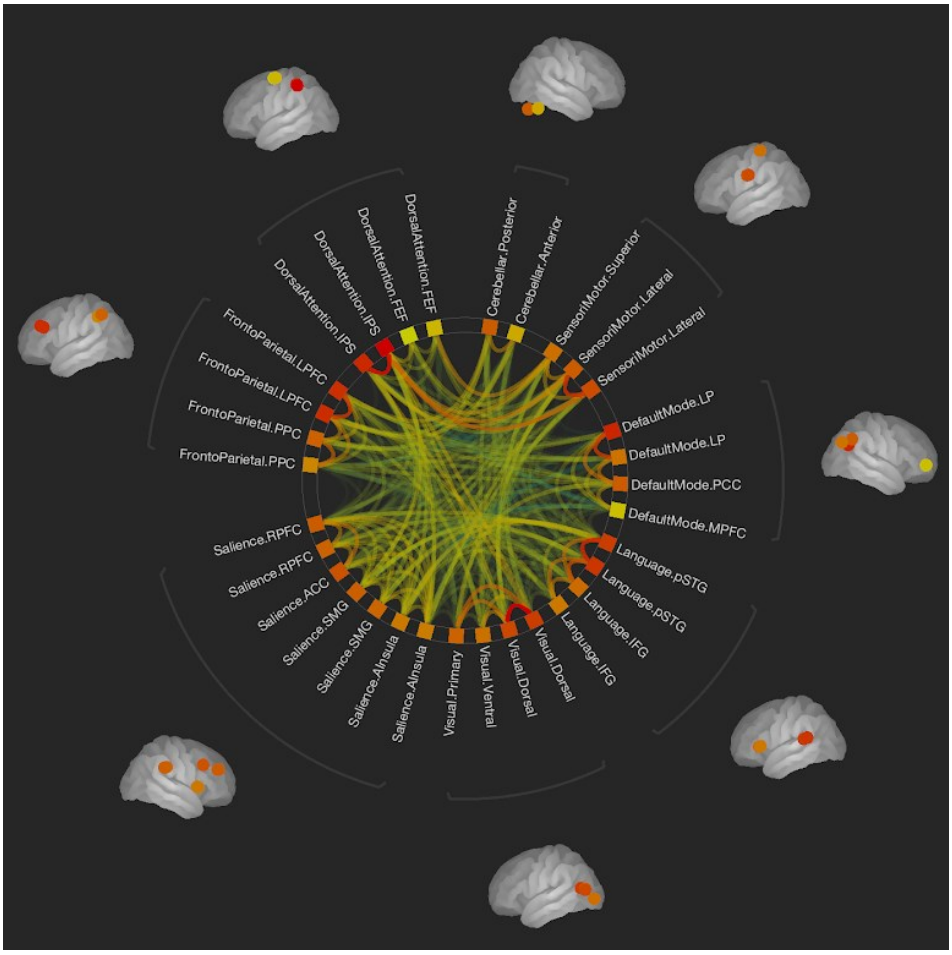


Figure 23 - Connectome Ring Display

patient populations. It can be considered intra network connectivity (the differences between FC in the synchronicity between different regions of the same network) or inter network connectivity (the differences between FC in the synchronicity of different regions between different networks). In Figure 23 it’s shown an example of a full connectome ring, in which every connection is represented in a round scheme, color coded by intensity of the correlation and grouped by the functional network of origin.

3.3.2 Graph Analysis

Once we evaluate the correlation between the timeseries of different regions, it is possible to build an undirected graph of connections. The resulting graph will be composed of a node for each region and edges between region whom timeseries correlation is above a certain threshold value[53].

On this built graph we can estimate topological measurements for each node and we can aggregate them across the network[54]:

- Degree: the degree of a node is the number of edges associated to the node, and estimates the network centrality and local connectivity of a node
- Average Path Length: Average length of the paths connecting a node with each other node in the network, gives an estimate of the compactness
- Clustering Coefficient: Fraction of edges among all possible edges in the local neighbouring sub-graph for each node. It is an estimate of the interconnectedness in sub-graphs, often measures the small-world characteristics of a network.
- Global and Local Efficiency: Average of inverse-distances between each node and all other nodes in the graph. The global efficiency measures the centrality of a node in a network, while the local efficiency is the same metric applied to neighbouring sub-graphs.
- Betweenness Centrality: Measure of node centrality in a graph, measures the percentage of optimal paths between other nodes in which the node is included.

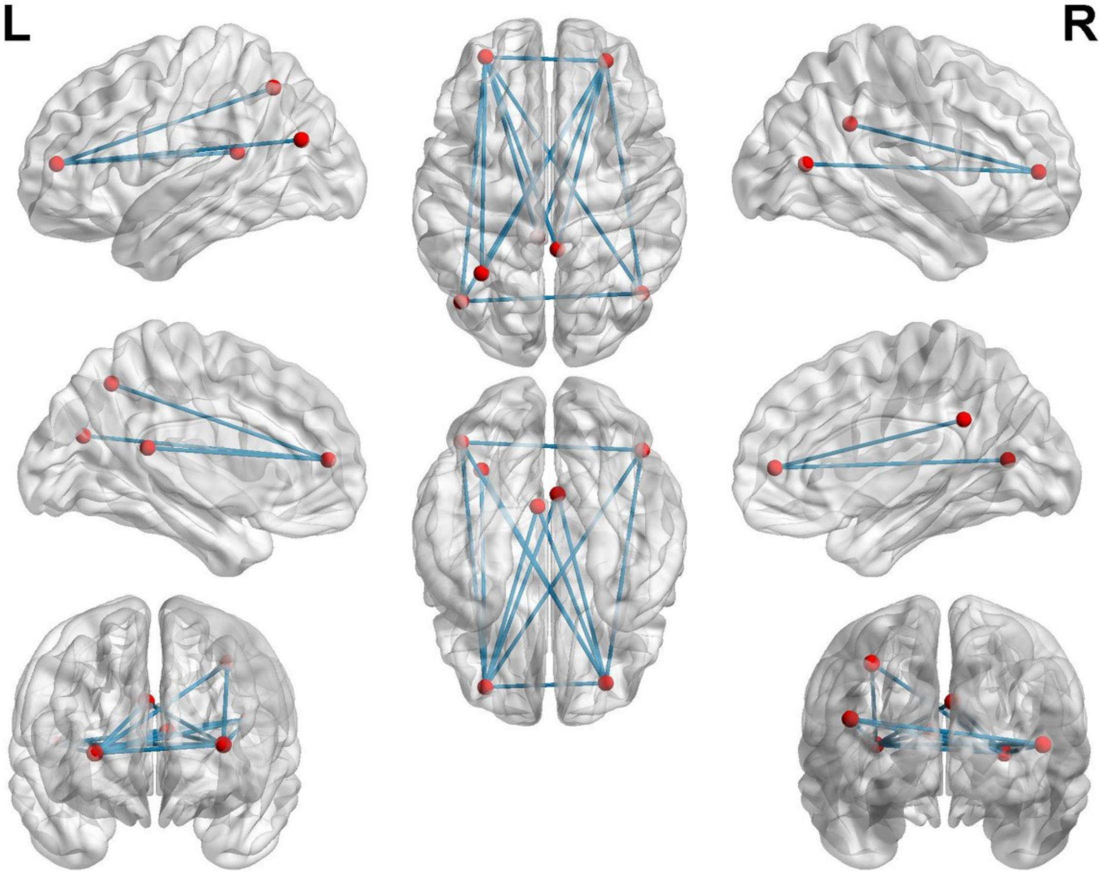


Figure 24 - Graph network built on rs-fMRI

3.3.3 Analysis Pipeline

In a similar manner for the DTI analysis pipeline, we needed to establish a pipeline to standardize the process to make comparisons between control and hypertensive population.

The preliminary step of the pipeline is to obtain a structural segmentation of the brain on T1-weighted images. After that, the functional image is denoised, low-pass filtered and co-registered to the structural one, to have a map of functional timeseries associated to structural region parcellation[55].

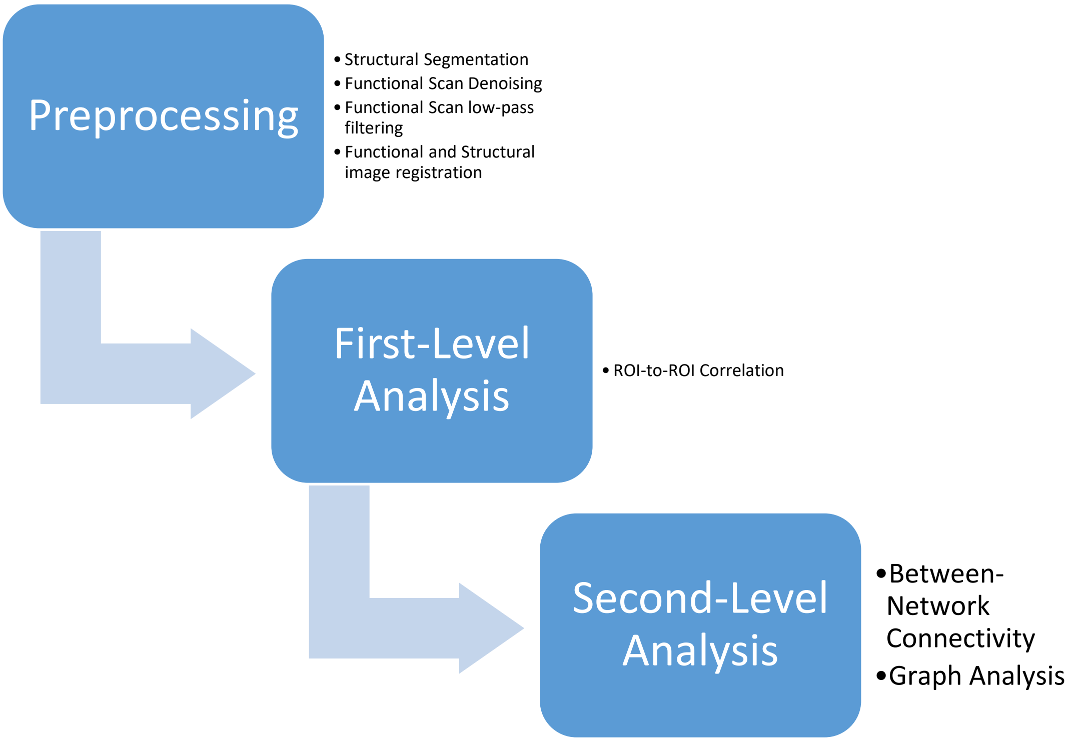


Figure 25 - rs-fMRI Analysis Pipeline

The subsequent stage is denominated First-Level Analysis, or subject level analysis. In this step all the operations are performed on a per-patient basis and are aimed at extracting functional connectivity measures from the functional scans. In our problem we are interested mainly to analyse whether the hypertension modifies the functional organization of the brain. Thus, we chose to implement a ROI-to-ROI, in which a set of previously validated networks are divided in ROIs and the timeseries in those ROIs are pairwise correlated to obtain an estimate of their functional connectivity. Once we characterized every patient, we carried on Second-Level Analysis, in which we perform group-level comparison between normotensive and hypertensive patients both on ROI-to-ROI functional connectivity and on topological metrics of the graph built by thresholding the connectivity matrices.

3.4 Results

Our multimodal characterization of the early brain damage revealed a characteristic pattern of damage, both in the microstructural integrity of the white matter fibers and in the functional organization between different regions.

3.4.1 Hypertension alters microstructural integrity of white matter

By assessing the diffusion parameters of FA, MD, AD and RD of the segmented tracts, we evidenced a pattern of significant alterations which pointed out a worse fascicle organization in hypertensive patients' white matter respect to the normotensive ones. This is supported by the concomitant variation of FA and MD in specific tracts, a characteristic usually associated to an ongoing pathological process.

Hypertensive patients show a reduced integrity of WM fiber bundles of right Anterior Thalamic Radiation (ATR) (Figure 26A), right Superior Longitudinal Fasciculus (SLF) (Figure 26B) and Forceps Minor (FMI) (Figure 26C).

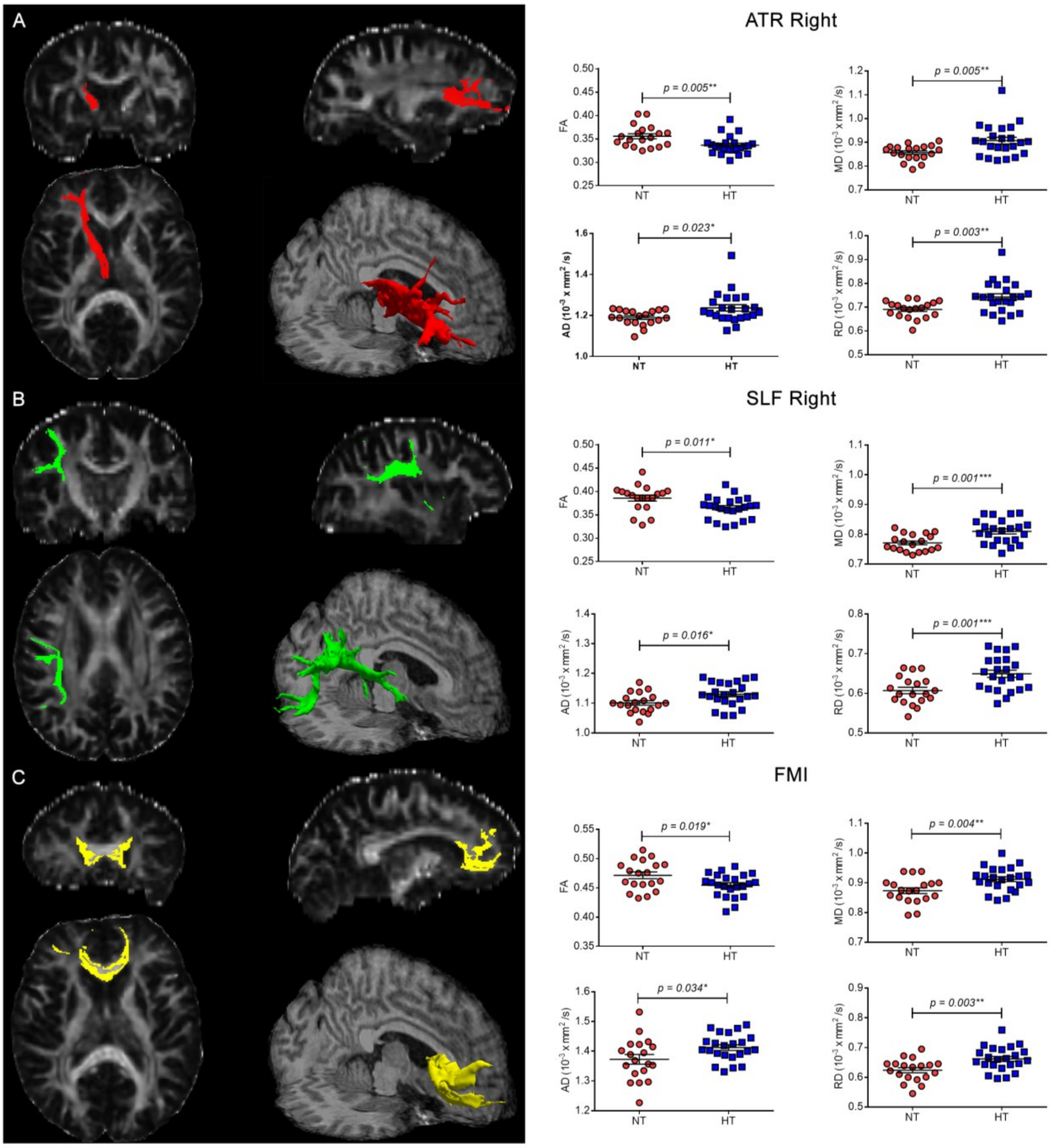


Figure 26 - WM altered tracts

While the damage on FA was evident only in the aforementioned tracts, the MD showed a broader pattern of alteration, suggesting an ongoing WM integrity loss which can be first measured by MD (Table 6). In addition, AD and RD DTI parameters were considered, to take into account the potential impact of different neural mechanisms on WM abnormalities. Conventionally, incremental variations in RD are associated with myelin breakdown, whereas in AD describe secondary processes of axon degeneration. Our data also report a significant alteration of RD and AD in the tracts of interest (Table 7).

	FA		**MD**	
Associative	**NT**	**HT**	**NT**	**HT**
SLF R	0.39 (0.03)	0.36 (0.02)*	0.77 (0.03)	0.81 (0.04)***
SLF L	0.39 (0.03)	0.39 (0.02)	0.76 (0.03)	0.80 (0.04)**
ILF R	0.39 (0.02)	0.38 (0.02)	0.86 (0.02)	0.87 (0.04)
ILF L	0.39 (0.02)	0.39 (0.02)	0.86 (0.02)	0.87 (0.03)
IFO R	0.40 (0.02)	0.39 (0.02)	0.85 (0.02)	0.88 (0.04)*
IFO L	0.40 (0.02)	0.39 (0.03)	0.86 (0.02)	0.88 (0.04)
UNC R	0.33 (0.02)	0.32 (0.02)	0.89 (0.04)	0.93 (0.05)**
UNC L	0.33 (0.03)	0.33 (0.02)	0.88 (0.04)	0.91 (0.06)*
Limbic				
CGC R	0.35 (0.03)	0.34 (0.02)	0.81 (0.03)	0.84 (0.04)*
CGC L	0.39 (0.04)	0.37 (0.03)	0.81 (0.03)	0.83 (0.05)
CGH R	0.23 (0.02)	0.22 (0.02)	0.99 (0.04)	1.03 (0.06)*
CGH L	0.23 (0.02)	0.21 (0.03)	0.99 (0.06)	1.02 (0.10)
Projection				
CST R	0.45 (0.02)	0.44 (0.03)	0.89 (0.05)	0.90 (0.06)
CST L	0.44 (0.03)	0.44 (0.03)	0.90 (0.06)	0.89 (0.06)
AR R	0.30 (0.02)	0.30 (0.02)	0.93 (0.07)	0.98 (0.07)
AR L	0.29 (0.02)	0.30 (0.02)	0.95 (0.06)	0.97 (0.06)
ATR R	0.36 (0.02)	0.34 (0.02)**	0.86 (0.03)	0.91 (0.07)**
ATR L	0.36 (0.02)	0.35 (0.02)	0.86 (0.04)	0.89 (0.05)
STR R	0.38 (0.02)	0.37 (0.02)	0.82 (0.05)	0.83 (0.06)
STR L	0.39 (0.02)	0.38 (0.02)	0.81 (0.04)	0.82 (0.05)
PTR R	0.37 (0.02)	0.36 (0.03)	0.88 (0.04)	0.88 (0.04)
PTR L	0.37 (0.02)	0.37 (0.02)	0.89 (0.04)	0.90 (0.04)
Callosal				
FMI	0.47 (0.02)	0.45 (0.02)*	0.87 (0.04)	0.91 (0.04)**
FMA	0.45 (0.03)	0.45 (0.03)	1.04 (0.08)	0.99 (0.07)*

Table 6 - Average FA and MD of segmented tracts

	AD		**RD**	
Associative	**NT**	**HT**	**NT**	**HT**
SLF R	1.10 (0.03)	1.13 (0.04)*	0.61 (0.04)	0.65 (0.04)**
SLF L	1.12 (0.03)	1.14 (0.04)*	0.61 (0.03)	0.63 (0.04)*
ILF R	1.24 (0.03)	1.25 (0.03)	0.66 (0.03)	0.68 (0.04)
ILF L	1.25 (0.03)	1.26 (0.03)	0.66 (0.02)	0.68 (0.04)
IFO R	1.24 (0.05)	1.27 (0.04)	0.66 (0.02)	0.69 (0.04)*
IFO L	1.25 (0.04)	1.28 (0.05)	0.67 (0.02)	0.69 (0.05)
UNC R	1.22 (0.05)	1.25 (0.05)*	0.73 (0.04)	0.77 (0.05)**
UNC L	1.20 (0.05)	1.24 (0.06)*	0.73 (0.05)	0.75 (0.06)
Limbic				
CGC R	1.12 (0.05)	1.15 (0.05)	0.65 (0.03)	0.68 (0.04)**
CGC L	1.17 (0.05)	1.18 (0.05)	0.63 (0.04)	0.66 (0.05)*
CGH R	1.23 (0.05)	1.26 (0.07)	0.88 (0.04)	0.92 (0.06)*
CGH L	1.22 (0.07)	1.24 (0.10)	0.88 (0.06)	0.92 (0.10)
Projection				
CST R	1.33 (0.06)	1.33 (0.06)	0.67 (0.06)	0.68 (0.06)
CST L	1.33 (0.06)	1.32 (0.05)	0.68 (0.07)	0.68 (0.06)
AR R	1.22 (0.08)	1.27 (0.07)*	0.79 (0.07)	0.83 (0.07)
AR L	1.22 (0.06)	1.27 (0.07)	0.81 (0.06)	0.83 (0.06)
ATR R	1.19 (0.04)	1.24 (0.08)*	0.69 (0.04)	0.74 (0.06)**
ATR L	1.19 (0.05)	1.23 (0.05)*	0.69 (0.04)	0.72 (0.05)*
STR R	1.16 (0.04)	1.17 (0.07)	0.64 (0.05)	0.66 (0.06)
STR L	1.15 (0.03)	1.16 (0.05)	0.64 (0.04)	0.65 (0.05)
PTR R	1.24 (0.05)	1.24 (0.05)	0.70 (0.04)	0.70 (0.04)
PTR L	1.26 (0.06)	1.26 (0.04)	0.71 (0.04)	0.71 (0.05)
Callosal				
FMI	1.37 (0.07)	1.41 (0.05)*	0.62 (0.04)	0.66 (0.04)**
FMA	1.57 (0.09)	1.52 (0.08)	0.78 (0.08)	0.73 (0.07)

Table 7 - Average AD and RD of segmented tracts

3.4.2 WM microstructural alterations scale with cognitive impairment and target organ damage

We generated correlation models among microstructural WM alterations, cardiac remodeling, hypertensive condition and cognitive profile. We found a significant positive correlation between MoCA scores and FA of the projection and association fibers (ATR and SLF) (Figure 27C and D). FA values of the same projection and association fibers negatively correlated with the estimated duration of hypertension (Figure 27E) as MoCA scores did (Figure 27F). It is interesting to notice that the sample of hypertensive patients (n=18), excluded because of already manifest neurological damage, had a significantly longer estimated duration of hypertension (mean ± SD = 9.61 ± 5.57) when compared to the group of included patients (mean ± SD = 6.13 ± 3.73) (*$p < 0.05$), thus suggesting that with disease advancement, brain damage progressively evolve toward an increasingly manifest injury. Further supporting this hypothesis, both the included and excluded patients had comparable good levels of blood pressure control.

Hence, in order to test whether the observed correlations between DTI parameters and cognitive scores was modulated by hypertensive condition, we also performed a partial correlation analyses controlling for SBP and years of hypertensive conditions. There were no significant influences of SBP levels and overall duration of hypertension in the correlation observed neither between MoCA and FA r-ATR (SBP: $p = .132$; years of hypertension: $p = .213$) nor between MoCA and FA r-SLF (SBP: $p = .069$; years of hypertension: $p = .081$).

Since altered FMI may be involved in impaired processing speed tasks41, we tested the correlation between Stroop interference time, Stroop test errors and FMI-FA, finding a significant relationship suggestive of an impact of hypertension in inhibiting interfering stimuli, represented by the time performance (Figure 27G and H). Even the correlation observed between FA-FMI and Stroop Interference Time was controlled for the interaction with both SBP, duration of hypertension and age, given that this

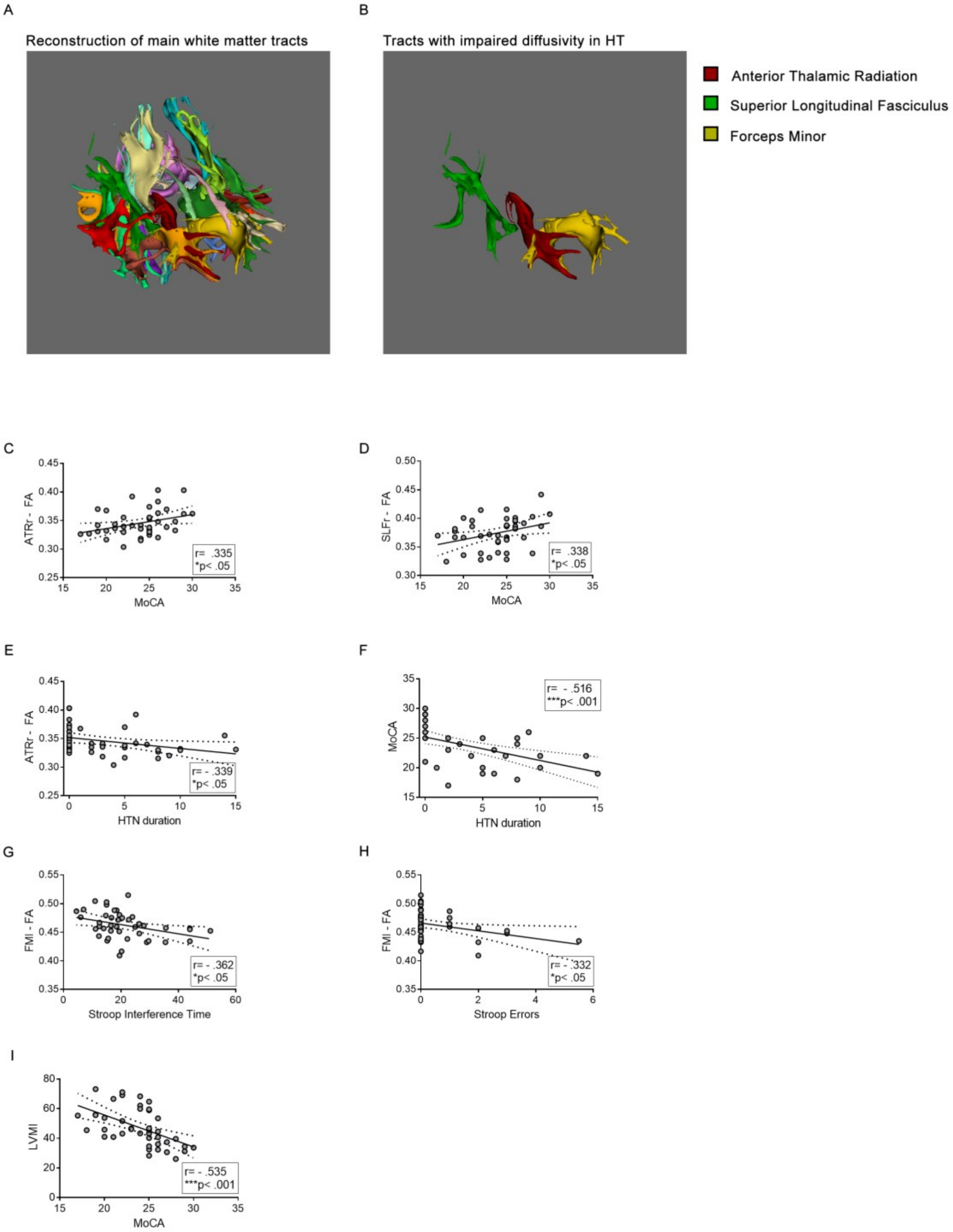


Figure 27 - Characteristic pattern of altered tracts and DTI-organ damage correlations

latter parameter emerged as influencing the deterioration of FA-FMI observed in HT. Interestingly, while there was a statistically significant negative partial correlation between Stroop Interference Time and FA-FMI whilst controlling for SBP (p = .040) no effect was observed for overall duration of hypertension (p = .074) or when controlling for age (p = .431). When controlling the correlation between Stroop test errors and FMI-FA for the same parameters it emerged a significant negative correlation when controlling for overall duration of hypertension (p = .047) while no effect were observed controlling for SBP (p = .057) or when controlling for age (p = .219).

In the end, the significant negative correlation observed between indexed LVMI and MoCA (Figure 27I) revealed parallel progression of early cognitive alterations and initial peripheral organ damage.

3.4.3 ROI-to-ROI analyses of Functional Connectivity show altered aberrant connections between task-positive networks

The ROI-to-ROI showed a pattern of altered connections in hypertensive patients respect to the normotensive controls between the Dorsal Attention Network, the Visual Network, the Sensorimotor Network and the FrontoParietal Network. Specifically, in the hypertensive sample the Intraparietal Sulcus (IPS) region of the right hemisphere shows an increased functional connectivity with the Superior (p = 0.0064) and the Lateral (R: p = 0.0379; L: p = 0.0427) regions of the SensoriMotor Network and increased functional connectivity with the Occipital region of the Visual Network (p = 0.0427). Conversely, the right hemisphere IPS presents a decreased functional

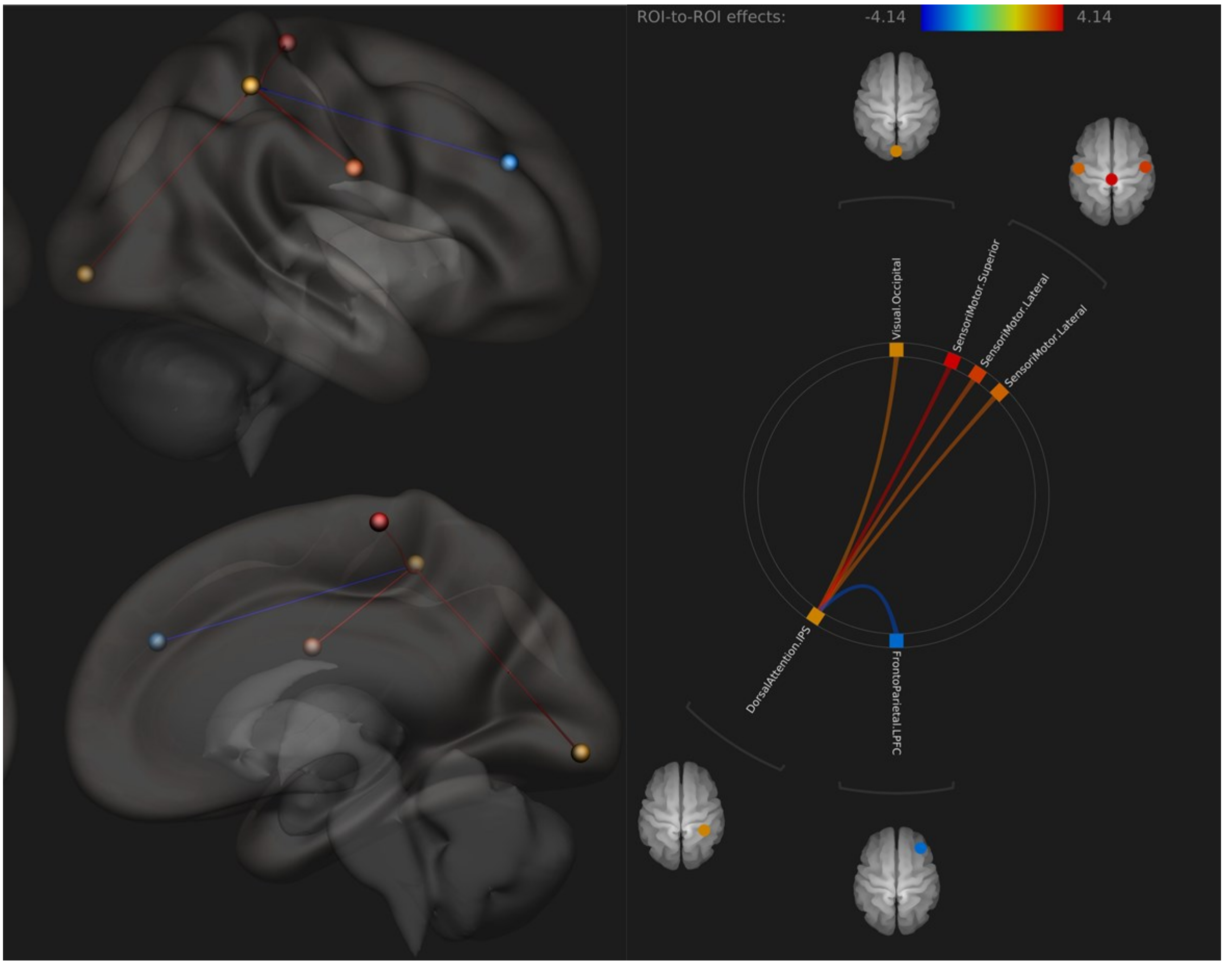


Figure 28 - Connectome ring showing altered subnetwork and its 3D reconstruction

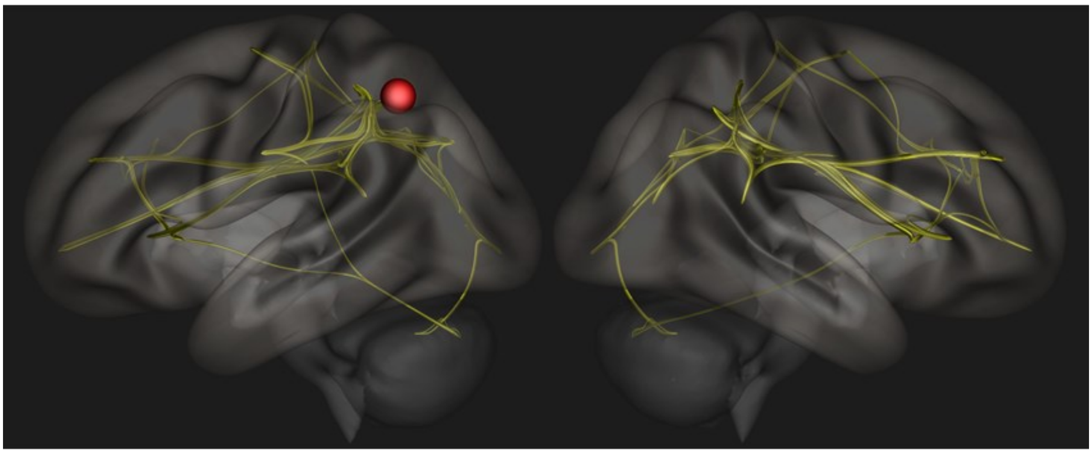
connectivity with the same hemisphere Lateral PreFrontal Cortex (LPFC) of the FrontoParietal network (p = 0.0379) (Figure 28).

3.4.4 Graph Theory Analyses of brain network connectivity

By thresholding the connectivity values between the different regions of interest we can obtain an undirected graph representing the global network of connections in the brains. Once obtained the network, graph theory analysis can highlight differences in the global organization of the network and the differences in the nodes of the network. While no metric shows significant differences at global network level, the Intraparietal Sulcus region of the FrontoParietal network in the left hemisphere shows a decreased node Degree (p = 0.0485) and an increased Average Path Length (p = 0.0428).

Moreover, the left hemisphere of the Salience network shows a similar decrease in Degree (p = 0.0485) (Figure 29).

AVERAGE PATH LENGTH



DEGREE

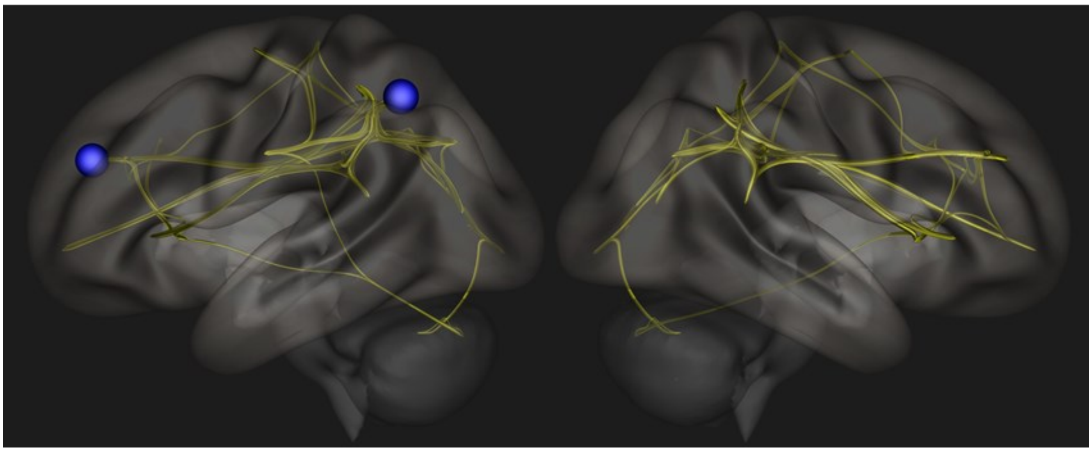


Figure 29 - Nodes altered in corresponding graph metrics

3.4.5 Diffusion parameters and cognitive performances correlates with the functional connectivity

Analyzing the effect of different covariates on functional connectivity, we tested whether the Fractional Anisotropy of altered connection identified in hypertensive patients and their scores in cognitive tests could be correlated to the functional connectivity. We tested Anterior Thalamic Radiation (ATR), Superior Longitudinal Fasciculus (SLF) and Forcep Minor (FMI) regarding the tracts, MoCA Score and

Stroop Test Execution Time for the cognitive performances scores. No significant correlation could be identified between FC and ATR FA nor between FC and FMI FA, while SLF FA was negatively correlated to the connection between the Precuneus Cortex of the Default Mode Network and the Posterior Parietal Cortex of the FrontoParietal Network ($p = 0.039$). Global cognitive performances, expressed as MoCA score, is correlated to various connectivity scores between different regions, mirroring the capability of the test to examine different cognitive domains regulated by different brain regions. Specifically, a significant correlation could be identified with the connectivity scores between the right IPS of the Dorsal Attention Network and the SupraMarginal Gyrus of the Salience Network ($p = 0.0463$) and the Superior region of Sensorimotor Network ($p = 0.0240$). Moreover, a significant correlation was found between MoCA Score and the connectivity of the right Frontal Eye Fields of the Dorsal Attention Network and the posterior Superior Temporal Gyrus (pSTG) of the Language network ($p = 0.0320$) and between pSTG and Accumbens of Salience Network ($p = 0.0320$).

3.5 Dataset Construction

All data has been collected in csv file format. A single database containing clinical, cognitive and tractographic data has been used for univariate statistical analysis. From the global data three separate datasets for machine learning analysis have been extracted:

Functional data: the dataset is composed of 407 attributes.

Advanced Neuroimaging data: contains all data from the aforementioned dataset plus the addition of tractographic data. Is composed of 542 attributes.

Advanced Neuroimaging + Cognitive data: contains all data from the aforementioned dataset with the addition of data from cognitive tests. Is composed of 549 attributes.

CHAPTER 4. A MACHINE LEARNING APPROACH TO CLASSIFY HYPERTENSION FROM ADVANCED NEUROIMAGING DATA

4.1 The curse of dimensionality

Richard Bellman introduced the term *curse of dimensionality* in 1961 to describe the problem of an exponential growth in volume when the dimensions of the input space increase. For example, the volume of a hypercube with 100 units (sampled intervals) in each of its *p* dimensions has a volume of $V = 100^p$. Regarding these problems Bellman said: “This does not mean that we cannot attack them. It merely means that we must employ more sophisticated methods”.

With the technological advancement in the biomedical field the capability to extract different data from a single observation increased, often presenting datasets with more variables than observation. The effort of the statisticians and machine learning experts have been very focused on this kind of problem, defined in its theoretical aspects only during the last decade. “*Large p, small n*” problems[56], with p number of features composing the feature vector and n number of observations, can be easily found in genetic analysis or functional MRI data[57]. One of the solutions to optimize the classification of such datasets has been the feature reduction with various techniques of feature selection. Following the example of Kuncheva, we wanted to investigate whether our dataset required a dimension reduction and what technique could achieve the best result. The experiment consisted of the comparison between various classification techniques with and without a feature selection strategy.

4.2 Weka Tools

Waikato Environment for Knowledge Analysis (WEKA), is a very popular open source software written in Java and developed at the University of Waikato, New Zealand, in 1997. It is available under the GNU General Public License. Contains a graphical user interface (GUI) very useful for interacting with data files and represent the results in an intuitive form like curves or graphics[58].

Some of WEKAs main features are the following:

Data preprocessing - WEKA supports a couple of popular text file formats such as CSV, JSON and Matlab ASCII files to import data along with their own file format ARFF. They also have support to import data from databases through JDBC. Beside importing data, they have a wide collection of supervised as well as unsupervised filters to apply on your data to facilitate further analysis.

Data classification - A huge collection of algorithms has been implemented to perform classification on data sets. These include Bayesian algorithms, mathematical functions such as support vector machines, lazy classifiers implementing nearest-neighbour calculations, meta based algorithms as well as rule and tree-based classifiers.

Data clustering - A couple of algorithms for clustering exist such as variations of the k-mean method as well as density and hierarchical based clustering algorithms.

Attribute selection - Methods to evaluate which attribute contribute the most when predicting an outcome.

Data visualization - Depending on the methods used to analyse the data, this view can to plot data against suitable variables as well as give tools to analyse specific points further.

4.3 Classifiers

In this section we present the most popular classifiers provided by WEKA tools which have been used in our experiments.

4.3.1 J48 Decision Tree

J48 is an implementation of the well-known C4.5 algorithm for producing either pruned or unpruned C4.5 tree[59]. The C4.5 algorithm was built based on the concept of information obtaining or entropy reduction to select the most efficient split. In general, it assumes that individual attributes of the data can be used to make a decision by splitting the original data into minor subsets. The main reason J48 decision tree was chosen to serve as a model for classification is that it produces simpler rules and remove insignificant parameters before it begins a process of tree induction. Usually, J48 decision trees happen to have a relatively higher accuracy than other classification algorithms. In addition, J48 also provides extremely fast and pretty powerful way of fast and powerful way to show structures for a data.

4.3.2 Naïve Bayes

Naïve Bayes classifier works under the assumption that the classes are all statistically independent[60]. The Naïve Bayes algorithm makes use of Bayes' Theorem, which is a formula that determines a probability by estimating the frequency of values and mixture of values in the previously collected data. It determines the probability of an event happening provided that the probability of another event that has already happened. The Bayes' theorem is stated as follows:

$$P\ (H/X) = P(X/H)\ P(H)\ /\ P(X)$$

The Naive Bayes algorithm provides a way to mix the prior probability and conditional probabilities within a single formula that can be used to determine the probability of each of the classifications in turn. After that, the class with the highest value will be chosen as the class of the new instance.

4.3.3 Logistic Regression

Logistic regression models the probabilities for classification problems with two possible outcomes. It's an extension of the linear regression model for classification problems[61].

The logistic function is defined as:

$$logistic(\eta) = \frac{1}{1 + \exp(-\eta)}$$

Logistic regression is used in various applications, mainly in the medical field and social sciences, as a simple model to perform binary prediction with one or more features (with the multivariate model).

4.3.4 SVM

Support vector machines, also called SVM, is a popular type of learning machine that use supervised learning models to analyse and classify data. One of the main areas of usage is to construct an optimal model that can distinguish new data points into one of two different classes. First presented by Vapnik and Lerner in 1963, the main concept is to construct a hyperplane as the separator of the two classes[62]. The clever thing about hyperplanes is that it can be with ease be applied in higher dimensions as well which makes them ideal to general solutions.

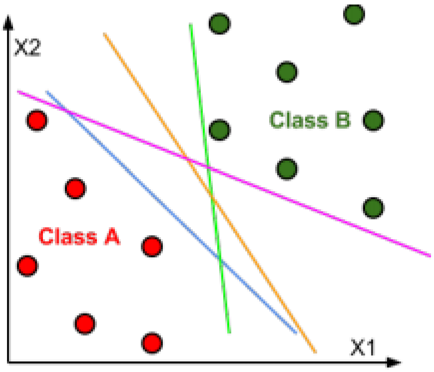


Figure 30 - Linearly separable data

When dealing with a linearly separable data set, there can be up to infinite ways to construct a hyperplane to correctly divide a data set into two classes as seen in Figure 30. However, this is where SVM excels since the method presented by Vapnik will present an optimal hyperplane such as the one seen below in Figure 31.

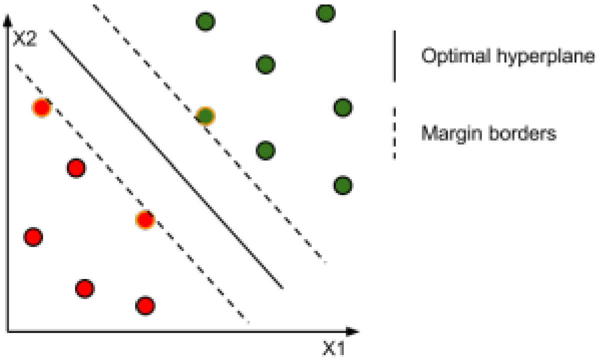


Figure 31 - Optimal hyperplane placement

Not all data sets are linearly separable as seen in, in fact that is usually the case when dealing with real data. A clever trick is that the above algorithm to maximize the margin still works in higher dimensions, so if you're faced with non-linear separable points in the x space, you can do a nonlinear transformation into a much higher dimensional space and solve the problem there with the linear SVM method.

4.4 Ensemble

A common optimization of the classifier performances is given by using ensembles techniques to exploit a combination of various *weak learners* (learners which have an accuracy at least of 50%+1) to obtain a strong learner.

4.4.1 AdaBoost

AdaBoost stands for *"Adaptive Boosting",* is one of the most used algorithms for the generation of a classifier ensemble[63]. AdaBoost builds a list of classifiers, assigning in an iterative procedure a weight to each of them, based on the capability to correctly classify samples which hadn't been correctly classified by other learners. Every classifier will contribute to the final decision proportionally to their weighting and the output will be decided on majority of voting.

4.4.2 Bagging

Bagging stand for *"Bootstrap Aggregating",* is an algorithm which improves the stability of the selected weak learner ensemble[64]. From a given training set *D,* this technique consists in generating *m* different training sets D_i by sampling from *D* uniformly and with replacement. A weak learner is trained on one of the *m* different training sets. After the different models are built, the different learners are combined in the classification stage by voting.

4.4.3 Random Forest

Random forest is an ensemble method in which multiple decision trees are grown independently and can be grown in parallel. Each tree is built using all the training samples sampled with replacement[65]. The forest is trained to optimize the parameters at each node of every tree. While training a tree, each node of the tree has access to only a randomly chosen subset of the entire set of features. Training consists of selecting at each internal node the feature in the randomly chosen feature subset that best separates the training samples. During testing, all decision trees in the forest classify the sample, and the predicted class label is simply most frequent output of the individual decision trees.

4.4.4 Random Subspace

Random Subspace is a technique in which we can construct an ensemble on *m* classifiers with *m* different training set obtained by a random sampling without replacement of the original training set *D*[66]. After the different models are built, the different learners are combined in the classification stage by voting. It is a technique similar to bagging, with the main difference that in the latter the sampling is with replacement.

4.5 Feature Selection

The strategies to reduce the number of examined features are fundamental to achieve a better classification result, especially in a "*Large p, small n*" situation. These strategies are used mainly to enhance the generalization capabilities of the classifiers, reducing the overfitting of the model. Hereby are described the used selection strategies.

4.5.1 Correlation-based Feature Subset

CFS evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Correlation coefficients is used to estimate correlation between subset of attributes and class, as well as inter-correlations between the features[67]. Relevance of a group of features grows with the correlation between features and classes and decreases with growing inter-correlation. CFS is used to determine the best feature subset and is usually combined with search strategies such as forward selection, backward elimination, bi-directional search, best-first search.

4.5.2 Information Gain and Gain Ratio

The information gain filter measures the attribute's information gain with respect to the class. Entropy was defined as follows:

$$H(D) = -\sum_{i=1}^{m} p_i log_2(p_i)$$

Where D is our data sample, pi is the proportion of D in respect to class C_i, m is the number of outcomes. *H(D)* is bound by 0 (perfectly classified) and 1 (totally random). We can define the expected information required to classify a sample of the dataset based on the partitioning of a feature A.

$$H_A(D) = \sum_{j=1}^{v} (|D_j|/|D|) \times H(D_j)$$

where D_j is the subset of D containing distinct value of A, and v is the number of distinct values in A[68]. The *information gain* is the difference between the prior entropy of classes and posterior entropy.

$$Gain = H(D) - H_A(D)$$

Gain Ratio is the Gain normalized by the *intrinsic information* of a feature.

4.5.3 Principal Component Analysis

PCA is one of the most commonly used dimensionality reduction methods. PCA finds a linear combination of the multivariate data that captures a maximum amount of variance[69]. However, the projections that PCA seeks are not necessarily related to the class labels; hence they may not be optimal for classification problems.

4.6 Results

Here is the global comparison among the top performing classifiers on the three different datasets (Functional, Advanced Neuroimaging, Advanced Neuroimaging and Cognitive data) and then the global ranking all the classifiers applied. The evaluated parameters are classification accuracy, AUC and F-measure. Each result is the average over 10 different runs of the same classifier, each run configured in 10-fold cross validation. (Total run: 18000)

4.6.1 Global Comparison

Functional data Classifier	**Accuracy**	**AUC**	**F Measure**
Random Forest - Infogain	**85.83% (19.37)**	**0.95 (0.13)**	**0.83 (0.26)**
Random Forest - GainRatio	85.18% (18.45)	0.93 (0.16)	0.81 (0.27)
Random Forest - CFS	85.08% (19.11)	0.92 (0.18)	0.81 (0.28)
Naïve Bayes - CFS	82.67% (18.90)	0.92 (.019)	0.80 (0.25)
J48 – Bagging - CFS	81.58% (20.25)	0.89 (0.22)	0.80 (0.25)

Table 8 - Top 5 classifiers on functional data

Neuroimaging data Classifier	Accuracy	AUC	F Measure
Naïve Bayes – CFS	**91.25% (15.95)**	**0.95 (0.15)**	**0.91 (0.19)**
Random Forest – CFS	89.50% (15.87)	0.94 (0.15)	0.89 (0.19)
Naïve Bayes – Infogain	86.75% (18.69)	0.93 (0.16)	0.86 (0.23)
Random Forest – Infogain	85.50% (18.41)	0.92 (0.16)	0.83 (0.25)
SVM – Random SubSpace – Infogain	84% (17.99)	0.90 (0.21)	0.83 (0.24)

Table 9 - Top 5 classsifiers on advanced neuroimaging data

All data Classifier	Accuracy	AUC	F Measure
Random Forest – CFS	**89.42% (15.97)**	**0.96 (0.12)**	**0.87 (0.23)**
Naïve Bayes – CFS	89.17% (17.14)	0.95 (0.13)	0.87 (0.23)
SVM – Random SubSpace – CFS	85.58% (18.95)	0.93 (0.17)	0.85 (0.23)
Naïve Bayes – Infogain	85.17% (17.67)	0.93 (0.17)	0.84 (0.23)
SVM – Bagging – CFS	83.75% (19.48)	0.89 (0.20)	0.81 (0.27)

Table 10 - Top 5 classifiers on all data

4.6.2 Global Ranking

Accuracy Classifier	**Better**	**Worse**	**Delta**
Naïve Bayes	**15**	**0**	**15**
Random Forest	9	0	9
SVM – Random SubSpace	3	0	3
SVM	3	0	3
SVM – Bagging	2	0	2
J48 – Bagging	0	0	0
J48 – Adaboost	1	1	0
MLP	0	2	-2
SVM – Adaboost	0	3	-3
J48 – Random Subspace	0	4	-4
J48	0	9	-9
Logistic Regression	0	14	-14

Table 11 - Accuracy Ranking

F Measure Classifier	**Better**	**Worse**	**Delta**
Naïve Bayes	**12**	**0**	**12**
Random Forest	6	0	6
SVM – Random SubSpace	4	0	4
SVM	1	0	1
SVM – Bagging	1	0	1
J48 – Bagging	0	0	0
J48 – Adaboost	0	0	0
MLP	0	2	-2
SVM – Adaboost	0	-2	-2
J48 – Random Subspace	0	3	-3
J48	0	5	-5
Logistic Regression	0	12	-12

Table 12 - F Measure Ranking

AUC Classifier	**Better**	**Worse**	**Delta**
Random Forest	**17**	**0**	**17**
Naïve Bayes	12	0	12
J48 – Bagging	4	0	4
J48 – Adaboost	4	0	4
J48 – Random Subspace	2	0	2
SVM – Random Subspace	2	0	2
SVM – Bagging	1	0	1
SVM – Adaboost	1	0	1
MLP	1	0	1
SVM	0	6	-6
Logistic Regression	0	9	-9
J48	0	29	-29

Table 13 - AUC Rankings

4.6.3 Discussion

With our experiment we demonstrated the efficacy of feature selection strategies in classification of hypertensive pathology from neuroimaging data. Every classifier tested ameliorated its performances when the input dataset was subjected to a feature selection strategy, with both Information based approaches and Correlation based approaches. On the other hand, feature extraction (PCA) always worsened the performances of the classifiers, probably because of the detrimental variance, not useful for classification, condensed in the extracted dimensions.

Naïve Bayes and Random Forest had better results overall, being non statistically inferior to any classifier in any condition and always in top 2 positions of the rankings for different metrics. Quite surprisingly the addition of cognitive tests scores worsened the performances of the classifiers, suggesting that the imaging marker of damage while correlated is not strictly linked to the cognitive tests, which can be altered also by emotive reaction during the tests and thus be less objective.

Our best setup can discriminate with accuracy greater than 90% the cardiovascular pathology only based on the signature of damage it exerts on the brain, both in functional and microstructural way. With this work we identified a candidate biomarker composed of the subsequent features:

- Right Superior Longitudinal Fasciculus – Fractional Anisotropy
- Forceps Minor – Fractional Anisotropy
- Right Anterior Thalamic Radiation – Mean Diffusivity
- Right Superior Longitudinal Fasciculus – Mean Diffusivity
- Forceps Major – Mean Diffusivity
- Right Anterior Thalamic Radiation – Radial Diffusivity
- Default Mode LP – Visual Occipital – Functional Connectivity
- Sensorimotor Lateral Right – Language pSTG Left – Functional Connectivity
- Sensorimotor Superior – Salience Acc. – Functional Connectivity
- Sensorimotor Superior – Dorsal Attention IPS Right – Functional Connectivity
- Visual Medial – Salience Acc. – Functional Connectivity
- Salience RPFC Left – FrontoParietal LPFC Left – Functional Connectivity
- Salience RPFC Left – FrontoParietal PPC Left – Functional Connectivity
- Salience RPFC Left – Language IFG Right – Functional Connectivity
- Dorsal Attention IPS Right – FrontoParietal LPFC Right – Functional Connectivity

We are performing follow up visits at one and four years, in order to monitor the progression of the cognitive impairment and of the parameters of our identified biomarker, with the addition of the absolute quantification of WMH for macrostructural damage with the system established in the first part of this thesis.

CONCLUSIONS

Application of our pipeline analysis to two clinical cases

The present body of work has been implemented in a continuous development together with the clinical recruitment of patients in the context of the heart and brain research program in I.R.C.C.S. Neuromed. In particular, this toolset was optimized for a new project in which we perform the global brain injury characterization. In this conclusion we will report the results obtained on the first two recruited patients, showing the results of the segmentation network proposed in Chapter 2, the diffusion and functional parameters obtained with the pipelines setup and presented in Chapter 3, with particular attention to the ensemble of features proposed in 4.6.3 which will be the subset of features evaluated at baseline and follow up to establish whether the hypertension associated brain injury has worsened or not after 1 year follow up.

RF004

The first patient which completed all the procedures for the ongoing project is a 43 year old female, presenting a controlled hypertension (132 SBP, 89 DBP) diagnosed 2 years ago, smoker, not presenting any carotid stenosis or significant wall thickening, no metabolic syndromes nor renal damage. The patient shows no significant cognitive decline, scoring 30/30 on Minimental cognitive test and 25/30 at MoCA test, suggesting preserved cognitive functions. Moreover, no errors were evidenced at stroop interference test, which was completed in 31.5 seconds. The clinical assessments suggest an hypertensive patient whom pressure control is optimal, with mostly preserved cognitive functions. We performed brain imaging to assess white matter hyperintensities, diffusion and functional connectivity parameters.

The analysis of WMH was carried on applying the same preprocessing steps reported in Chapter 2 to images acquired by our scanner, then we applied the FractalNet trained with Dice Normalized loss and obtained our WMH segmentation. One first issue raised by this process is the difference in basal noise and contrast between challenge images and our acquired images. This resulted in many spurious lesion segmentations located on the white matter – gray matter interfaces and gray matter – background interface. While we are working on the sequence to optimize the FLAIR contrast, we applied the workaround to mask the lesion segmentation on the white matter. This ensures that WMH lesions are correctly localized and spurious segmentations are avoided. An example of this results is shown in Figure 32.

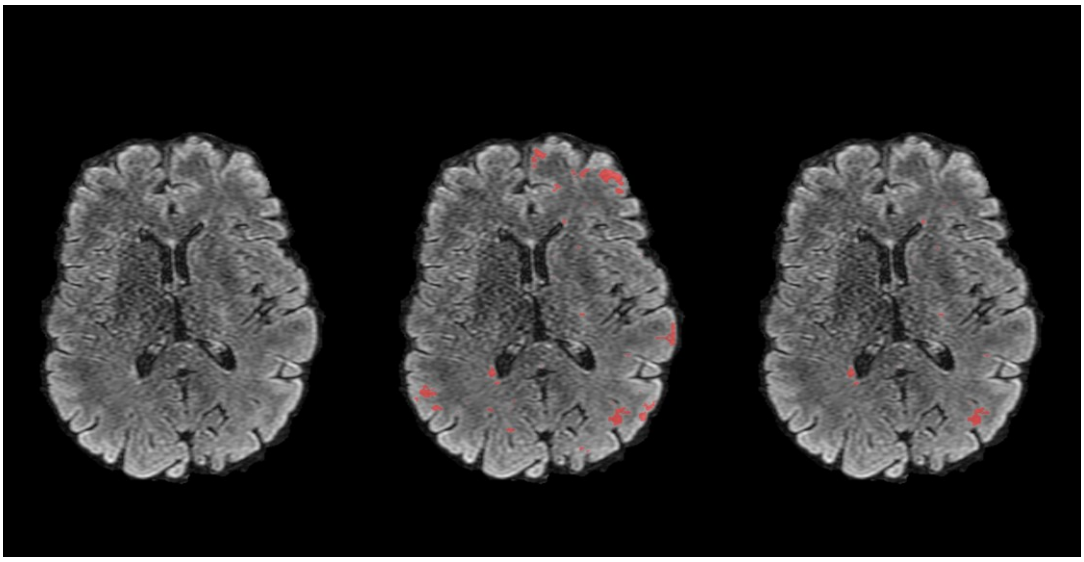


Figure 32 - Input image, raw WMH segmentation and masked WMH segmentation of patient RF004

After WMH, we characterized the brain for its diffusion connectivity and functional connectivity, extracting and reporting here the proposed feature ensemble:

- Right SLF – FA: 0.331
- FMI – FA: 0.409
- Right ATR – MD: $0.916 * 10^{-3}$
- Right SLF – MD: $0.871 * 10^{-3}$
- FMA – MD: $1.047 * 10^{-3}$

- Right ATR – RD: $0.742 * 10^{-3}$
- Default Mode LP – Visual Occipital – FC: -0.061
- Sensorimotor Lateral Right – Language pSTG Left – FC: 0.035
- Sensorimotor Superior – Salience Acc. – FC: -0.213
- Sensorimotor Superior – Dorsal Attention IPS Right – FC: 0.504
- Visual Medial – Salience Acc. – FC: 0.150
- Salience RPFC Left – FrontoParietal LPFC Left – FC: 0.323
- Salience RPFC Left – FrontoParietal PPC Left – FC: 0.123
- Salience RPFC Left – Language IFG Right – FC: 0.161
- Dorsal Attention IPS Right – FrontoParietal LPFC Right – FC: 0.779

RF009

The second patient which completed all the procedures for the project is a 47 year old female, presenting controlled hypertension (126 SBP, 77 DBP) diagnosed 5 years ago, no smoking, not presenting any carotid stenosis or significant wall thickening, no metabolic syndromes nor renal damage. The patient shows hints of cognitive decline, scoring 26/30 on Minimental cognitive test and 24/30 at MoCA test, suggesting declining cognitive functions mainly ascribed to the low score on Minimental test. Moreover, 3 errors were evidenced at stroop interference test, which was completed in 58 seconds. The clinical assessments suggest an hypertensive patient whom pressure control is optimal, with declining cognitive functions. We performed brain imaging to assess white matter hyperintensities, diffusion and functional connectivity parameters.

We applied the same workaround used in RF004 to ensure an optimal WMH segmentation, as shown in Figure 33.

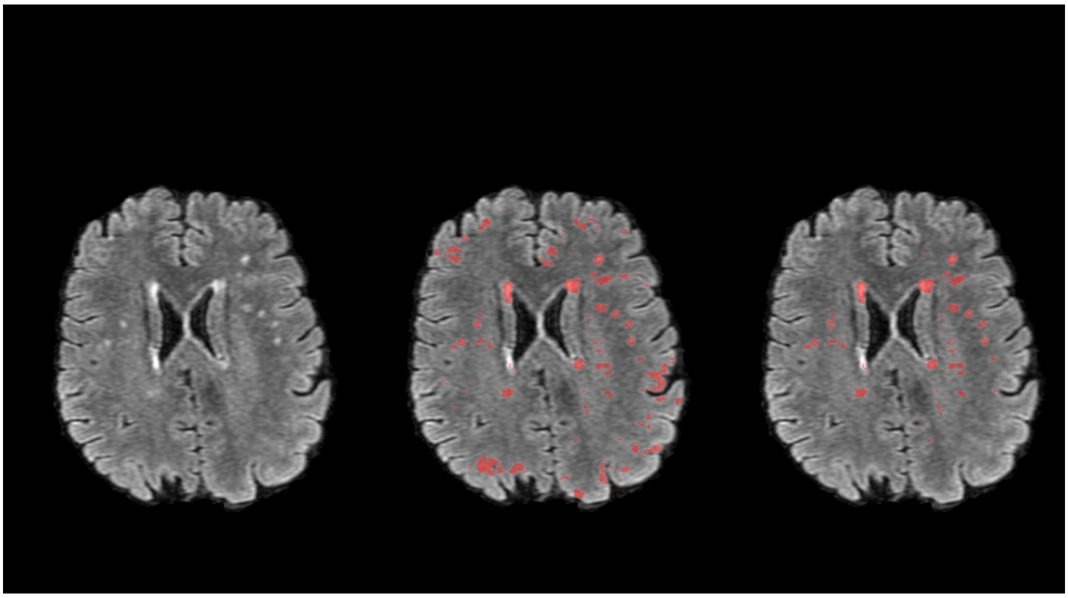


Figure 33 - Input image, raw WMH segmentation and masked WMH segmentation of patient RF009

After WMH, we characterized the brain for its diffusion connectivity and functional connectivity, extracting and reporting here the proposed feature ensemble:

- Right SLF – FA: 0.324
- FMI – FA: 0.447
- Right ATR – MD: 0.958 * 10^{-3}
- Right SLF – MD: 0.867 * 10^{-3}
- FMA – MD: 1.112 * 10^{-3}
- Right ATR – RD: 0.794 * 10^{-3}
- Default Mode LP – Visual Occipital – FC: -0.400
- Sensorimotor Lateral Right – Language pSTG Left – FC: -0.009
- Sensorimotor Superior – Salience Acc. – FC: -0.194
- Sensorimotor Superior – Dorsal Attention IPS Right – FC: -0.035
- Visual Medial – Salience Acc. – FC: 0.315
- Salience RPFC Left – FrontoParietal LPFC Left – FC: 0.068
- Salience RPFC Left – FrontoParietal PPC Left – FC: 0.286
- Salience RPFC Left – Language IFG Right – FC: -0.092

- Dorsal Attention IPS Right – FrontoParietal LPFC Right – FC: -0.049

Conclusion

We applied our brain MRI analysis pipeline to two hypertensive patients recruited in a prospective study to evaluate the effects of hypertension on cognitive functions and brain structure and connectivity. The reported analysis are the baseline exams for the two patients which will be monitored yearly, to analyze how hypertension progression can impact the brain on a macrostructural (evaluated as WMH load) or microstructural and functional basis (evaluated by the proposed feature ensemble). From this baseline analysis, what is evident from on a per patient analysis is that a patient in which the cognitive functions are more challenged and which hypertension has been sustained for more years show a higher load of white matter hyperintensities and increased diffusion parameters as MD and RD, representing a less organized fiber structure in the white matter. The functional connectivity differences are less obvious, as the FC parameter is an index of synchronicity between different areas. These parameters will be more useful to perform group based analysis in a longitudinal setting.

APPENDIX A – MRI SEQUENCES

MRI Principles

Magnetic Resonance Imaging (MRI) was developed following the principle of magnetic resonance, discovered in 1946 independently by Felix Block and Edward Purcell. This principle leverages the intrinsic characteristic of some atoms, the one presenting a nuclear spin, which if are immersed in an external magnetic field and irradiated with radiofrequencies, they start resonating and give back energy in the form of radiofrequency. The characteristics of the emitted signal let us infer information about physical and chemical composition of the excited element. This principle has been first exploited to perform Magnetic Resonance spectroscopy and then, following technological upgrades which included the modulation of the external magnetic field to add spatial encoding of exciting signal, MR evolved to MRI and largely penetrated the clinical practice. After the introduction of the imaging sequences in 1973 by Lauterbur, the field acquired a substantial stability in 1975 thanks to Richard Ernst and the introduction of the Fourier Imaging, whose principle is still now fundamentally unaltered and at the foundation of modern MRI.

Every atom is composed by a certain number of particles and each of them has a property called spin, which is the angular momentum of each particle rotation. Depending on the direction of the momentum the spin can be positive or negative, and they can be "summed" to get the total spin of the nucleus. Non-zero spin nuclei can be investigated by NMR and by extension by MRI. The global behaviour of these nuclei is assumed as the one of small dipoles with a random orientation, thus having a random orientation the global magnetization cancels out and the net magnetization is zero. If the nuclei are immersed in an external magnetic field, the small dipoles will align to the external field and produce a net magnetization aligned with the direction of the magnetic field.

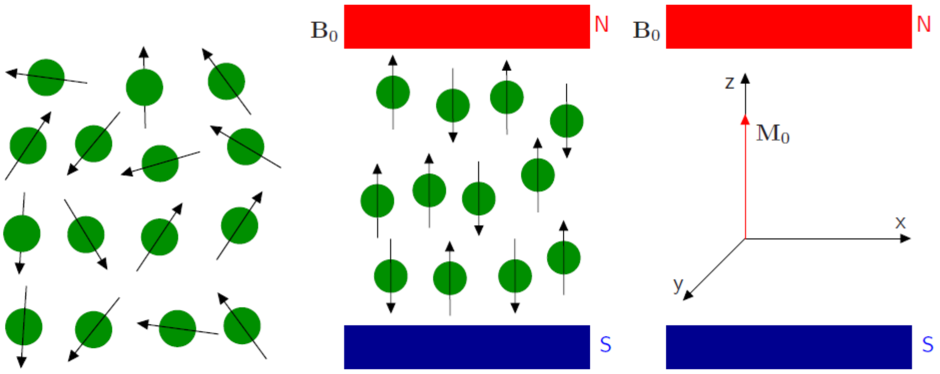


Figure 34

Defined M_0 the vector representing the net magnetization of the volume taken into account, and B0 the external magnetic field, we can define the longitudinal component of M_0 as M_z, aligned along B0, and the transversal component of M_0 as M_T, which is perpendicular M_0 direction.

When the nuclei are supplied with energy the longitudinal part of net magnetization becomes lower than the value at equilibrium state, due to a dealignment of the nuclei dipoles. The time in which the system goes back to the equilibrium is defined as T1 Relaxation time or spin-lattice relaxation time.

When considering the single protons in the nuclei, while the global small dipole is aligned to B0, we have a certain angle of dealignment between them and B0. This angle produces a torque between the magnetic field B and the proton magnetic moment, causing a precession of the proton magnetic moment and subsequently of the net magnetization vectors M around the z-axis at the frequency of resonance of the system. The time needed after an excitation pulse to restore the equilibrium magnitude of the transverse plane magnetization influenced by this precession process is called T2-relaxation time or spin-spin relaxation time. This parameter is highly influenced by the inhomogeneity of the magnetic field and the susceptibility variation across the different tissues.

By transmitting a radio-frequency pulse (RF pulse) through a transmitting RF coil at the resonance frequency energy is added to the system. A correct amount of energy at the resonance frequency flips the magnetization vector M0 from the longitudinal orientation to the transverse plane (xy-plane), flipping the net magnetization of 90°.

Precession of the magnetization vector in the xy-plane is measured through the oscillating electromagnetic field it generates.

After the 90◦ pulse, the precessing motion of M generates an alternating current in a receiver coil placed around the x-axis and perpendicular to the y-axis, see Figure 2.6. The measured signal is called free induction decay (FID). Combinations of RF pulses and FID signal reading are performed to read the NMR signal. The addition of gradient magnet system let us alterate the punctual magnetic field, changing the frequency of resonance in different spatial points, thus exciting selectively a volume of interest (the voxels in the resulting images). The combination of gradient field modulation and MR signal in each different volume of interest makes possible reconstructing images leveraging spatial codification and FT of the received signal.

T2-FLAIR

T2-weighted fluid attenuated inversion recovery is a sequence of the inversion recovery family. This kind of pulse sequence leverage a succession of 180° and 90° excitation pulses. The first pulse is a 180°, which induces an inversion of magnetization and a subsequent realignment along B0 as shown in Figure 35. Then, during the relaxation process, a subsequent 90° pulse is provided, after the Time of

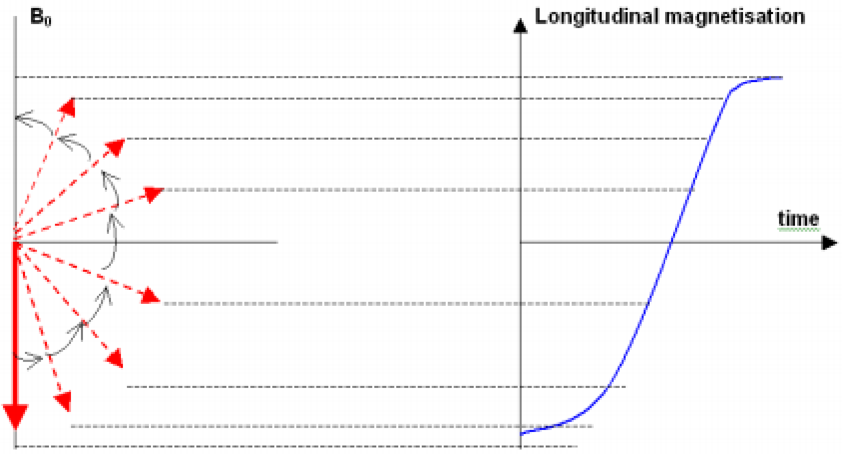


Figure 35

Inversion (TI). This pulse will delay the magnetization of relaxing tissues, and selecting accurately the TI we will be able to bring to 0 the magnetization for tissues of choice (as in Figure 36 for water).

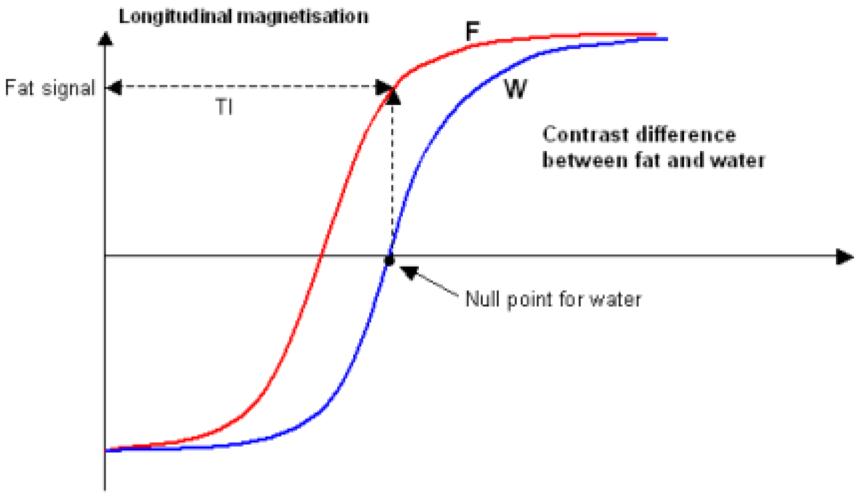


Figure 36

T2-FLAIR uses a TI accurately selected to nullify the magnetization of cerebrospinal fluid, to highlight the fluid component from the inner region of white matter.

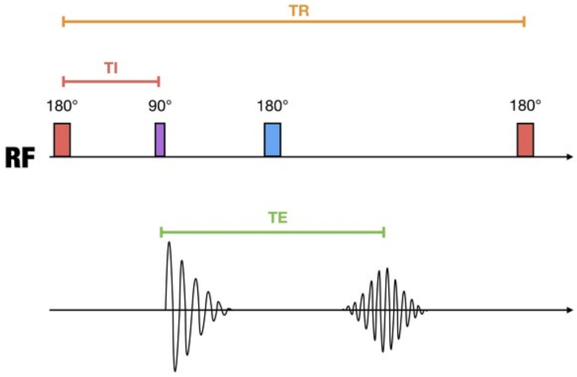


Figure 37 – Pulse sequence of FLAIR sequence

DIFFUSION TENSOR IMAGING (DTI)

The diffusion tensor was originally proposed for use in magnetic resonance imaging (MRI) by Peter Basser in 1994[39]. Before DTI, diffusion MRI had developed from research in diffusion nuclear magnetic resonance. Prior to the introduction of the diffusion tensor model, to measure anisotropic diffusion the orientation of the axons in a tissue sample had to be known. The introduction of the diffusion tensor model allowed, for the first time, a rotationally invariant description of the shape of water diffusion. The invariance to rotation was crucial because it enabled application of the DTI method to the complex anatomy of the fiber tracts in the human brain.

The popularity of DTI has been enormous. It has been applied to a tremendous variety of neuroscientific studies including schizophrenia, traumatic brain injury, multiple sclerosis, autism, and aging. Anatomical investigations have been undertaken regarding for example the structure of the language network, the asymmetry of the white matter in twins and siblings, and the location, asymmetry, and variability of the fiber tracts. Recent investigations have attempted to model the human "connectome" by analysing structural versus functional brain connectivity as measured by DTI and functional MRI. DTI has also been applied for neurosurgical planning and navigation. The addition of preoperative DTI to neuronavigation has been shown, in a large prospective study, to increase tumor resection and survival and to decrease neurologic morbidity. Our work represents one of the first study to apply DTI to investigate the consequences of hypertension of brain structural connectivity.

Tensorial model

Diffusion is a random transport phenomenon, which describes the transfer of material (e.g., water molecules) from one spatial location to other locations over time. In three dimensions, the Einstein diffusion equation:

$$D = \frac{\langle \Delta r^2 \rangle}{2n\Delta t}$$

states that the diffusion coefficient, D (in mm2/s), is proportional to the mean squared displacement, $\langle \Delta r^2 \rangle$ divided by the number of dimensions, n, and the diffusion time, Δt. The diffusion coefficient of pure water at 20°C is roughly 2.0 × 10-3mm2/s and increases at higher temperatures. In the absence of boundaries, the molecular water displacement is described by a Gaussian probability density

$$P(\Delta r, \Delta t) = \frac{1}{\sqrt{(2\pi D \Delta t)^3}} * e^{\frac{-\Delta r^2}{4D\Delta t}}$$

The diffusion of water in biological tissues occurs inside, outside, around, and through cellular structures. Water diffusion is primarily caused by random thermal fluctuations. The behaviour is further modulated by the interactions with cellular membranes, and subcellular and organelles. Cellular membranes hinder the diffusion of water, causing water to take more tortuous paths, thereby decreasing the mean squared displacement. The diffusion tortuosity and corresponding apparent diffusivity may be increased by either cellular swelling or increased cellular density. Conversely, necrosis, which results in a breakdown of cellular membranes, decreases tortuosity and increases the apparent diffusivity.

In fibrous tissues including white matter, water diffusion is relatively unimpeded in the direction parallel to the fiber orientation. Conversely, water diffusion is highly restricted and hindered in the directions perpendicular to the fibers. Thus, the diffusion in fibrous tissues is anisotropic.

The application of the diffusion tensor to describe anisotropic diffusion behaviour was introduced by Basser. In this model, diffusion is described by a multivariate normal distribution

$$P(\Delta r, \Delta t) = \frac{1}{\sqrt{(4\pi \Delta t)^3 \, |\boldsymbol{D}|}} * e^{\frac{-\Delta r^T \boldsymbol{D}^{-1} \Delta r}{4\Delta t}}$$

where the diffusion tensor is a 3x3 covariance matrix

$$\boldsymbol{D} = \begin{bmatrix} D_{XX} & D_{XY} & D_{XZ} \\ D_{YX} & D_{YY} & D_{YZ} \\ D_{ZX} & D_{ZY} & D_{ZZ} \end{bmatrix}$$

which describes the covariance of diffusion displacements in 3D normalized by the diffusion time. The diagonal elements ($D_{ii} > 0$) are the diffusion variances along the x, y and z axes, and the off-diagonal elements are the covariance terms and are symmetric about the diagonal ($D_{ij} = D_{ji}$). Diagonalization of the diffusion tensor yields the eigenvalues (λ_1, λ_2, λ_3) and corresponding eigenvectors (ε_1, ε_2, ε_3) of the diffusion tensor, which describe the directions and apparent diffusivities along the axes of principle diffusion. The diffusion tensor may be visualized using an ellipsoid with the eigenvectors defining the directions of the principle axes and the ellipsoidal radii defined by the eigenvalues (Figure 38). Diffusion is considered isotropic when the eigenvalues are nearly equal (e.g., $\lambda_1 \sim \lambda_2 \sim \lambda_3$). Conversely, the diffusion tensor is anisotropic when the eigenvalues are significantly different in magnitude (e.g., $\lambda_1 > \lambda_2 > \lambda_3$). The eigenvalue magnitudes may be affected by changes in local tissue microstructure with many types of tissue injury, disease or normal physiological changes (i.e., aging). Thus, the diffusion tensor is a sensitive probe for characterizing both normal and abnormal tissue microstructure.

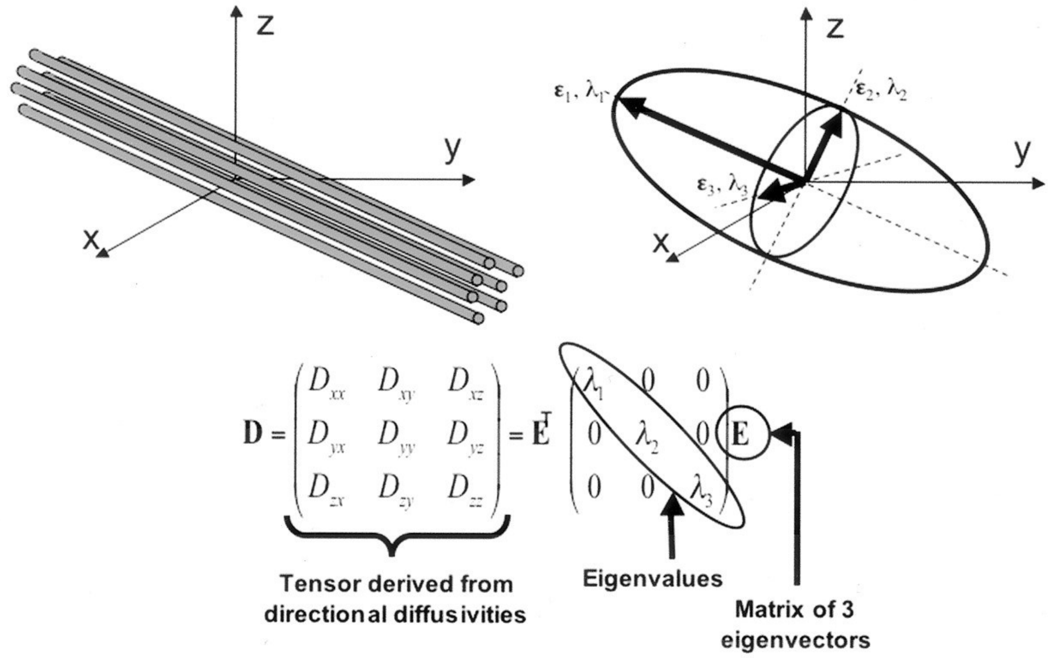


Figure 38 - Diffusion Tensor

Applying the tensorial model to the raw diffusion weighted images, we can obtain maps of the DTI measures. Being unpractical to visualize a 3x3 matrix for each voxel, we have break down the diffusion matrix in multiple measures.

region.

RESTING STATE FUNCTIONAL MRI (rs-fMRI)

Functional MRI is a technique developed in the early nineties following the studies of Ogawa, which discovered the blood-oxygen-level dependent signal (BOLD signal)[70]. This signal is derived from the ratio between deoxyhemoglobine and oxyhemoglobine in a given volume of space and their differential magnetic susceptibility let us estimate their relative concentrations. This ratio is fundamental in the brains to understand the regions in which is undergoing a neuronal activation. The principle of hemodynamic response (HDR) states that neuronal activity, for which is needed an increase of nutrients flow, is reflected in a variation of oxygenated blood flow, even though with a temporal delay due to vascular system regulation (Figure 39).

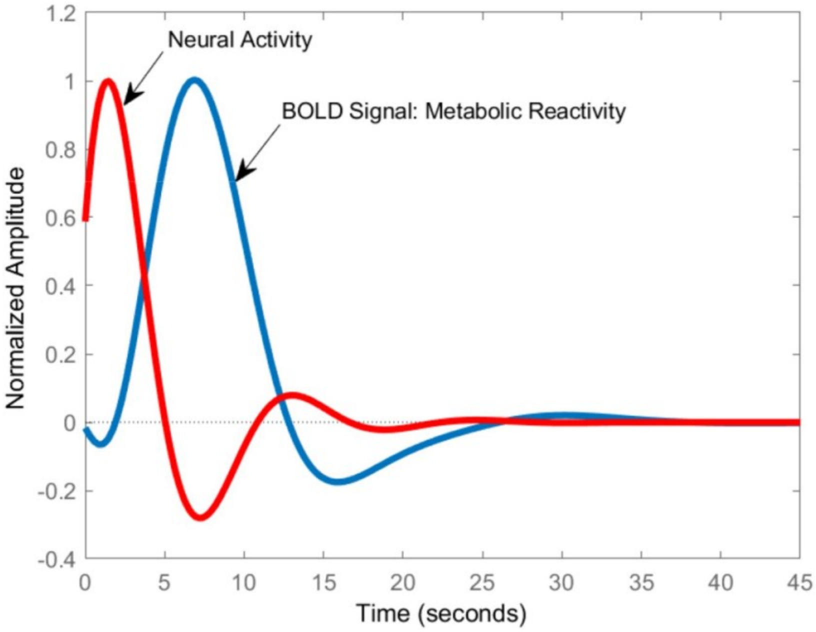


Figure 39 - Hemodynamic Response (HDR)

Exploiting this neurovascular coupling, it is possible to map the regional activations of the human brain and understand the functions associated to the different brain regions. The first experiments using this technique were aimed at understanding which were the regions whom activity was evoked in response to simple tasks which could be conducted while lying in the MRI scanner, such as finger tapping for motor mapping, image recognition for visual cortex mapping[47].

Another potent tool has been the introduction of resting-state functional MRI (rs-fMRI). The concept behind it is that even when the brain is not challenged with any active task, there is a spontaneous brain activity which can be reconstructed and associated to the steady state activity of different brain circuits which are eventually recruited when we are challenged with different tasks.

By acquiring several minutes of functional MRI scans, we can obtain timeseries of spontaneous brain activity and, by evaluating the synchronicity and correlation of the activation time course, we can estimate the synchronous regions and the networks they compose.

APPENDIX B – NETWORK STRUCTURES

Unet3 Structure

```
def get_conv(f=16):
    img_rows = 80
    img_cols = 80
    smooth = 1.
    learning_rate = 0.001
    inputs = Input((img_rows, img_cols, 1))
    conv1 = Convolution2D(f, 5, 5, activation='relu', border_mode='same')(inputs)
    conv1 = BatchNormalization()(conv1)
    conv1 = Convolution2D(f, 5, 5, activation='relu', border_mode='same')(conv1)
    conv2 = MaxPooling2D(pool_size=(2, 2))(conv1)
    conv2 = BatchNormalization()(conv2)
    conv2 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same')(conv2)
    conv2 = BatchNormalization()(conv2)
    conv2 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same')(conv2)
    conv3 = MaxPooling2D(pool_size=(2, 2))(conv2)
    conv3 = BatchNormalization()(conv3)
    conv3 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same')(conv3)
    conv3 = BatchNormalization()(conv3)
    conv3 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same')(conv3)
    conv4 = MaxPooling2D(pool_size=(2, 2))(conv3)
    conv4 = BatchNormalization()(conv4)
    conv4 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same')(conv4)
    conv4 = BatchNormalization()(conv4)
    conv4 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same')(conv4)
    conv5 = MaxPooling2D(pool_size=(2, 2))(conv4)
    conv5 = BatchNormalization()(conv5)
    conv5 = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same')(conv5)
    conv5 = BatchNormalization()(conv5)
    conv5 = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same')(conv5)
    up1 = merge([UpSampling2D(size=(2, 2))(conv5), conv4], mode='concat', concat_axis=3)
    conv6 = BatchNormalization()(up1)
    conv6 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same')(conv6)
    conv6 = BatchNormalization()(conv6)
    conv6 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same')(conv6)
    up2 = merge([UpSampling2D(size=(2, 2))(conv6), conv3], mode='concat', concat_axis=3)
    conv7 = BatchNormalization()(up2)
    conv7 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same')(conv7)
    conv7 = BatchNormalization()(conv7)
    conv7 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same')(conv7)
    up3 = merge([UpSampling2D(size=(2, 2))(conv7), conv2], mode='concat', concat_axis=3)
    conv8 = BatchNormalization()(up3)
    conv8 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same')(conv8)
    conv8 = BatchNormalization()(conv8)
    conv8 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same')(conv8)
    up4 = merge([UpSampling2D(size=(2, 2))(conv8), conv1], mode='concat', concat_axis=3)
    conv9 = BatchNormalization()(up4)
    conv9 = Convolution2D(f, 5, 5, activation='relu', border_mode='same')(conv9)
    conv9 = BatchNormalization()(conv9)
    outputs = Convolution2D(1, 1, 1, activation='hard_sigmoid',

                              border_mode='same')(conv9)
    outputs = Convolution2D(1, 2, 2, activation='hard_sigmoid',
                              border_mode='same')(outputs)
    adam = optimizers.Adam(lr=learning_rate, beta_1=0.9, beta_2=0.999, epsilon=None)
    net = Model(inputs=inputs, outputs=outputs)
    net.compile(loss=dice_coef_loss, optimizer=adam)
    net.summary()
    return net
```

FractalNet Structure

```
def get_fractalunet(f=16):
    img_rows = 80
    img_cols = 80
    inputs = Input((img_rows, img_cols, 1))

    conv1 = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(inputs)
    conv1 = BatchNormalization()(conv1)
    conv1 = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv1)

    down1 = MaxPooling2D(pool_size=(2, 2))(conv1)

    conv2 = BatchNormalization()(down1)
    conv2 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv2)
    conv2 = BatchNormalization()(conv2)
    conv2 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv2)

    down2 = MaxPooling2D(pool_size=(2, 2))(conv2)

    conv3 = BatchNormalization()(down2)
    conv3 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv3)
    conv3 = BatchNormalization()(conv3)
    conv3 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv3)

    down3 = MaxPooling2D(pool_size=(2, 2))(conv3)

    conv4 = BatchNormalization()(down3)
    conv4 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv4)
    conv4 = BatchNormalization()(conv4)
    conv4 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv4)

    down4 = MaxPooling2D(pool_size=(2, 2))(conv4)

    conv5 = BatchNormalization()(down4)
    conv5 = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv5)
    conv5 = BatchNormalization()(conv5)
    conv5 = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv5)

    up1 = merge([UpSampling2D(size=(2, 2))(conv5), conv4], mode='concat', concat_axis=3)

    conv6 = BatchNormalization()(up1)
```

```
conv6 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv6)
conv6 = BatchNormalization()(conv6)
conv6 = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv6)

up2 = merge([UpSampling2D(size=(2, 2))(conv6), conv3], mode='concat', concat_axis=3)

conv7 = BatchNormalization()(up2)
conv7 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv7)
conv7 = BatchNormalization()(conv7)
conv7 = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv7)

up3 = merge([UpSampling2D(size=(2, 2))(conv7), conv2], mode='concat', concat_axis=3)

conv8 = BatchNormalization()(up3)
conv8 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv8)
conv8 = BatchNormalization()(conv8)
conv8 = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv8)

up4 = merge([UpSampling2D(size=(2, 2))(conv8), conv1], mode='concat', concat_axis=3)

conv9 = BatchNormalization()(up4)
conv9 = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv9)
conv9 = BatchNormalization()(conv9)
conv9 = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv9)

# --- end first u block

down1b = MaxPooling2D(pool_size=(2, 2))(conv9)
down1b = merge([down1b, conv8], mode='concat', concat_axis=3)

conv2b = BatchNormalization()(down1b)
conv2b = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv2b)
conv2b = BatchNormalization()(conv2b)
conv2b = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv2b)

down2b = MaxPooling2D(pool_size=(2, 2))(conv2b)
down2b = merge([down2b, conv7], mode='concat', concat_axis=3)

conv3b = BatchNormalization()(down2b)
conv3b = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv3b)
conv3b = BatchNormalization()(conv3b)
conv3b = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                      kernel_initializer='glorot_normal',
                      bias_initializer='normal')(conv3b)
down3b = MaxPooling2D(pool_size=(2, 2))(conv3b)
down3b = merge([down3b, conv6], mode='concat', concat_axis=3)
```

```
conv4b = BatchNormalization()(down3b)
    conv4b = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv4b)
    conv4b = BatchNormalization()(conv4b)
    conv4b = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv4b)

    down4b = MaxPooling2D(pool_size=(2, 2))(conv4b)
    down4b = merge([down4b, conv5], mode='concat', concat_axis=3)

    conv5b = BatchNormalization()(down4b)
    conv5b = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv5b)
    conv5b = BatchNormalization()(conv5b)
    conv5b = Convolution2D(16 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv5b)

    up1b = merge([UpSampling2D(size=(2, 2))(conv5b), conv4b], mode='concat',
                          concat_axis=3)

    conv6b = BatchNormalization()(up1b)
    conv6b = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv6b)
    conv6b = BatchNormalization()(conv6b)
    conv6b = Convolution2D(8 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv6b)

    up2b = merge([UpSampling2D(size=(2, 2))(conv6b), conv3b], mode='concat',
                          concat_axis=3)

    conv7b = BatchNormalization()(up2b)
    conv7b = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv7b)
    conv7b = BatchNormalization()(conv7b)
    conv7b = Convolution2D(4 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv7b)

    up3b = merge([UpSampling2D(size=(2, 2))(conv7b), conv2b], mode='concat',
                          concat_axis=3)

    conv8b = BatchNormalization()(up3b)
    conv8b = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv8b)
    conv8b = BatchNormalization()(conv8b)
    conv8b = Convolution2D(2 * f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv8b)

    up4b = merge([UpSampling2D(size=(2, 2))(conv8b), conv9], mode='concat',
                          concat_axis=3)
    conv9b = BatchNormalization()(up4b)
    conv9b = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv9b)
    conv9b = BatchNormalization()(conv9b)
    conv9b = Convolution2D(f, 5, 5, activation='relu', border_mode='same',
                          kernel_initializer='glorot_normal',
                          bias_initializer='normal')(conv9b)
    conv9b = BatchNormalization()(conv9b)
```

```
outputs = Convolution2D(1, 1, 1, activation='hard_sigmoid',
                     border_mode='same')(conv9b)
adam = optimizers.Adam(lr=learning_rate, beta_1=0.9, beta_2=0.999, epsilon=None)

net = Model(inputs=inputs, outputs=outputs)
net.compile(loss=dice_coef_loss, optimizer=adam)
net.summary()

return net
```

REFERENCES

[1] W. H. Organization, *Dementia: a public health priority*. World Health Organization, 2012.

[2] G. Faraco and C. Iadecola, "Hypertension: a harbinger of stroke and dementia," *Hypertension,* vol. 62, no. 5, pp. 810-817, 2013.

[3] M. M. Santisteban and C. Iadecola, "Hypertension, dietary salt and cognitive impairment," *Journal of Cerebral Blood Flow & Metabolism,* vol. 38, no. 12, pp. 2112-2128, 2018.

[4] C. Iadecola *et al.*, "Impact of hypertension on cognitive function: a scientific statement from the American Heart Association," *Hypertension,* vol. 68, no. 6, pp. e67-e94, 2016.

[5] P. B. Gorelick *et al.*, "Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association," *Stroke,* vol. 42, no. 9, pp. 2672-2713, 2011.

[6] P. K. Whelton *et al.*, "2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines," (in eng), *Hypertension,* vol. 71, no. 6, pp. e13-e115, Jun 2018.

[7] B. Williams *et al.*, "2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology (ESC) and the European Society of Hypertension (ESH)," *European heart journal,* vol. 39, no. 33, pp. 3021-3104, 2018.

[8] T. S. M. I. f. t. S. R. Group, "Effect of Intensive vs Standard Blood Pressure Control on Probable Dementia: A Randomized Clinical Trial," *JAMA,* vol. 321, no. 6, pp. 553-561, 2019.

[9] L. Carnevale *et al.*, "Brain MRI fiber-tracking reveals white matter alterations in hypertensive patients without damage at conventional neuroimaging," *Cardiovascular research,* vol. 114, no. 11, pp. 1536-1546, 2018.

[10] C. Iadecola and R. F. Gottesman, "Neurovascular and cognitive dysfunction in hypertension: epidemiology, pathobiology, and treatment," *Circulation research,* vol. 124, no. 7, pp. 1025-1044, 2019.

[11] C. Iadecola, "The neurovascular unit coming of age: a journey through neurovascular coupling in health and disease," *Neuron,* vol. 96, no. 1, pp. 17-42, 2017.

[12] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith, "Fsl," *Neuroimage,* vol. 62, no. 2, pp. 782-790, 2012.

[13] A. M. Dale, B. Fischl, and M. I. Sereno, "Cortical surface-based analysis: I. Segmentation and surface reconstruction," *Neuroimage,* vol. 9, no. 2, pp. 179-194, 1999.

[14] P. Scheltens *et al.*, "Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates," *Journal of Neurology, Neurosurgery and Psychiatry,* Article vol. 55, no. 10, pp. 967-972, 1992.

[15] S. O. Casey, R. C. Sampaio, E. Michel, and C. L. Truwit, "Posterior reversible encephalopathy syndrome: Utility of fluid-attenuated inversion recovery mr imaging in the detection of cortical and subcortical lesions," *American Journal of Neuroradiology,* Article vol. 21, no. 7, pp. 1199-1206, 2000.

[16] A. A. Gouw *et al.*, "Heterogeneity of small vessel disease: a systematic review of MRI and histopathology correlations," *Journal of Neurology, Neurosurgery & Psychiatry,* vol. 82, no. 2, pp. 126-135, 2011.

[17] P. Scheltens *et al.*, "A semiquantative rating scale for the assessment of signal hyperintensities on magnetic resonance imaging," *Journal of the Neurological Sciences,* vol. 114, no. 1, pp. 7-12, 1993/01/01/ 1993.

[18] L. Griffanti *et al.*, "BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities," *NeuroImage,* vol. 141, pp. 191-205, 2016/11/01/ 2016.

[19] H. J. Kuijf *et al.*, "Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities; Results of the WMH Segmentation Challenge," *IEEE Transactions on Medical Imaging,* vol. 0062, no. c, pp. 1-1, 2019.

[20] N. Raz, K. M. Rodrigue, and J. D. Acker, "Hypertension and the brain: vulnerability of the prefrontal regions and executive functions," *Behavioral neuroscience,* vol. 117, no. 6, p. 1169, 2003.

[21] R. Schmidt *et al.*, "Magnetic resonance imaging cerebral abnormalities and neuropsychologic test performance in elderly hypertensive subjects: a case-control study," *Archives of Neurology,* vol. 52, no. 9, pp. 905-910, 1995.

[22] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, 2012.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026-1034.

[25] A. Bria, N. Karssemeijer, and F. Tortorella, "Learning from unbalanced data: A cascade-based approach for detecting clustered microcalcifications," *Medical Image Analysis,* Article vol. 18, no. 2, pp. 241-252, 2014.

[26] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature,* vol. 542, no. 7639, p. 115, 2017.

[27] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med Image Anal,* vol. 42, pp. 60-88, Dec 2017.

[28] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik,* vol. 29, no. 2, pp. 102-127, 2019/05/01/ 2019.

[29] F. Fazekas, J. B. Chawluk, A. Alavi, H. I. Hurtig, and R. A. Zimmerman, "MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging," *American journal of roentgenology,* vol. 149, no. 2, pp. 351-356, 1987.

[30] H. J. Kuijf *et al.*, "Standardized assessment of automatic segmentation of white matter hyperintensities; results of the wmh segmentation challenge," *IEEE transactions on medical imaging,* 2019.

[31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241: Springer.

[32] D. Liciotti, M. Paolanti, R. Pietrini, E. Frontoni, and P. Zingaretti, "Convolutional networks for semantic heads segmentation using top-view depth data in crowded environment," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 1384-1389: IEEE.

[33] G. Larsson, M. Maire, and G. Shakhnarovich, "Fractalnet: Ultra-deep neural networks without residuals," *arXiv preprint arXiv:1605.07648,* 2016.

[34] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249-256.

[35] S. Sadegh Mohseni Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," *arXiv,* p. arXiv: 1706.05721, 2017.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980,* 2014.

[37] Z. S. Nasreddine *et al.*, "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society,* vol. 53, no. 4, pp. 695-699, 2005.

[38] P. J. Basser and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI," *Journal of Magnetic Resonance - Series B,* Article vol. 111, no. 3, pp. 209-219, 1996.

[39] P. J. Basser, J. Mattiello, and D. LeBihan, "MR diffusion tensor spectroscopy and imaging," *Biophysical journal,* vol. 66, no. 1, pp. 259-267, 1994.

[40] A. L. Alexander, J. E. Lee, M. Lazar, and A. S. Field, "Diffusion tensor imaging of the brain," *Neurotherapeutics,* vol. 4, no. 3, pp. 316-29, Jul 2007.

[41] S. Jbabdi, S. N. Sotiropoulos, A. M. Savio, M. Graña, and T. E. Behrens, "Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems," *Magnetic resonance in medicine,* vol. 68, no. 6, pp. 1846-1855, 2012.

[42] T. E. Behrens, H. J. Berg, S. Jbabdi, M. F. Rushworth, and M. W. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?," *Neuroimage,* vol. 34, no. 1, pp. 144-155, 2007.

[43] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping,* vol. 17, no. 3, pp. 143-155, 2002.

[44] J. L. Andersson and S. N. Sotiropoulos, "An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging," *Neuroimage,* vol. 125, pp. 1063-1078, 2016.

[45] M. de Groot *et al.*, "Improving alignment in tract-based spatial statistics: evaluation and optimization of image registration," *Neuroimage,* vol. 76, pp. 400-411, 2013.

[46] M. de Groot *et al.*, "Tract-specific white matter degeneration in aging: the Rotterdam Study," *Alzheimer's & Dementia,* vol. 11, no. 3, pp. 321-330, 2015.

[47] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magnetic resonance in medicine,* vol. 34, no. 4, pp. 537-541, 1995.

[48] F. Agosta, M. Pievani, C. Geroldi, M. Copetti, G. B. Frisoni, and M. Filippi, "Resting state fMRI in Alzheimer's disease: beyond the default mode network," *Neurobiology of aging,* vol. 33, no. 8, pp. 1564-1578, 2012.

[49] M. E. Raichle, "The brain's default mode network," *Annual review of neuroscience,* vol. 38, pp. 433-447, 2015.

[50] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon, "Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI," *Proceedings of the National Academy of Sciences,* vol. 101, no. 13, pp. 4637-4642, 2004.

[51] B. T. T. Yeo *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," (in eng), *Journal of neurophysiology,* vol. 106, no. 3, pp. 1125-1165, 2011.

[52] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE transactions on medical imaging,* vol. 23, no. 2, pp. 137-152, 2004.

[53] J. Wang *et al.*, "The Role of Disturbed Small-World Networks in Patients with White Matter Lesions and Cognitive Impairment Revealed by Resting State Function Magnetic Resonance Images (rs-fMRI)," *Medical science monitor: international medical journal of experimental and clinical research,* vol. 25, p. 341, 2019.

[54] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage,* vol. 52, no. 3, pp. 1059-1069, 2010.

[55] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain connectivity,* vol. 2, no. 3, pp. 125-141, 2012.

[56] P. J. Bickel and E. Levina, "Some theory for Fisher's linear discriminant function,naive Bayes', and some alternatives when there are many more variables than observations," *Bernoulli,* vol. 10, no. 6, pp. 989-1010, 2004.

[57] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fMRI classification," *IEEE transactions on medical imaging,* vol. 29, no. 2, pp. 531-542, 2010.

[58] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter,* vol. 11, no. 1, pp. 10-18, 2009.

[59] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., 1993, p. 302.

[60] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, vol. 3, no. 22, pp. 41-46.

[61] R. E. Wright, "Logistic regression," 1995.

[62] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning,* vol. 20, no. 3, pp. 273-297, 1995.

[63] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *icml*, 1996, vol. 96, pp. 148-156: Citeseer.

[64] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, no. 2, pp. 123-140, 1996.

[65] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.

[66] T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 8, pp. 832-844, 1998.

[67] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.

[68] C. Lee and G. G. Lee, "Information gain and divergence-based feature selection for machine learning-based text categorization," *Information processing & management,* vol. 42, no. 1, pp. 155-165, 2006.

[69] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Transactions on Instrumentation and Measurement,* vol. 53, no. 6, pp. 1517-1525, 2004.

[70] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *proceedings of the National Academy of Sciences,* vol. 87, no. 24, pp. 9868-9872, 1990.